

Introduction-Numpy

December 25, 2021

1 Introduction to Numpy, Scipy, SciKit-Learn

In this tutorial we will look into the Numpy library: <http://www.numpy.org/>

Numpy is a very important library for numerical computations and matrix manipulation. It has a lot of the functionality of Matlab, and some of the functionality of Pandas

We will also use the Scipy library for scientific computation: <http://docs.scipy.org/doc/numpy/reference/index.html>

```
[15]: import numpy as np
import scipy as sp
import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns
%matplotlib inline
```

```
[16]: import scipy.sparse as sp_sparse
import scipy.spatial.distance as sp_dist
import sklearn as sk
import sklearn.datasets as sk_data
import sklearn.metrics as metrics
from sklearn import preprocessing
import scipy.sparse.linalg as linalg
```

1.0.1 Why Numpy?

```
[20]: import time

def trad_version():
    t1 = time.time()
    X = range(10000000)
    Y = range(10000000)
    Z = [x+y for x,y in zip(X,Y)]
    return time.time() - t1
```

```

def naive_numpy_version():
    t1 = time.time()
    X = np.arange(10000000)
    Y = np.arange(10000000)
    Z = np.zeros(10000000)
    for i in range(10000000):
        Z[i] = X[i]+Y[i]
    return time.time() - t1

def numpy_version():
    t1 = time.time()
    X = np.arange(10000000)
    Y = np.arange(10000000)
    Z = X + Y
    return time.time() - t1

traditional_time = trad_version()
naive_numpy_time = naive_numpy_version()
numpy_time = numpy_version()
print ("Traditional time = "+ str(traditional_time))
print ("Naive numpy time      = "+ str(naive_numpy_time))
print ("Numpy time           = "+ str(numpy_time))

```

```

Traditional time = 1.6193649768829346
Numpy time      = 0.05197000503540039
Naive numpy time = 5.760146617889404

```

2 Arrays

2.0.1 Creating Arrays

In Numpy data is organized into arrays. There are many different ways to create a numpy array.

For the following we will use the random library of Numpy: <http://docs.scipy.org/doc/numpy-1.10.0/reference/routines.random.html>

Creating arrays from lists

```

[22]: #1-dimensional arrays
x = np.array([2,5,18,14,4])
print ("\n Deterministic 1-dimensional array \n")
print (x)

#2-dimensional arrays
x = np.array([[2,5,18,14,4], [12,15,1,2,8]])
print ("\n Deterministic 2-dimensional array \n")
print (x)

```

Deterministic 1-dimensional array

```
[ 2  5 18 14  4]
```

Deterministic 2-dimensional array

```
[[ 2  5 18 14  4]
 [12 15  1  2  8]]
```

We can also create Numpy arrays from Pandas DataFrames

```
[23]: d = {'A':[1., 2., 3., 4.],
          'B':[4., 3., 2., 1.]}
df = pd.DataFrame(d)
x = np.array(df)
print(x)
```

```
[[1. 4.]
 [2. 3.]
 [3. 2.]
 [4. 1.]]
```

Creating random arrays

```
[24]: #1-dimensional arrays
x = np.random.rand(5)
print ("\n Random 1-dimensional array \n")
print (x)

#2-dimensional arrays

x = np.random.rand(5,5)
print ("\n Random 5x5 2-dimensional array \n")
print (x)

x = np.random.randint(10,size=(2,3))
print("\n Random 2x3 array with integers")
print(x)
```

Random 1-dimensional array

```
[0.17892262 0.98767388 0.29423887 0.53986072 0.47561553]
```

Random 5x5 2-dimensional array

```
[[0.5306491  0.1662201  0.68729121 0.79026673 0.11813716]
 [0.4122728  0.09298485 0.43648845 0.04428801 0.37746864]
```

```
[0.51302356 0.1566327 0.31313685 0.95676673 0.09095682]
[0.65414023 0.5373658 0.55889802 0.78685662 0.43209138]
[0.75747506 0.05336422 0.32338388 0.80326922 0.55393489]]
```

Random 2x3 array with integers

```
[[1 3 9]
 [5 6 7]]
```

Transpose and get array dimensions

```
[25]: print("\n Matrix Dimensions \n")
print(x.shape)
print ("\n Transpose of the matrix \n")
print (x.T)
print (x.T.shape)
```

Matrix Dimensions

```
(2, 3)
```

Transpose of the matrix

```
[[1 5]
 [3 6]
 [9 7]]
(3, 2)
```

Special Arrays

```
[26]: x = np.zeros((4,4))
print ("\n 4x4 array with zeros \n")
print(x)

x = np.ones((4,4))
print ("\n 4x4 array with ones \n")
print (x)

x = np.eye(4)
print ("\n Identity matrix of size 4\n")
print(x)

x = np.diag([1,2,3])
print ("\n Diagonal matrix\n")
print(x)
```

4x4 array with zeros

```
[[0. 0. 0. 0.]
 [0. 0. 0. 0.]
 [0. 0. 0. 0.]
 [0. 0. 0. 0.]]
```

4x4 array with ones

```
[[1. 1. 1. 1.]
 [1. 1. 1. 1.]
 [1. 1. 1. 1.]
 [1. 1. 1. 1.]]
```

Identity matrix of size 4

```
[[1. 0. 0. 0.]
 [0. 1. 0. 0.]
 [0. 0. 1. 0.]
 [0. 0. 0. 1.]]
```

Diagonal matrix

```
[[1 0 0]
 [0 2 0]
 [0 0 3]]
```

```
[29]: A = np.random.randint(10,size=(2,3))
A
```

```
[29]: array([[6, 4, 4],
           [1, 8, 3]])
```

```
[32]: v = np.array([2,3])
D = np.diag(v)
print(D@A)
```

```
[[12  8  8]
 [ 3 24  9]]
```

2.0.2 Operations on arrays.

These are very similar to what we did with Pandas

```
[33]: x = np.random.randint(10, size = (2,4))
print (x)
print('\n mean value of all elements')
print (np.mean(x))
print('\n vector of mean values for columns')
print (np.mean(x,0)) #0 signifies the dimension meaning columns
```

```
print('\n vector of mean values for rows')
print (np.mean(x,1)) #1 signifies the dimension meaning rows
```

```
[[5 3 1 0]
 [3 1 4 8]]
```

```
mean value of all elements
3.125
```

```
vector of mean values for columns
[4.  2.  2.5 4. ]
```

```
vector of mean values for rows
[2.25 4.  ]
```

```
[34]: print('\n standard deviation of all elements')
print (np.std(x))
print('\n vector of std values for rows')
print (np.std(x,1)) #1 signifies the dimension meaning rows
print('\n median value of all elements')
print (np.median(x))
print('\n vector of median values for rows')
print (np.median(x,1))
print('\n sum of all elements')
print (np.sum(x))
print('\n vector of column sums')
print (np.sum(x,0))
print('\n product of all elements')
print (np.prod(x))
print('\n vector of row products')
print (np.prod(x,1))
```

```
standard deviation of all elements
2.4206145913796355
```

```
vector of std values for rows
[1.92028644 2.54950976]
```

```
median value of all elements
3.0
```

```
vector of median values for rows
[2.  3.5]
```

```
sum of all elements
25
```

```
vector of column sums  
[8 4 5 8]
```

```
product of all elements  
0
```

```
vector of row products  
[ 0 96]
```

2.0.3 Manipulating arrays

Accessing and Slicing

```
[35]: x = np.random.rand(4,3)  
print(x)  
print("\n element\n")  
print(x[1,2])  
print("\n row zero \n")  
print(x[0,:])  
print("\n column 2 \n")  
print(x[:,2])  
print("\n submatrix \n")  
print(x[1:3,0:2])  
print("\n entries > 0.5 \n")  
print(x[x>0.5])
```

```
[[0.36815974 0.64115793 0.36430617]  
 [0.88835153 0.8317612  0.95864869]  
 [0.55102392 0.07780769 0.69667718]  
 [0.39576696 0.09889422 0.8918579 ]]
```

element

```
0.9586486865600558
```

row zero

```
[0.36815974 0.64115793 0.36430617]
```

column 2

```
[0.36430617 0.95864869 0.69667718 0.8918579 ]
```

submatrix

```
[[0.88835153 0.8317612 ]  
 [0.55102392 0.07780769]]
```

```
entries > 0.5
```

```
[0.64115793 0.88835153 0.8317612  0.95864869 0.55102392 0.69667718  
0.8918579 ]
```

Changing entries

```
[36]: x = np.random.rand(4,3)  
print(x)  
  
x[1,2] = -5 #change an entry  
x[0:2,:] += 1 #change a set of rows  
x[2:4,1:3] = 0.5 #change a block  
print(x)  
  
print('\n Set entries > 0.5 to zero')  
x[x>0.5] = 0  
print(x)
```

```
[[0.49145468 0.84667038 0.20579891]  
 [0.00605387 0.64702729 0.19377095]  
 [0.35023522 0.24101048 0.23858619]  
 [0.24054741 0.67847084 0.30491831]]  
[[ 1.49145468  1.84667038  1.20579891]  
 [ 1.00605387  1.64702729 -4.         ]  
 [ 0.35023522  0.5         0.5        ]  
 [ 0.24054741  0.5         0.5        ]]
```

Set entries > 0.5 to zero

```
[[ 0.         0.         0.         ]  
 [ 0.         0.        -4.         ]  
 [ 0.35023522  0.5         0.5        ]  
 [ 0.24054741  0.5         0.5        ]]
```

```
[37]: print('\n Diagonal \n')  
x = np.random.rand(4,4)  
print(x)  
print('\n Read Diagonal \n')  
print(x.diagonal())  
print('\n Fill Diagonal with 1s \n')  
np.fill_diagonal(x,1)  
print(x)  
print('\n Fill Diagonal with vector \n')  
x[np.diag_indices_from(x)] = [1,2,3,4]  
print(x)
```

Diagonal


```
[0.33523301 0.29499204 0.36811354 0.97835507]
[0.51947023 0.74329564 0.31912965 0.73288542]
[0.75533618 0.86859457 0.52179594 0.13864346]
[0.13475995 0.46200594 0.32269882 0.27173642]]
```

Read Diagonal

```
[0.33523301 0.74329564 0.52179594 0.27173642]
```

Fill Diagonal with 1s

```
[[1.          0.29499204 0.36811354 0.97835507]
 [0.51947023 1.          0.31912965 0.73288542]
 [0.75533618 0.86859457 1.          0.13864346]
 [0.13475995 0.46200594 0.32269882 1.          ]]
```

Fill Diagonal with vector

```
[[1.          0.29499204 0.36811354 0.97835507]
 [0.51947023 2.          0.31912965 0.73288542]
 [0.75533618 0.86859457 3.          0.13864346]
 [0.13475995 0.46200594 0.32269882 4.          ]]
```

2.0.4 Quiz

We want to create a dataset of 10 users and 5 items, where each user i has selects an item j with probability 0.3.

How can we do this with matrix operations?

```
[43]: A = np.random.rand(10,5)
print(1-A)
A[A > 0.7] = 1
A[A < 0.7] = 0
print(A)
```

```
[[0.28824896 0.33268429 0.26909108 0.26352427 0.38341292]
 [0.34901154 0.9792799  0.03548684 0.07583914 0.24513767]
 [0.69225676 0.95232765 0.30882957 0.79114677 0.21338595]
 [0.4114604  0.47878534 0.57968324 0.02102232 0.69368738]
 [0.46424439 0.51146893 0.75088599 0.80881415 0.2536838 ]
 [0.59365372 0.30749492 0.79005323 0.35878169 0.46375286]
 [0.81813728 0.08004125 0.22605602 0.40383428 0.11326191]
 [0.18942507 0.29471383 0.06855792 0.93926606 0.68355041]
 [0.35269547 0.04931675 0.19131868 0.80425355 0.15989763]
 [0.88555675 0.37674654 0.30886494 0.75146318 0.45113655]]
[[1. 0. 1. 1. 0.]
 [0. 0. 1. 1. 1.]
 [0. 0. 0. 0. 1.]
```

```
[0. 0. 0. 1. 0.]
[0. 0. 0. 0. 1.]
[0. 0. 0. 0. 0.]
[0. 1. 1. 0. 1.]
[1. 1. 1. 0. 0.]
[0. 1. 1. 0. 1.]
[0. 0. 0. 0. 0.]
```

```
[4]: D = np.random.rand(10,5)
print(D)
D[D>=0.7] = 1
D[D< 0.7] = 0
#D[D <= 0.3] = 1
#D[D != 1] = 0
D
```

```
[[0.04063855 0.17115142 0.24436856 0.80370788 0.31885233]
 [0.39838706 0.08784638 0.66632184 0.85173152 0.3740655 ]
 [0.85515578 0.70191999 0.13256389 0.86430743 0.09837548]
 [0.26761904 0.80279186 0.95095196 0.39176937 0.22266263]
 [0.10887049 0.22309118 0.21955218 0.62408993 0.88571848]
 [0.48239737 0.01359243 0.41686108 0.66344021 0.24138179]
 [0.36301805 0.08943147 0.88049689 0.06076662 0.27845554]
 [0.5678915  0.79725246 0.57642486 0.21327003 0.82019054]
 [0.3566846  0.94917938 0.01890199 0.95145893 0.13594288]
 [0.4938302  0.26780825 0.20105711 0.52518667 0.1776111 ]]
```

```
[4]: array([[0., 0., 0., 1., 0.],
          [0., 0., 0., 1., 0.],
          [1., 1., 0., 1., 0.],
          [0., 1., 1., 0., 0.],
          [0., 0., 0., 0., 1.],
          [0., 0., 0., 0., 0.],
          [0., 0., 1., 0., 0.],
          [0., 1., 0., 0., 1.],
          [0., 1., 0., 1., 0.],
          [0., 0., 0., 0., 0.]])
```

2.0.5 Operations with Arrays

Multiplication and addition with scalar

```
[79]: x = np.random.rand(4,3)
print(x)

#multiplication and addition with scalar value
print("\n Matrix 2x+1 \n")
print(2*x+1)
```

```
[0.26655923 0.4976653 0.04086411]
[0.37654774 0.72029154 0.19286209]
[0.94842235 0.45388593 0.331548 ]
[0.93141818 0.46236218 0.67553373]]
```

Matrix 2x+1

```
[1.53311846 1.99533059 1.08172822]
[1.75309547 2.44058308 1.38572418]
[2.8968447 1.90777185 1.66309599]
[2.86283635 1.92472436 2.35106747]]
```

Vector-vector dot product

There are three ways to get the dot product of two vectors:

Using the method `.dot` of an array

Using the method `dot` of the numpy library

Using the '@' operator

```
[44]: y = np.array([2,-1,3])
z = np.array([-1,2,2])
print('\n y:',y)
print(' z:',z)
print('\n vector-vector dot product')
print(y.dot(z))
print(np.dot(y,z))
print(y@z)
```

```
y: [ 2 -1  3]
z: [-1  2  2]
```

vector-vector dot product

```
2
2
2
```

External product

The external product between two vectors x,y of size $(n,1)$ and $(m,1)$ results in a matrix M of size (n,m) with entries $M(i,j) = x(i)*y(j)$

```
[45]: print('\n y:',y)
print(' z:',z)
print('\n vector-vector external product')
print(np.outer(y,z))
```

```
y: [ 2 -1  3]
```

```
z: [-1  2  2]
```

```
vector-vector external product
```

```
[-2  4  4]  
[ 1 -2 -2]  
[-3  6  6]]
```

Element-wise operations

```
[46]: print('\n y:',y)  
print(' z:',z)  
print('\n element-wise addition')  
print(y+z)  
print('\n element-wise product')  
print(y*z)  
print('\n element-wise division')  
print(y/z)
```

```
y: [ 2 -1  3]  
z: [-1  2  2]
```

```
element-wise addition
```

```
[1 1 5]
```

```
element-wise product
```

```
[-2 -2  6]
```

```
element-wise division
```

```
[-2.  -0.5  1.5]
```

Matrix-Vector multiplication

Again we can do the multiplication either using the dot method or the '@' operator

```
[48]: X = np.random.randint(10, size = (4,3))  
print('Matrix X:\n',X)  
y = np.array([1,0,0])  
print("\n Matrix-vector right multiplication with",y,"\n")  
print(X.dot(y))  
print(np.dot(X,y))  
print(X@y)  
y = np.array([1,0,1,0])  
print("\n Matrix-vector left multiplication with",y,"\n")  
print(y.dot(X))  
print(np.dot(y,X))  
print(y@X)
```

Matrix X:

```
[[4 4 5]
```

```
[5 4 1]
[6 1 5]
[0 2 5]]
```

Matrix-vector right multiplication with [1 0 0]

```
[4 5 6 0]
[4 5 6 0]
[4 5 6 0]
```

Matrix-vector left multiplication with [1 0 1 0]

```
[10 5 10]
[10 5 10]
[10 5 10]
```

Matrix-Matrix multiplication

Same for the matrix-matrix operation

```
[49]: Y = np.random.randint(10, size=(3,2))
print("\n Matrix-matrix multiplication\n")
print('Matrix X:\n',X)
print('Matrix Y:\n',Y)
print('Product:\n',X.dot(Y))
print('Product:\n',X@Y)
```

Matrix-matrix multiplication

Matrix X:

```
[[4 4 5]
 [5 4 1]
 [6 1 5]
 [0 2 5]]
```

Matrix Y:

```
[[2 7]
 [0 5]
 [5 9]]
```

Product:

```
[[33 93]
 [15 64]
 [37 92]
 [25 55]]
```

Product:

```
[[33 93]
 [15 64]
 [37 92]
 [25 55]]
```

Matrix-Matrix element-wise operations

```
[50]: Z = np.random.randint(10, size=(3,2))+1
print('Matrix Y:\n',Y)
print('Matrix Z:\n',Z)
print("\n Matrix-matrix element-wise addition\n")
print(Y+Z)
print("\n Matrix-matrix element-wise multiplication\n")
print(Y*Z)
print("\n Matrix-matrix element-wise division\n")
print(Y/Z)
```

Matrix Y:

```
[[2 7]
 [0 5]
 [5 9]]
```

Matrix Z:

```
[[ 9  5]
 [ 7  9]
 [10  3]]
```

Matrix-matrix element-wise addition

```
[[11 12]
 [ 7 14]
 [15 12]]
```

Matrix-matrix element-wise multiplication

```
[[18 35]
 [ 0 45]
 [50 27]]
```

Matrix-matrix element-wise division

```
[[0.22222222 1.4      ]
 [0.         0.55555556]
 [0.5        3.        ]]
```

2.0.6 Creating Sparse Arrays

For sparse arrays we need to use the `sp_sparse` library from SciPy: <http://docs.scipy.org/doc/scipy/reference/sparse.html>

There are three types of sparse matrices:

csr: compressed row format

csc: compressed column format

lil: list of lists format

coo: coordinates format

These different types have to do with the matrix implementation and the data structures used.

The csr and csc formats are fast for arithmetic operations, but slow for slicing and incremental changes.

The lil format is fast for slicing and incremental construction, but slow for arithmetic operations.

The coo format does not support arithmetic operations and slicing, but it is very fast for constructing a matrix incrementally. You should then transform it to some other format for operations.

Creation of matrix from triplets

Triplets are of the form (row, column, value)

```
[53]: import scipy.sparse as sp_sparse

d = np.array([[0, 0, 12],
              [0, 1, 1],
              [0, 5, 34],
              [1, 3, 12],
              [1, 2, 6],
              [2, 0, 23],
              [3, 4, 14],
              ])

row = d[:,0]
col = d[:,1]
data = d[:,2]
# a matrix M with M[row[i],col[i]] = data[i] will be created
M = sp_sparse.csr_matrix((data,(row,col)), shape=(5,6))
print(M)
print(M.toarray()) #transforms back to full matrix
```

```
(0, 0)      12
(0, 1)       1
(0, 5)      34
(1, 2)       6
(1, 3)      12
(2, 0)      23
(3, 4)      14
[[12  1  0  0  0 34]
 [ 0  0  6 12  0  0]
 [23  0  0  0  0  0]
 [ 0  0  0  0 14  0]
 [ 0  0  0  0  0  0]]
```

Making a full matrix sparse

```
[54]: x = np.random.randint(2,size = (3,4))
print(x)
print('\n make x sparce')
A = sp_sparse.csr_matrix(x)
print(A)
```

```
[[1 1 1 1]
 [1 0 0 0]
 [1 1 0 1]]
```

```
make x sparce
(0, 0)      1
(0, 1)      1
(0, 2)      1
(0, 3)      1
(1, 0)      1
(2, 0)      1
(2, 1)      1
(2, 3)      1
```

Creating a sparse matrix incrementally

```
[55]: # Use lil (list of lists) representation, you can also use coo (coordinates)
A = sp_sparse.lil_matrix((10, 10))
A[0, :5] = np.random.randint(10,size = 5)
A[1, 5:10] = A[0, :5]
A.setdiag(np.random.randint(10,size = 10))
A[9,9] = 99
A[9,0]=1
print(A)
print(A.toarray())
print(A.diagonal())
A = A.tocsr() # makes it a compressed column format. better for dot product.
B = A.dot(np.ones(10))
print(B)
```

```
(0, 0)      5.0
(0, 1)      4.0
(0, 2)      9.0
(0, 3)      3.0
(0, 4)      6.0
(1, 1)      4.0
(1, 5)      1.0
(1, 6)      4.0
(1, 7)      9.0
(1, 8)      3.0
(1, 9)      6.0
(2, 2)      1.0
```



```

(3, 3)      3.0
(5, 5)      6.0
(6, 6)      9.0
(7, 7)      7.0
(8, 8)      3.0
(9, 0)      1.0
(9, 9)      99.0
[[ 5.  4.  9.  3.  6.  0.  0.  0.  0.  0.]
 [ 0.  4.  0.  0.  0.  1.  4.  9.  3.  6.]
 [ 0.  0.  1.  0.  0.  0.  0.  0.  0.  0.]
 [ 0.  0.  0.  3.  0.  0.  0.  0.  0.  0.]
 [ 0.  0.  0.  0.  0.  0.  0.  0.  0.  0.]
 [ 0.  0.  0.  0.  0.  6.  0.  0.  0.  0.]
 [ 0.  0.  0.  0.  0.  0.  9.  0.  0.  0.]
 [ 0.  0.  0.  0.  0.  0.  0.  7.  0.  0.]
 [ 0.  0.  0.  0.  0.  0.  0.  0.  3.  0.]
 [ 1.  0.  0.  0.  0.  0.  0.  0.  0.  99.]]
[ 5.  4.  1.  3.  0.  6.  9.  7.  3.  99.]
[ 27. 27.  1.  3.  0.  6.  9.  7.  3. 100.]

```

All operations work like before

```
[56]: print(A.dot(A.T))
```

```

(0, 3)      9.0
(0, 2)      9.0
(0, 1)     16.0
(0, 9)      5.0
(0, 0)    167.0
(1, 9)    594.0
(1, 8)      9.0
(1, 7)     63.0
(1, 6)     36.0
(1, 5)      6.0
(1, 1)    159.0
(1, 0)     16.0
(2, 2)      1.0
(2, 0)      9.0
(3, 3)      9.0
(3, 0)      9.0
(5, 5)     36.0
(5, 1)      6.0
(6, 6)     81.0
(6, 1)     36.0
(7, 7)     49.0
(7, 1)     63.0
(8, 8)      9.0
(8, 1)      9.0

```

```
(9, 1)      594.0
(9, 9)      9802.0
(9, 0)      5.0
```

```
[59]: A[0].mean()
```

```
[59]: 2.7
```

2.1 Singular Value Decomposition

For the singular value decomposition we will use the libraries from Numpy and SciPy and SciKit Learn

We use sklearn to create a low-rank matrix (https://scikit-learn.org/stable/modules/generated/sklearn.datasets.make_low_rank_matrix.html). We will create a matrix with *effective rank* 2.

```
[60]: import sklearn.datasets as sk_data

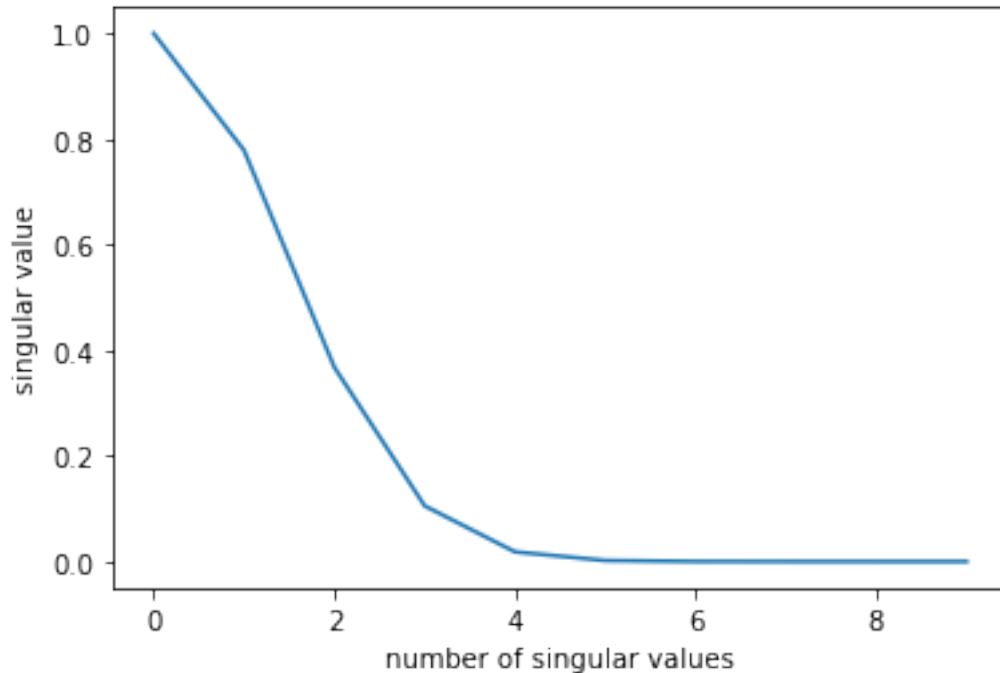
data = sk_data.make_low_rank_matrix(n_samples=100, n_features=50,
    ↪effective_rank=2, tail_strength=0.0, random_state=None)
#sns.heatmap(data, xticklabels=False, yticklabels=False, linewidths=0)
```

We will use the `numpy.linalg.svd` function to compute the Singular Value Decomposition of the matrix we created (<http://docs.scipy.org/doc/numpy/reference/generated/numpy.linalg.svd.html>).

```
[69]: U, s, V = np.linalg.svd(data, full_matrices = False)
print (U.shape, s.shape, V.shape)
print(s)
plt.plot(s[0:10])
plt.ylabel('singular value')
plt.xlabel('number of singular values')
```

```
(100, 50) (50,) (50, 50)
[1.00000000e+00 7.78800783e-01 3.67879441e-01 1.05399225e-01
 1.83156389e-02 1.93045414e-03 1.23409804e-04 4.78511739e-06
 1.12535175e-07 1.60522805e-09 1.38879490e-11 7.28775676e-14
 2.34683864e-16 8.93169829e-17 8.93169829e-17 8.93169829e-17
 8.93169829e-17 8.93169829e-17 8.93169829e-17 8.93169829e-17
 8.93169829e-17 8.93169829e-17 8.93169829e-17 8.93169829e-17
 8.93169829e-17 8.93169829e-17 8.93169829e-17 8.93169829e-17
 8.93169829e-17 8.93169829e-17 8.93169829e-17 8.93169829e-17
 8.93169829e-17 8.93169829e-17 8.93169829e-17 8.93169829e-17
 8.93169829e-17 8.93169829e-17 8.93169829e-17 8.93169829e-17
 8.93169829e-17 8.93169829e-17 8.93169829e-17 8.93169829e-17
 8.93169829e-17 2.16246651e-17]
```

```
[69]: Text(0.5, 0, 'number of singular values')
```



We can also use the `scipy.sparse.linalg` library to compute the SVD for sparse matrices (<http://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.sparse.linalg.svds.html>)

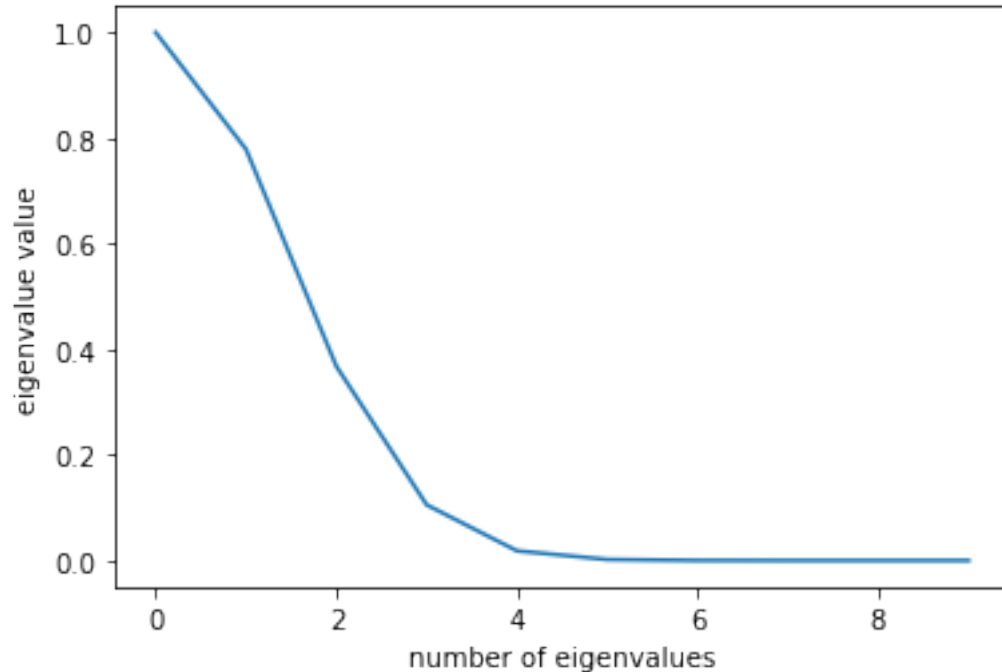
We need to specify the number of components, otherwise it is by default $k = 6$. The singular values are in increasing order.

```
[70]: import scipy.sparse.linalg as sp_linalg

data2 = sp_sparse.csc_matrix(data)
print(data2.shape)
U,s,V = sp_linalg.svds(data2, k = 10) #by default returns k=6 singular values
print (U.shape, s.shape, V.shape)
print(s)
plt.plot(s[::-1])
plt.ylabel('eigenvalue value')
plt.xlabel('number of eigenvalues')
```

```
(100, 50)
(100, 10) (10,) (10, 50)
[0.00000000e+00 0.00000000e+00 0.00000000e+00 1.23409804e-04
 1.93045414e-03 1.83156389e-02 1.05399225e-01 3.67879441e-01
 7.78800783e-01 1.00000000e+00]
```

```
[70]: Text(0.5, 0, 'number of eigenvalues')
```



We can also compute SVD using the library of SciKit Learn (<https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.TruncatedSVD.html>)

```
[71]: from sklearn.decomposition import TruncatedSVD

K = 10
svd = TruncatedSVD(n_components=K)
svd.fit(data2)
print(svd.components_.shape) # the V vectors
print(svd.transform(data2).shape) # the U vectors
print(svd.singular_values_)
```

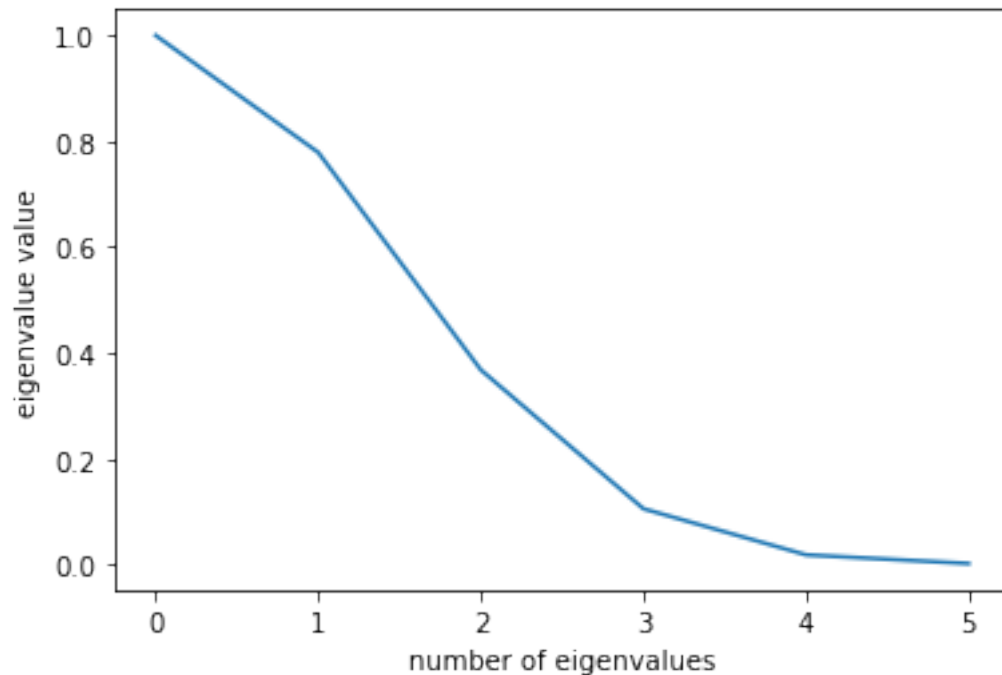
```
(10, 50)
(100, 10)
[1.00000000e+00 7.78800783e-01 3.67879441e-01 1.05399225e-01
 1.83156389e-02 1.93045414e-03 1.23409804e-04 4.78511739e-06
 1.12535175e-07 1.60522805e-09]
```

2.1.1 Obtaining a low rank approximation of the data

To obtain a rank-k approximation of the matrix we multiply the k first columns of U, with the diagonal matrix with the k first (largest) singular values, with the matrix with the first k rows of V transpose

```
[73]: K = 6
U_k,s_k,V_k = sp_linalg.svds(data2, K, which = 'LM')
print (U_k.shape, s_k.shape, V_k.shape)
print(s_k)
plt.plot(s_k[::-1])
plt.ylabel('eigenvalue value')
plt.xlabel('number of eigenvalues')
S_k = np.diag(s_k)
```

```
(100, 6) (6,) (6, 50)
[0.00193045 0.01831564 0.10539922 0.36787944 0.77880078 1.      ]
```

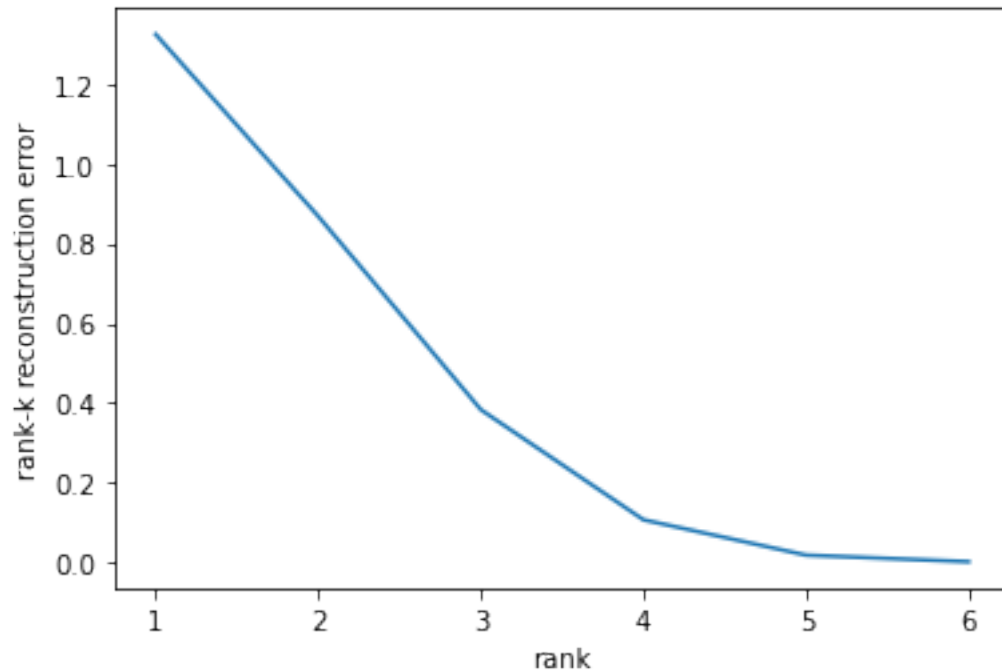


```
[75]: reconstruction_error = []
for k in range(K,0,-1):
    data_k = U_k[:,k:] .dot(S_k[k:,k:]) .dot(V_k[k:,:])
    error = np.linalg.norm(data_k-data2,ord='fro')
    reconstruction_error.append(error)
    print(error)
data_k = U_k.dot(S_k) .dot(V_k)
print(np.linalg.norm(data_k-data2,ord='fro'))

plt.plot(1+np.array(range(6)),reconstruction_error)
plt.ylabel('rank-k reconstruction error')
plt.xlabel('rank')
```

```
1.3241276917357527
0.8679367165994607
0.3831233278055767
0.10699626662742327
0.018417506182009293
0.0019344006983659258
0.00012350259009393685
```

```
[75]: Text(0.5, 0, 'rank')
```



2.2 An example

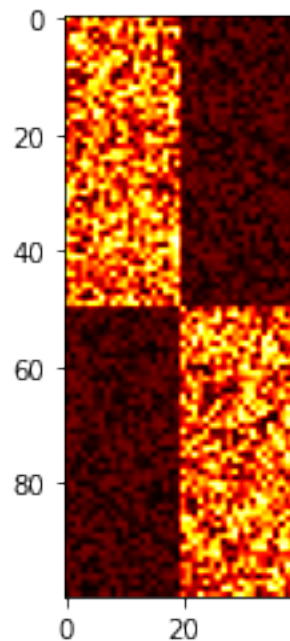
We will create a block diagonal matrix, with blocks of different “intensity” of values

```
[76]: import numpy as np

M1 = np.random.randint(1,50,(50,20))
M2 = np.random.randint(1,10,(50,20))
M3 = np.random.randint(1,10,(50,20))
M4 = np.random.randint(1,50,(50,20))

T = np.concatenate((M1,M2),axis=1)
B = np.concatenate((M3,M4),axis=1)
M = np.concatenate([T,B],axis = 0)
```

```
[77]: plt.imshow(M, cmap='hot')
plt.show()
```



We observe that there is a correlation between the column and row sums and the left and right singular vectors

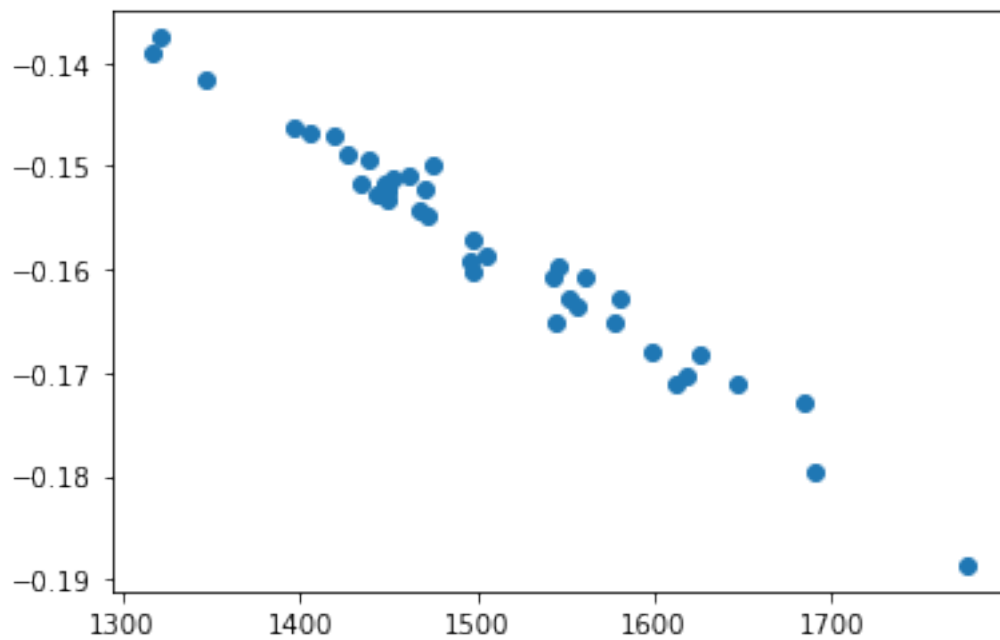
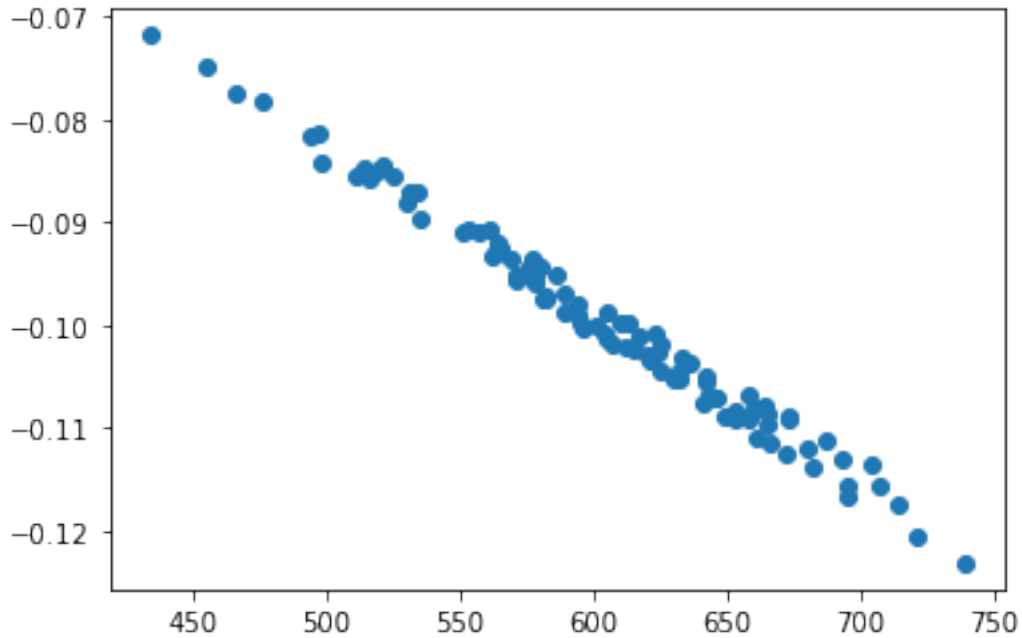
Note: The values of the vectors are negative. We would get the same result if we make them positive.

```
[78]: import scipy.stats as stats
import matplotlib.pyplot as plt

(U,S,V) = np.linalg.svd(M,full_matrices = False)
#print(S)
c = M.sum(0)
r = M.sum(1)
print(stats.pearsonr(r,U[:,0]))
print(stats.pearsonr(c,V[0]))
plt.scatter(r,U[:,0])
plt.figure()
plt.scatter(c,V[0])
```

```
(-0.9934144337312895, 5.00468040382164e-94)
(-0.9850767934010615, 1.1927248146327699e-30)
```

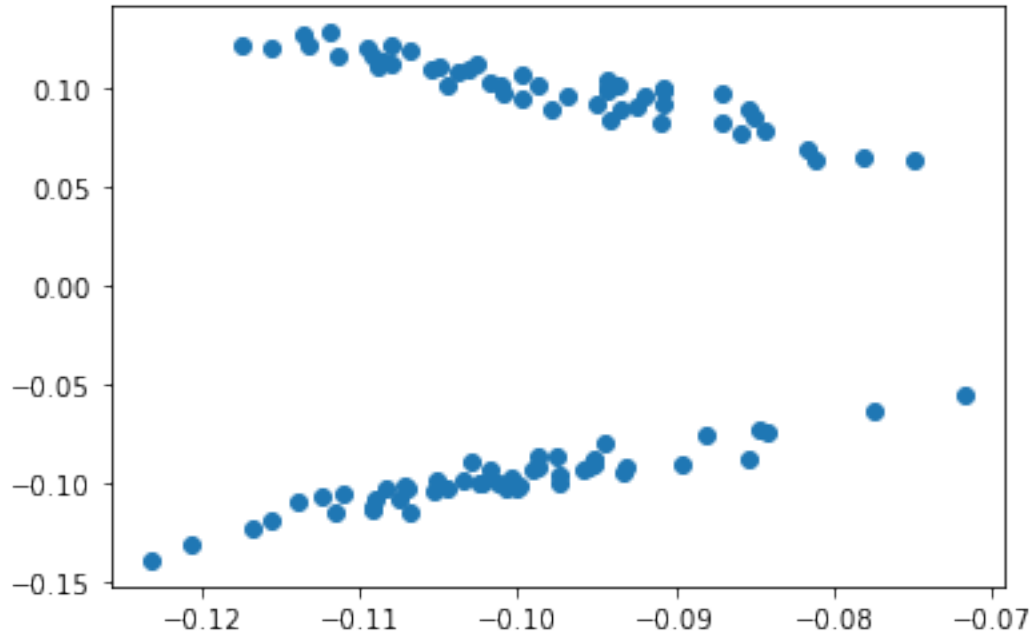
```
[78]: <matplotlib.collections.PathCollection at 0x16e3381eca0>
```



Using the first two singular vectors we can clearly differentiate the two blocks of rows

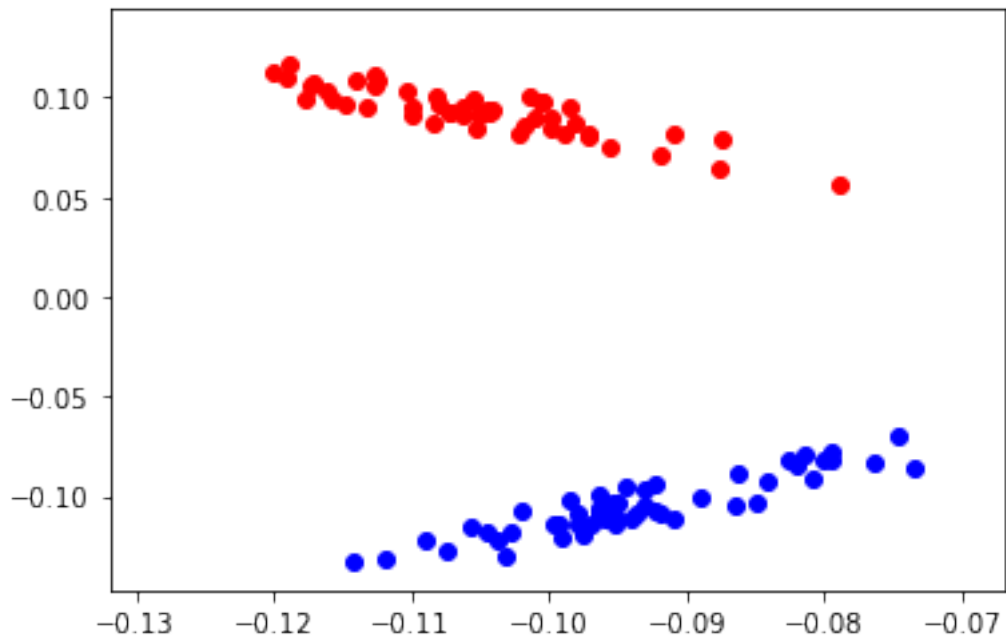
```
[79]: plt.scatter(U[:,0],U[:,1])
```

```
[79]: <matplotlib.collections.PathCollection at 0x16e337c0790>
```

```
[163]: plt.scatter(x = U[:50,0],y = U[:50,1],color='r')
plt.scatter(x = U[50:,0],y = U[50:,1], color = 'b')
```

```
[163]: <matplotlib.collections.PathCollection at 0x147065aeeb8>
```



2.3 PCA using SciKit Learn

We will now use the PCA package from the SciKit Learn (sklearn) library. PCA is the same as SVD but now the matrix is centered: the mean is removed from the columns of the matrix.

You can read more here: <https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html>

```
[80]: from sklearn.decomposition import PCA
      pca = PCA(n_components=2)
      pca.fit(M)
```

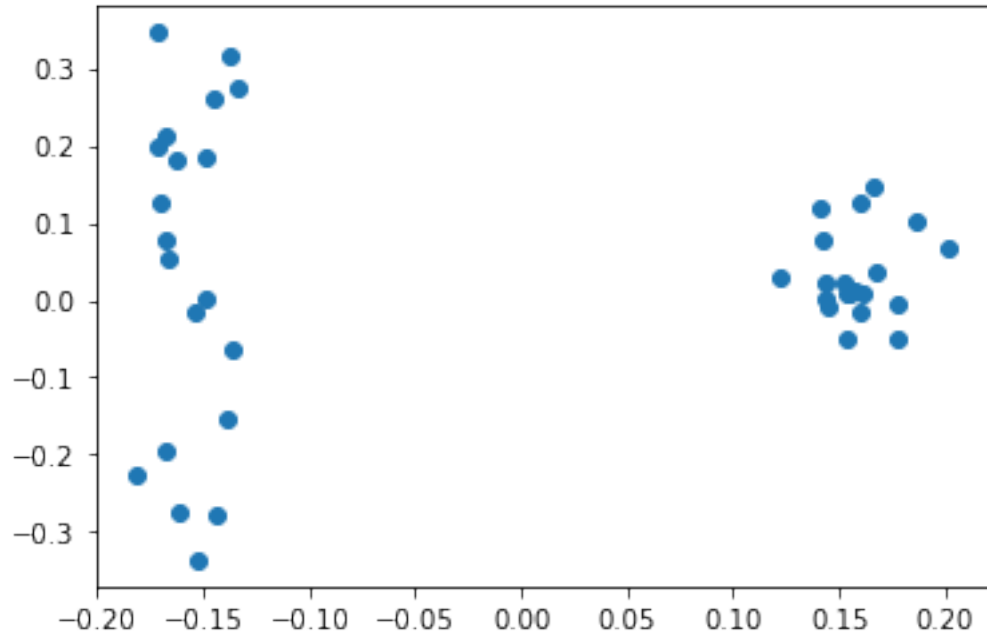
```
[80]: PCA(n_components=2)
```

```
[82]: pca.components_
```

```
[82]: array([[ 0.14124327,  0.15722362,  0.12228396,  0.15367817,  0.15942344,
            0.17796923,  0.16796    ,  0.14311024,  0.16665599,  0.17718799,
            0.15324126,  0.14189726,  0.15979208,  0.1522675 ,  0.20173536,
            0.15536385,  0.14302179,  0.14447337,  0.16088513,  0.18603015,
           -0.16092718, -0.14470302, -0.13857141, -0.16307114, -0.13340357,
           -0.13714955, -0.17123051, -0.13569162, -0.14348402, -0.15397093,
           -0.14895391, -0.18189587, -0.1676614 , -0.14898163, -0.17193101,
           -0.1671171 , -0.15245651, -0.16778115, -0.17032356, -0.16758541],
          [ 0.11860349,  0.0120605 ,  0.02996427,  0.00920918, -0.01600018,
           -0.00535662,  0.03517382,  0.02117632,  0.14693107, -0.04985701,
           -0.05011393,  0.07806108,  0.12484136,  0.02149389,  0.06889166,
            0.00975357,  0.00113441, -0.00772076,  0.00786862,  0.10276518,
           -0.27367207,  0.26090065, -0.15506362,  0.17944452,  0.27281674,
            0.31594807,  0.34593268, -0.06270117, -0.2764924 , -0.01472123,
            0.00311597, -0.22452511,  0.07605783,  0.18344832,  0.19952305,
            0.05277773, -0.33610051,  0.21305541,  0.12678607, -0.19485843]])
```

```
[83]: plt.scatter(pca.components_[0],pca.components_[1])
```

```
[83]: <matplotlib.collections.PathCollection at 0x16e33a2c790>
```



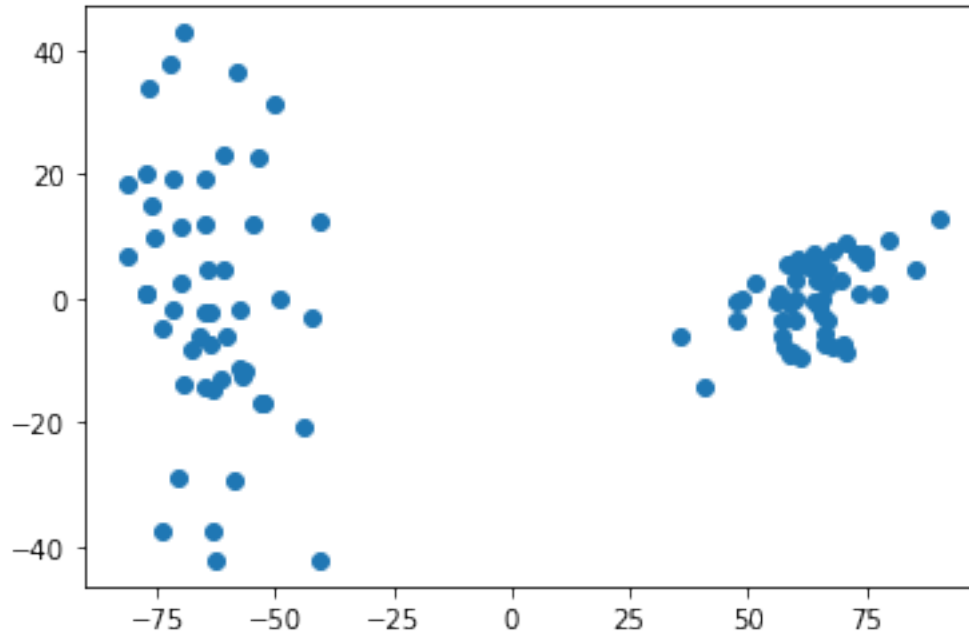
Using the operation transform we can transform the data directly to the lower-dimensional space

```
[89]: MPCA = pca.transform(M)  
      print(MPCA.shape)
```

```
(100, 2)
```

```
[90]: plt.scatter(MPCA[:,0],MPCA[:,1])
```

```
[90]: <matplotlib.collections.PathCollection at 0x16e36a761f0>
```



We will now experiment with a well-known dataset of data analysis, the iris dataset:

https://scikit-learn.org/stable/auto_examples/datasets/plot_iris_dataset.html

```
[91]: from sklearn import datasets
iris = datasets.load_iris()
X = iris.data
y = iris.target #contains the labels of the data

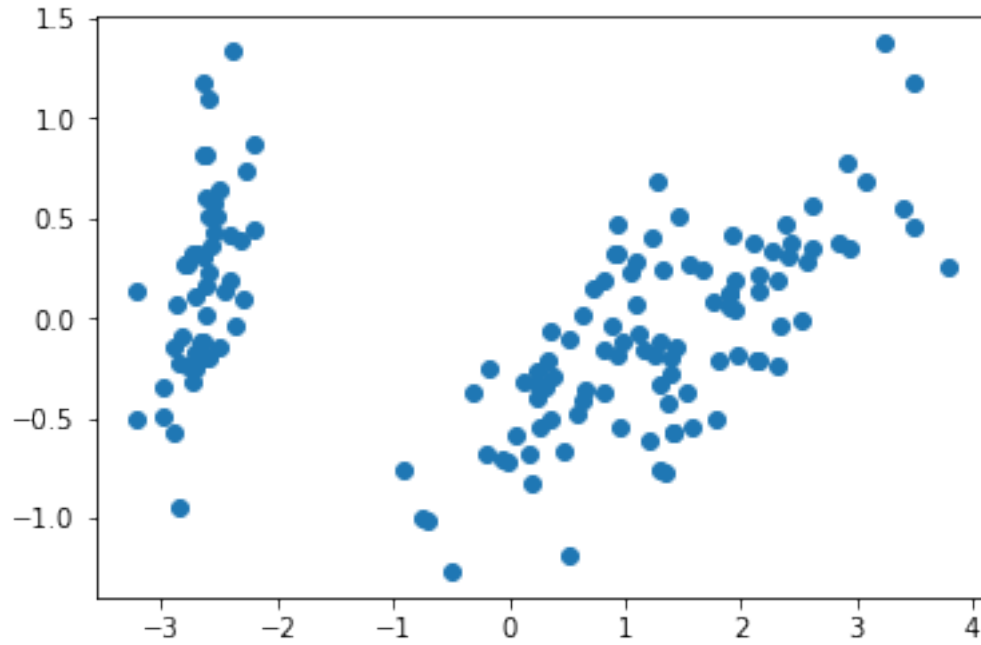
pca = PCA(n_components=3)
pca.fit(X)
X = pca.transform(X)
```

```
[92]: pca.explained_variance_
```

```
[92]: array([4.22824171, 0.24267075, 0.0782095 ])
```

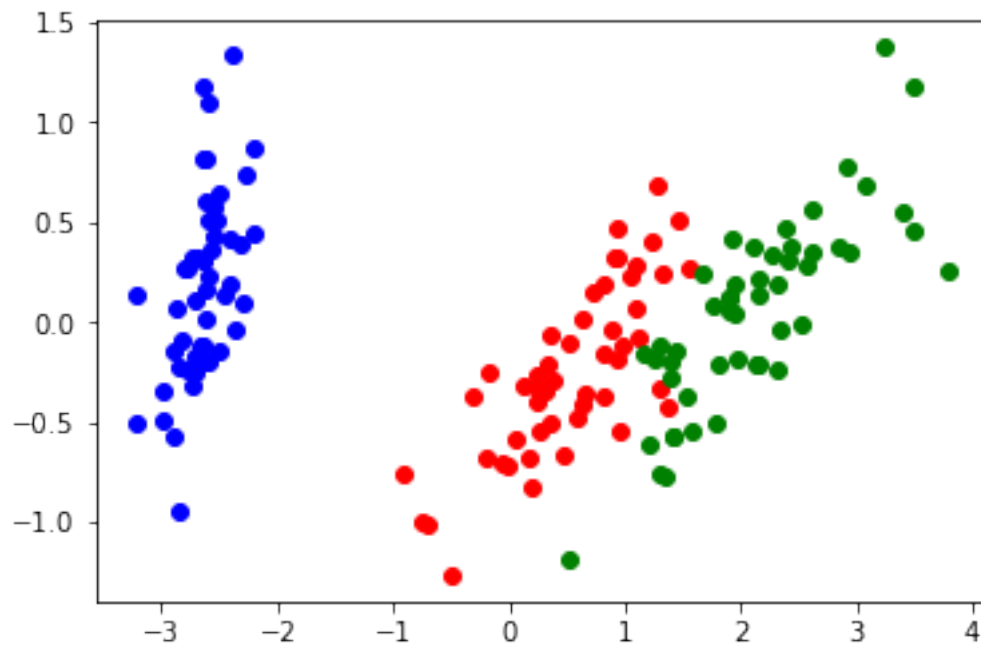
```
[93]: plt.scatter(X[:,0],X[:,1])
```

```
[93]: <matplotlib.collections.PathCollection at 0x16e36ac79d0>
```



```
[94]: plt.scatter(X[y==0,0],X[y==0,1], color='b')  
plt.scatter(X[y==1,0],X[y==1,1], color='r')  
plt.scatter(X[y==2,0],X[y==2,1], color='g')
```

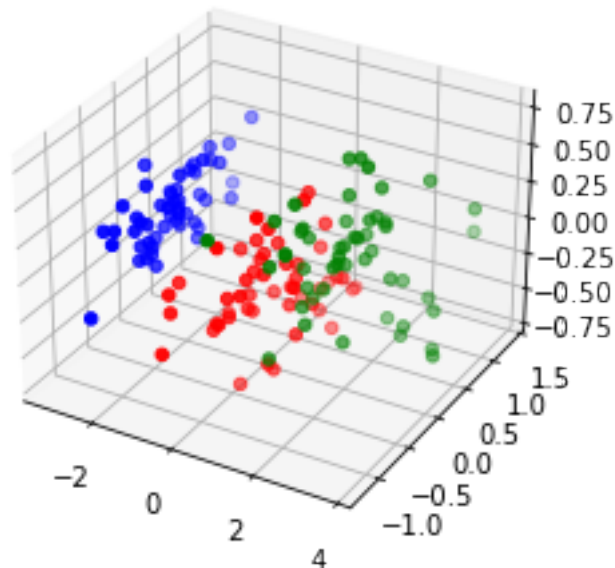
[94]: <matplotlib.collections.PathCollection at 0x16e36b26d90>



```
[95]: from mpl_toolkits.mplot3d import Axes3D
fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')

ax.scatter(X[y==0,0],X[y==0,1], X[y==0,2], color='b')
ax.scatter(X[y==1,0],X[y==1,1], X[y==1,2], color='r')
ax.scatter(X[y==2,0],X[y==2,1], X[y==2,2], color='g')
```

```
[95]: <mpl_toolkits.mplot3d.art3d.Path3DCollection at 0x16e36b9a8e0>
```



```
[96]: from sklearn.datasets import fetch_20newsgroups

categories = ['comp.os.ms-windows.misc', 'sci.space', 'rec.sport.baseball']
news_data = fetch_20newsgroups(subset='train', categories=categories)
```

```
[97]: from sklearn.feature_extraction.text import TfidfVectorizer

vectorizer = TfidfVectorizer(stop_words='english', min_df=4,max_df=0.8)
dtm = vectorizer.fit_transform(news_data.data)
```

```
[ ]: import nltk
nltk.download('punkt')
nltk.download('stopwords')
from nltk.stem.snowball import SnowballStemmer
from nltk.tokenize import word_tokenize, sent_tokenize
```

```

stemmed_data = [" ".join(SnowballStemmer("english", ignore_stopwords=True).
    ↪stem(word)
        for sent in sent_tokenize(message)
        for word in word_tokenize(sent))
    for message in news_data.data]

# stemmed_data = news_data.data

```

```

[15]: dtm = vectorizer.fit_transform(stemmed_data)
      terms = vectorizer.get_feature_names()
      print(terms)

```

```

['00', '000', '0005', '0062', '0096b0f0', '00bjgood', '00mbstultz', '01',
'0114', '01wb', '02', '02138', '023017', '024222', '03', '033', '0362', '03i',
'03k', '04', '0400', '05', '051746', '06', '0688', '07', '08', '086', '09',
'091051', '093231', '0_', '0a', '0b', '0c', '0d', '0e', '0ek', '0f', '0g', '0h',
'0hd', '0i', '0iv', '0j', '0k', '0l', '0m', '0m75u', '0n', '0o', '0p', '0q',
'0qax', '0qq', '0r', '0s', '0t', '0tbxn', '0tbxom', '0tq', '0tq6', '0u', '0v',
'0va', '0w', '0x', '0y', '0z', '10', '100', '1000', '100m', '101', '101044',
'102', '1024x768', '1024x768x256', '103', '104', '105', '106', '107', '108',
'109', '11', '110', '1100', '111', '11181', '112', '113', '114', '115',
'115313', '117', '118', '119', '11th', '12', '120', '1200', '120399', '121',
'122', '123', '124', '125', '126', '127', '128', '1280x1024', '129', '13',
'130', '131', '132', '133', '134', '135', '136', '137', '138', '139', '13p',
'13q', '13qs', '13s', '14', '140', '1400', '141', '141824', '142', '143', '144',
'14496', '145', '145s', '146', '14691229', '147', '148', '14853', '149', '14di',
'15', '150', '1500', '151', '152', '153', '154', '155', '1550', '158',
'15apr93', '15th', '16', '160', '1607', '161', '162', '164', '165', '16550',
'167', '168730', '16m', '16mb', '17', '170', '170mb', '171', '172', '17301',
'175', '176', '177', '17986', '18', '180', '18084tm', '182', '185', '186',
'187', '188', '189', '19', '190', '192', '1933', '194', '195', '1950', '1957',
'1958', '195853', '1960', '1960s', '1961', '1962', '1963', '1964', '1965',
'1968', '1969', '197', '1970', '1971', '1972', '1973', '1975', '1976', '1977',
'1979', '198', '1980', '1981', '1982', '1983', '1984', '1985', '1986', '1987',
'1988', '1989', '199', '1990', '1991', '1992', '1993', '1993apr06', '1993apr12',
'1993apr13', '1993apr14', '1993apr15', '1993apr16', '1993apr17', '1993apr18',
'1993apr19', '1993apr2', '1993apr20', '1993apr21', '1993apr22', '1993apr23',
'1993apr25', '1993apr26', '1993apr3', '1993apr5', '1993apr6', '1993e', '1994',
'1995', '1996', '1999', '19th', '1_', '1a', '1b', '1c', '1d', '1d2', '1d3hz',
'1d9', '1d9l', '1e', '1eq', '1eqtct', '1eqtm', '1f', '1f9', '1f9f8', '1f9f9f',
'1f9l3', '1fp', '1fp4', '1fp4u', '1fpl', '1g', '1h', '1i', '1j', '1k', '1l',
'1m', '1mb', '1n', '1o', '1oy', '1p', '1q', '1qi', '1qkkodinn5f5', '1qr',
'1qu8ud', '1qx', '1r', '1r46o9inn14j', '1r6aqr', '1ri', '1s', '1st', '1t',
'1t7', '1u', '1v', '1w', '1x', '1xl', '1xu', '1y', '1z', '1z4', '1z5', '1z6e',
'1z6eit', '1z6ei', '1z6ei0l', '20', '200', '2000', '2001', '2005', '201',

```

'2010', '202', '2048', '204845', '205', '208', '209', '20mb', '20th', '21',
'210', '211638', '212202', '213', '2130', '214', '216', '2178', '219', '22',
'220', '221', '222', '223', '224', '225', '226', '227', '2291', '22di', '23',
'230', '231', '232', '233', '234', '235', '23536', '235430', '236', '237',
'238', '24', '240', '2400', '241', '243', '2436', '244', '245', '246', '249',
'24939', '24e', '24g', '24th', '24x', '24xx', '25', '250', '252', '253', '254',
'255', '256', '257', '258', '259', '25k', '26', '260', '261', '262', '264',
'265', '267', '268', '2697', '27', '270', '272', '273', '274', '275', '276',
'277', '278', '28', '280', '282', '284', '285', '286', '287', '288', '28i',
'28th', '29', '290', '292', '293', '294', '295', '296', '297', '29848', '299',
'2_', '2a', '2a42dubinski', '2b', '2c', '2d', '2d1', '2de145', '2di', '2dm75u',
'2e', '2f', '2g', '2gt', '2h', '2hd', '2i', '2j', '2k', '2l', '2lk', '2m',
'2mb', '2n', '2nd', '2o', '2p', '2pl', '2pm', '2pu', '2q', '2r', '2s', '2t',
'2tct', '2tcv9', '2tcv9f0', '2tg', '2tm', '2u', '2ud', '2v', '2w', '2x', '2y',
'2z', '30', '300', '3000', '3006', '301', '303', '30332', '304', '306', '308',
'30th', '31', '310', '312', '313', '314', '315', '316', '318', '32', '320',
'321', '324', '325', '327', '328', '32bit', '32m', '33', '330', '3300', '333',
'334', '335', '336', '337', '339', '33mhz', '34', '343', '344', '345', '348',
'34deg', '34l', '34r', '34u', '35', '350', '351', '3511', '3539', '354', '355',
'3556', '356', '35611', '3578', '359', '35i', '35t', '36', '360', '361', '3610',
'363', '364', '365', '366', '367', '368', '3684', '369', '37', '370', '373',
'374', '375', '378', '38', '380', '381', '384', '386', '386dx', '386sx', '389',
'39', '390', '391', '394', '3a', '3b', '3c', '3d', '3di', '3dy', '3e', '3f',
'3g', '3h', '3hz', '3i', '3is', '3j', '3k', '3l', '3m', '3n', '3o', '3p', '3q',
'3qs', '3r', '3r6', '3rd', '3s', '3t', '3tc', '3u', '3v', '3v9', '3v9f0',
'3v9f9d', '3v9f9f9', '3v9f9f9f9f9', '3w', '3w2', '3w2tg', '3w2tm', '3x', '3z',
'40', '400', '4000', '4001', '404', '405', '406', '407', '408', '41', '410',
'411', '412', '415', '416', '42', '420', '422', '423', '424', '427', '428',
'43', '430', '43011', '433', '435', '4368', '439', '44', '440', '442', '443',
'444', '448', '45', '450', '453', '4544', '455', '458', '45u', '46', '460',
'462', '464', '4650', '4656', '467', '469', '47', '471', '475', '476', '48',
'480', '4800', '481', '482', '483', '486', '486dx', '487', '489', '49', '490',
'4916384', '492', '498', '4_', '4a', '4b', '4c', '4d', '4e', '4f', '4g', '4h',
'4i', '4j', '4k', '4l', '4m', '4mb', '4mn', '4n', '4o', '4p', '4q', '4r', '4s',
'4t', '4tb', '4tbxom', '4th', '4u', '4v', '4w', '4wa', '4x', '4y', '4z', '50',
'500', '5000', '500c', '503', '507', '508', '51', '510', '512', '514', '5145',
'5148', '515', '517', '518', '51s', '52', '520', '525', '53', '530', '532',
'534', '54', '542', '545', '546', '549', '54909', '54e', '55', '550', '552',
'556', '56', '560', '5600', '562', '567', '57', '5700', '571', '5744', '575',
'577', '578', '58', '580', '583', '586', '588', '59', '590', '594', '595',
'596', '5_', '5a', '5at', '5c', '5d', '5e', '5e8', '5e9', '5f', '5g', '5g9',
'5g9p', '5g9v', '5h', '5i', '5j', '5k', '5l', '5m', '5mb', '5o', '5p', '5q',
'5s', '5t', '5th', '5u', '5vz', '5w', '5x', '5z', '60', '600', '6000', '601',
'602', '603', '604', '605', '606', '6097', '60s', '61', '610', '612', '614',
'615', '616', '617', '619', '61d', '61d9', '62', '620', '621', '625', '628',
'629', '63', '630', '631', '632', '635', '636', '637', '638', '639', '63q',
'64', '640x480', '641', '643', '644', '646', '647', '649', '64k', '65', '650',
'653', '655', '656', '657', '65k', '66', '661', '664', '666', '667', '66mhz',

'afg', 'afit', 'afraid', 'africa', 'afternoon', 'afterward', 'ag', 'agate',
'agc', 'age', 'agenc', 'agenda', 'agent', 'aggrav', 'agian', 'ago', 'agre',
'agreement', 'agricultur', 'ah', 'ahead', 'ahf', 'ahh', 'ahl', 'ai', 'aiaa',
'aid', 'aim', 'aio', 'air', 'aircraft', 'airfram', 'airlg', 'airlin', 'airplan',
'airport', 'airwav', 'aix', 'aiyar', 'aj', 'ajr', 'ajz', 'ak', 'ak333', 'al',
'ala', 'alabama', 'alan', 'alaska', 'alavi', 'albany', 'albeit', 'albert',
'alberta', 'albnyvms', 'album', 'alcohol', 'ale', 'alec', 'alex', 'alexand',
'algorithm', 'alias', 'alicea', 'alien', 'align', 'alird', 'aliv', 'alleg',
'allegheni', 'allen', 'alloc', 'allow', 'almaden', 'alomar', 'alon', 'alot',
'alpha', 'alphabet', 'alphacdc', 'alreadi', 'alt', 'alter', 'altern', 'altitud',
'aluminum', 'alway', 'amateur', 'amaz', 'ambiti', 'amd', 'ame', 'amend',
'america', 'american', 'ames', 'amherst', 'ami', 'amipro', 'amsat', 'amsterdam',
'amv', 'analog', 'analys', 'analysisi', 'analyst', 'analyz', 'ancient',
'andersen', 'anderson', 'andi', 'andr', 'andresen', 'andrew', 'andy', 'angel',
'anger', 'angl', 'angular', 'anim', 'annes', 'announc', 'annoy', 'annual',
'anomali', 'anonym', 'anoth', 'answer', 'antarct', 'antenna', 'anthoni', 'anti',
'anybodi', 'anymor', 'anyon', 'anyth', 'anytim', 'anywher', 'ao', 'ap', 'apana',
'apart', 'apertur', 'api', 'apollo', 'apolog', 'app', 'appar', 'appeal',
'appear', 'appl', 'apple', 'appli', 'applic', 'appoint', 'appreci', 'approach',
'appropri', 'approv', 'approx', 'approxim', 'apr', 'april', 'apt', 'aq', 'ar',
'arbitr', 'arbitrari', 'arc', 'archi', 'architectur', 'archiv', 'archive',
'arden', 'ardua', 'area', 'aren', 'argu', 'argument', 'arian', 'arianespac',
'ariel', 'aris', 'arizona', 'arizvm', 'arlington', 'arm', 'armi', 'arnold',
'arrang', 'array', 'arriv', 'arrow', 'art', 'arthur', 'artic', 'articl',
'article', 'artifici', 'artin', 'ascend', 'ascii', 'asd', 'ashok', 'asid',
'ask', 'aspect', 'ass', 'assembl', 'assess', 'asset', 'asshol', 'assign',
'assist', 'associ', 'assum', 'assumpt', 'assur', 'asteroid', 'astor', 'astra',
'astro', 'astronaut', 'astronom', 'astronomi', 'astrophys', 'atc', 'ate',
'athen', 'athena', 'athlet', 'ati', 'atl', 'atlant', 'atlanta', 'atlas',
'atleast', 'atm', 'atmanag', 'atmospher', 'atom', 'att', 'attach', 'attack',
'attempt', 'attend', 'attent', 'attitud', 'attract', 'attribut', 'au', 'auburn',
'audienc', 'audio', 'aug', 'august', 'aurora', 'auspic', 'austin', 'australia',
'author', 'auto', 'autoexec', 'autom', 'automat', 'av', 'avail', 'ave', 'avenu',
'averag', 'avg', 'aviat', 'avoid', 'aw', 'awar', 'award', 'away', 'awesom',
'aweto', 'ax', 'ax2', 'axa12', 'axe', 'axi', 'ay', 'az', 'b0', 'b1', 'b2', 'b3',
'b4', 'b4q', 'b5', 'b6', 'b7', 'b8', 'b8e', 'b8f', 'b8g', 'b9', 'b9r', 'b_',
'ba', 'baalk', 'babe', 'babi', 'babylon', 'bach', 'backer', 'background',
'backup', 'bacteria', 'bad', 'baerga', 'baeyer', 'bag', 'bagwel', 'bailey',
'baker', 'balanc', 'bald', 'ball', 'ballist', 'balloon', 'ballpark', 'ballplay',
'ballyard', 'baltimor', 'ban', 'band', 'bandwidth', 'bang', 'bangl', 'bank',
'banner', 'bar', 'barberi', 'barman', 'barn', 'barrel', 'barri', 'bas', 'base',
'basebal', 'baselin', 'baseman', 'basemen', 'baserun', 'baserunn', 'bash',
'basi', 'basic', 'basketbal', 'bass', 'bat', 'batch', 'batter', 'batti',
'battl', 'baub', 'baud', 'bay', 'baylor', 'bb', 'bbddd', 'bbs', 'bby', 'bbzx',
'bc', 'bchm', 'bd', 'beach', 'beacon', 'beam', 'bear', 'beast', 'beastmast',
'beat', 'beaten', 'beauti', 'becam', 'beck', 'becom', 'bedfellow', 'beef',
'beer', 'beg', 'began', 'begin', 'begun', 'behalf', 'behavior', 'bel', 'belat',
'belgium', 'belief', 'believ', 'bell', 'belmont', 'belong', 'belt', 'bem',

'ben', 'bench', 'benchmark', 'bend', 'benefit', 'benjamin', 'bennett', 'berg',
'berkeley', 'berlin', 'berman', 'bernard', 'bernstein', 'berryhil', 'besid',
'best', 'bet', 'beta', 'better', 'beverag', 'bewar', 'bf', 'bg', 'bgsu', 'bh',
'bhi', 'bhj', 'bhjn', 'bhjnux', 'bhjnuy', 'bi', 'bias', 'bichett', 'big',
'bigger', 'biggest', 'biggio', 'billboard', 'billi', 'billion', 'binari', 'bio',
'biochem', 'biochemistri', 'biochemistry', 'biographi', 'biolog', 'bird',
'bishop', 'bison', 'bit', 'bitch', 'bitmap', 'bitnet', 'biz', 'bizarr', 'bizw',
'bizwt', 'bj', 'bj200', 'bk', 'bl', 'black', 'blah', 'blame', 'blank', 'blast',
'blauser', 'bleacher', 'bleah', 'bless', 'blew', 'blind', 'blip', 'block',
'blomberg', 'blood', 'blow', 'blown', 'blue', 'blurb', 'blvd', 'bm', 'bmdhh286',
'bmp', 'bn', 'bnh', 'bnkjz', 'bnr', 'bnsc', 'bo', 'board', 'bob', 'bobbi',
'boca', 'bode', 'bodi', 'boe', 'boeing', 'boell', 'bogg', 'boi', 'bois', 'bolt',
'bomb', 'bond', 'bone', 'bongo', 'bonilla', 'bonn', 'bonus', 'book', 'boom',
'boon', 'boost', 'booster', 'boot', 'booth', 'bootup', 'borden', 'border',
'bore', 'borland', 'born', 'borrow', 'bosco', 'bosox', 'boss', 'boston',
'bother', 'bought', 'boulder', 'bounc', 'bound', 'bout', 'bowl', 'box', 'boy',
'bozo', 'bp', 'bps', 'bpu', 'bq', 'bql', 'br', 'brad', 'brader', 'bradley',
'brahms', 'brain', 'branch', 'brandei', 'brandeis', 'bratt', 'brave', 'bread',
'break', 'breakup', 'bream', 'breath', 'breeder', 'brendan', 'bret', 'brett',
'brewer', 'brian', 'brief', 'briefli', 'brigham', 'bright', 'bring', 'british',
'broad', 'broadcast', 'brochur', 'broke', 'broken', 'bronx', 'brook',
'brooklin', 'brother', 'brought', 'brown', 'brows', 'bruce', 'bruchner',
'brunel', 'bruno', 'brunswick', 'brute', 'bryan', 'bs', 'bs0t', 'bs0tq',
'bs0tq6', 'bs3', 'bsu', 'bsuvc', 'bt', 'btw', 'bu', 'bubblejet', 'buc', 'buck',
'buckeridg', 'bud', 'buddi', 'budget', 'buffalo', 'buffer', 'bug', 'build',
'builder', 'built', 'bulk', 'bull', 'bulletin', 'bullpen', 'bun', 'bunch',
'bundl', 'bunker', 'bunt', 'burdett', 'bureau', 'buri', 'burn', 'burnabi',
'burst', 'burstein', 'burster', 'bus', 'bush', 'busi', 'butler', 'butt',
'button', 'buy', 'bv', 'bw', 'bx', 'bxlt', 'bxn', 'bxom', 'bye', 'byte', 'byu',
'bz', 'c0', 'c2', 'c3', 'c4', 'c4r', 'c4u', 'c5', 'c5de05a0', 'c5o1yq',
'c5t05k', 'c6', 'c7', 'c7ff', 'c8', 'c8v', 'c9', 'c_', 'ca', 'cabell', 'cabl',
'cach', 'cacs', 'cadlab', 'cae', 'cage', 'cal', 'calcul', 'calderon', 'calgari',
'calib', 'calif', 'california', 'caller', 'calpoly', 'caltech', 'cam',
'cambridg', 'camden', 'came', 'camelot', 'camera', 'camin', 'camp', 'campaign',
'campanella', 'campbel', 'campus', 'canada', 'canadian', 'cancel', 'candi',
'candid', 'canon', 'canseco', 'canyon', 'cap', 'capabl', 'capac', 'cape',
'capit', 'capsul', 'captain', 'captur', 'car', 'carbon', 'card', 'cardin',
'care', 'career', 'carew', 'cargo', 'carl', 'carleton', 'carlo', 'carlosn',
'carlton', 'carnegi', 'carolina', 'carri', 'carrier', 'carter', 'cartridg',
'casar', 'casbah', 'case', 'casey', 'cash', 'cassini', 'cast', 'castillo',
'cat', 'catalog', 'catalogu', 'catch', 'catcher', 'categori', 'caterpillar',
'caught', 'caus', 'caution', 'cb', 'cbis', 'cbnewsg', 'cc', 'cc_sysk',
'ccastco', 'ccd', 'ccit', 'cck', 'cco', 'ccs', 'ccsvax', 'cd', 'cda90038',
'cdc', 'cdy', 'ce', 'celesti', 'cem', 'cent', 'centaur', 'center', 'centr',
'central', 'centuri', 'ceo', 'certain', 'certif', 'certifi', 'ces', 'cf', 'cfa',
'cg', 'ch', 'chahin', 'chain', 'chairman', 'challeng', 'chamberlain', 'champ',
'champion', 'chanc', 'chang', 'channel', 'chapel', 'chapter', 'char', 'charact',
'character', 'characterist', 'charg', 'charl', 'charli', 'charon', 'chart',

'che', 'cheap', 'cheaper', 'cheapest', 'cheapli', 'cheat', 'check', 'checker',
'cheer', 'chem', 'chemic', 'chemistri', 'chen', 'chest', 'chi', 'chiao',
'chicago', 'chicogo', 'chief', 'child', 'children', 'chines', 'chip', 'chlorin',
'cho', 'chocol', 'choic', 'choke', 'choos', 'chop', 'chose', 'chosen', 'chp',
'chris', 'christian', 'christoph', 'chu', 'chuck', 'churchil', 'chz', 'chzd9',
'ci', 'ci2j', 'cica', 'cii', 'cim', 'cinci', 'cincinnati', 'circa', 'circuit',
'circular', 'circumfer', 'circumst', 'cirrus', 'cis', 'cisc', 'cit', 'cite',
'citi', 'citizen', 'civil', 'civilian', 'cj', 'cj1', 'ck', 'ck8', 'cka52397',
'cl', 'claebaur', 'claim', 'claremont', 'clarif', 'clarifi', 'clark', 'clarku',
'class', 'classic', 'claudio', 'claus', 'clay', 'clayton', 'clean', 'clear',
'clemen', 'clementin', 'cleveland', 'clever', 'clich', 'click', 'client',
'cliff', 'clinton', 'clipper', 'clive', 'clobber', 'clock', 'clone', 'close',
'closer', 'closest', 'closeup', 'cloud', 'clove', 'club', 'clubhous', 'clue',
'clueless', 'cluster', 'clutch', 'cm', 'cmk', 'cmu', 'cn', 'coach', 'coast',
'coat', 'cobb', 'cod', 'code', 'coe', 'coffe', 'coffman', 'cogsci', 'cohen',
'coincid', 'col', 'cold', 'cole', 'coleman', 'coliseum', 'collaps', 'colleagu',
'collect', 'colleg', 'collid', 'collin', 'collis', 'coloni', 'color',
'colorado', 'colostate', 'colour', 'columbia', 'columbus', 'column', 'com',
'com1', 'com2', 'com3', 'comb', 'combin', 'combo', 'comdex', 'come', 'comeback',
'comet', 'comfort', 'comiskey', 'comm', 'command', 'comment', 'commer',
'commerci', 'commerici', 'commit', 'committe', 'common', 'commonwealth',
'communic', 'communiti', 'comp', 'compact', 'compani', 'company', 'compar',
'comparison', 'compat', 'compens', 'compet', 'competit', 'compil', 'complain',
'complaint', 'complet', 'complex', 'complic', 'compon', 'compos', 'composit',
'compound', 'comprehens', 'compress', 'compromis', 'compton', 'compuserv',
'compuserve', 'comput', 'conceiv', 'concentr', 'concept', 'concern', 'conclud',
'conclus', 'concordia', 'condit', 'condor', 'conduct', 'cone', 'confer',
'confernc', 'confid', 'config', 'configur', 'confin', 'confirm', 'conflict',
'confus', 'congress', 'congression', 'connect', 'connor', 'consecut', 'conserv',
'consid', 'consider', 'consist', 'consrt', 'constant', 'constantino', 'constel',
'constitut', 'constraint', 'construct', 'consult', 'consum', 'contact',
'contain', 'contend', 'content', 'contest', 'context', 'contigu', 'contin',
'continent', 'continu', 'continuin', 'contract', 'contractor', 'contractu',
'contradict', 'contrari', 'contrast', 'contribut', 'contributor', 'control',
'controversi', 'conveni', 'convent', 'convers', 'convert', 'convex', 'convinc',
'cook', 'cool', 'cooper', 'coordin', 'copi', 'copyright', 'core', 'corel',
'coreldraw', 'corelmark', 'cornel', 'cornell', 'corner', 'corp', 'corpor',
'correct', 'correl', 'correspond', 'corrupt', 'corval', 'cosmic', 'cosmos',
'cost', 'costa', 'costar', 'couldn', 'council', 'count', 'counter', 'counti',
'countri', 'coupl', 'cours', 'court', 'courtesi', 'cov', 'coventri', 'coventry',
'cover', 'coverag', 'cowboy', 'cox', 'cp', 'cpa', 'cps', 'cpu', 'cpus', 'cq',
'cr', 'crack', 'cradl', 'craft', 'craig', 'crap', 'crash', 'crater', 'crazi',
'crchh7a9', 'creat', 'creation', 'creativ', 'credit', 'crew', 'cri', 'crisi',
'critic', 'cross', 'crosspost', 'crowd', 'crucial', 'cruis', 'crush', 'cruz',
'cruz', 'cryogen', 'cryptic', 'crystal', 'cs', 'cs902043', 'csc', 'csc2imd',
'csci', 'csd', 'csd4', 'cse', 'csie', 'csn', 'cso', 'cspara', 'ct', 'cti',
'ctl', 'ctrl', 'cts', 'cu', 'cub', 'cubic', 'cue', 'cultur', 'cum', 'cup',
'cupertino', 'cure', 'curios', 'curious', 'current', 'curs', 'cursor', 'curti',

'curv', 'custom', 'cut', 'cutoff', 'cv', 'cw', 'cwru', 'cx', 'cy', 'cycl',
'cygnus', 'cylind', 'cynic', 'cz', 'd0', 'd1', 'd2', 'd2_', 'd3', 'd3hz', 'd3k',
'd4', 'd5', 'd6', 'd6f', 'd6g', 'd7', 'd8', 'd9', 'da', 'dabl2', 'dad', 'daili',
'dak988', 'dakota', 'dal', 'dale', 'dalhousi', 'dalla', 'damag', 'dame', 'damn',
'dan', 'danc', 'dandi', 'danger', 'dani', 'daniel', 'danni', 'dannyb', 'dant',
'dark', 'darn', 'darrel', 'darren', 'darryl', 'das', 'dat', 'data', 'databas',
'dataproduct', 'date', 'daughter', 'daulton', 'dave', 'davewood', 'davi',
'david', 'daviss', 'dawson', 'day', 'daylight', 'daytim', 'dayton', 'db', 'db6',
'dbldisk', 'dbm0000', 'dc', 'dcx', 'dead', 'deadlin', 'deal', 'dean', 'dear',
'death', 'debat', 'debri', 'dec', 'decad', 'deceler', 'decemb', 'decent',
'decid', 'decis', 'deck', 'declar', 'declin', 'decnet', 'decreas', 'decvax',
'dedic', 'deep', 'deeper', 'deer', 'default', 'defeat', 'defend', 'defenit',
'defens', 'defin', 'definit', 'deflat', 'defrag', 'defray', 'degrad', 'degre',
'deion', 'deja', 'del', 'delawar', 'delay', 'delet', 'deliber', 'delino',
'deliv', 'deliveri', 'delmarva', 'delta', 'deluca', 'demand', 'demer', 'demo',
'demon', 'demonstr', 'deni', 'denni', 'dennism', 'dens', 'densiti', 'denver',
'depart', 'depend', 'deploy', 'depress', 'dept', 'depth', 'deputi', 'der',
'derbi', 'derek', 'deriv', 'des', 'descend', 'describ', 'descript', 'desert',
'deserv', 'deshield', 'design', 'desir', 'desk', 'deskjet', 'desktop', 'despit',
'destin', 'destruct', 'detect', 'detector', 'determin', 'deton', 'detroit',
'detweil', 'dev', 'develop', 'devic', 'devineni', 'devon', 'devot', 'dewint',
'df', 'dfrf', 'dg', 'dh', 'di', 'dia', 'diagnost', 'diagram', 'dial', 'dialog',
'diamet', 'diamond', 'dick', 'did', 'didn', 'die', 'diego', 'dietz', 'diff',
'differ', 'differenti', 'difficult', 'difficulti', 'dig', 'digest', 'digex',
'digit', 'dil', 'dillon', 'dime', 'dimens', 'dimension', 'dinger', 'dir',
'direct', 'director', 'directori', 'dirti', 'dis', 'dis_pkt', 'disabl',
'disagre', 'disappear', 'disappoint', 'disastr', 'discharges', 'disclaim',
'discount', 'discov', 'discoveri', 'discretionari', 'discrimin', 'discuss',
'dishard', 'disk', 'diskett', 'display', 'dispos', 'disput', 'distanc',
'distant', 'distinguish', 'distort', 'distress', 'distribut', 'disturb', 'dive',
'divid', 'divin', 'divis', 'division', 'divison', 'dj', 'djf', 'dk', 'dl',
'dle', 'dll', 'dm', 'dmcaloon', 'dmn', 'dn', 'dnt', 'doc', 'dock', 'doctor',
'document', 'dod', 'dodger', 'doe', 'does', 'doesn', 'dog', 'doing', 'doka',
'dollar', 'dolven', 'domain', 'dome', 'domin', 'don', 'donald', 'donat', 'dong',
'dont', 'doom', 'door', 'doppelganger', 'dos', 'dos6', 'dot', 'doubl', 'double',
'doublespac', 'doubt', 'doug', 'dougl', 'downey', 'download', 'downward',
'dozen', 'dp', 'dpi', 'dr', 'draft', 'drag', 'dragon', 'drake', 'drastic',
'draw', 'dream', 'drexel', 'dri', 'drift', 'drink', 'drive', 'driven', 'driver',
'drivers', 'drop', 'drove', 'drv', 'drw3l', 'dryden', 'ds', 'dseg', 'dsg',
'dsn', 'dswartz', 'dt', 'dtate', 'dtp', 'du', 'dual', 'dubinski', 'dubious',
'dublin', 'duc', 'dud', 'dug', 'duh', 'duk', 'duke', 'dumb', 'dump', 'dunn',
'dunno', 'duplic', 'durat', 'durham', 'dust', 'duti', 'dv', 'dw', 'dwarf',
'dwarner', 'dx', 'dx2', 'dy', 'dykstra', 'dylan', 'dynam', 'dynamit', 'dz',
'e0', 'e1', 'e145', 'e14di', 'e14dm75u', 'e1t', 'e2', 'e4', 'e5', 'e7', 'e8',
'e9c', 'e9l0qax', 'e_', 'ea', 'eagl', 'ear', 'earl', 'earli', 'earlier', 'earn',
'earnest', 'earth', 'eas', 'easi', 'easier', 'easili', 'east', 'eastern',
'eastgat', 'easy', 'eat', 'eb', 'ebosco', 'ec', 'eccentr', 'ece', 'ecf', 'echo',
'echonolog', 'eci', 'eckton', 'eclips', 'ecn', 'eco', 'econ', 'econom',

'economi', 'ecs', 'ed', 'eddi', 'eder', 'edg', 'edit', 'editor', 'edmoor',
'edo', 'edu', 'educ', 'edward', 'ee', 'eec', 'eecg', 'ef', 'effect', 'effici',
'effort', 'egalon', 'eh', 'ei', 'ei0l', 'ei0mfq', 'ei4', 'eid', 'eighth',
'eindhoven', 'einstein', 'eir', 'eisa', 'ej', 'ek', 'el', 'elabor', 'eldr',
'elect', 'electr', 'electric', 'electron', 'element', 'eli', 'elimin', 'elk',
'els', 'elsewher', 'elv', 'elvi', 'elwood', 'em', 'email', 'emb', 'embarass',
'embarrass', 'embed', 'emerg', 'emiss', 'emm386', 'emori', 'emphasi', 'empir',
'employ', 'employe', 'empti', 'emr', 'emu', 'emul', 'emx', 'en', 'enabl',
'encod', 'encount', 'encourag', 'end', 'endeavour', 'endor', 'energet',
'energi', 'energia', 'energietechnik', 'enfant', 'eng', 'engin', 'england',
'english', 'engrg', 'enhanc', 'enjoy', 'enlighten', 'enorm', 'enrich', 'ensur',
'enter', 'enterpris', 'entertain', 'enthusiast', 'entir', 'entitl', 'entri',
'envelop', 'enviro', 'environ', 'environment', 'enzo', 'eo', 'eos', 'eosvcr',
'ep', 'ephemeri', 'episod', 'epson', 'eq', 'equal', 'equat', 'equip', 'equival',
'er', 'era', 'erect', 'eric', 'erickson', 'erni', 'error', 'es', 'esa', 'esc',
'escap', 'eso', 'esoter', 'esp', 'especi', 'espn', 'essenc', 'essenti', 'est',
'establish', 'estim', 'et', 'et4000', 'etdesg', 'etern', 'etf', 'ether',
'ethic', 'etowrk', 'etrat', 'eu', 'eugen', 'europ', 'european', 'euv', 'ev',
'eva', 'evalu', 'evan', 'event', 'eventu', 'everi', 'everybodi', 'everyday',
'everyon', 'everyth', 'everytim', 'evid', 'evil', 'evolut', 'evsc', 'evz', 'ew',
'ex', 'exact', 'examin', 'exampl', 'exceed', 'excel', 'excerpt', 'exchang',
'excit', 'exclud', 'exclus', 'excus', 'execut', 'exercis', 'exhaust', 'exhibit',
'exist', 'exit', 'expand', 'expans', 'expect', 'expend', 'expensitur', 'expens',
'experi', 'experienc', 'experiment', 'expert', 'expir', 'explain', 'explan',
'explod', 'exploit', 'explor', 'explos', 'expo', 'export', 'expos', 'exposur',
'express', 'ext', 'extend', 'extens', 'extent', 'extern', 'extinguish', 'extra',
'extract', 'extrem', 'ey', 'eye', 'ez', 'ez027993', 'f0', 'f1', 'f2', 'f23',
'f3', 'f349', 'f3t', 'f3w2', 'f3w2tg', 'f4', 'f5', 'f6', 'f7', 'f8', 'f9',
'f9d', 'f9f9', 'f9f9f0', 'f9f9f3t', 'f9f9f9d', 'f9f9f9f9', 'f9f9f9f9f0',
'f9f9f9f9f9d', 'f_', 'fa', 'faa', 'fac', 'face', 'facil', 'fact', 'factor',
'faculti', 'fade', 'fahrenheit', 'fail', 'failur', 'faint', 'fair', 'fairbank',
'faith', 'fall', 'fallaci', 'fame', 'famer', 'famili', 'familiar', 'famous',
'fan', 'fanat', 'fanci', 'fantasi', 'faq', 'far', 'farley', 'farm', 'farr',
'farsid', 'fascin', 'fashion', 'fast', 'fastbal', 'faster', 'fastest', 'fat',
'father', 'fault', 'fav', 'favor', 'favorit', 'fax', 'fb', 'fc', 'fd', 'fdo',
'fdz', 'fe', 'fear', 'feasabl', 'feasibl', 'feat', 'featur', 'feb', 'februari',
'fed', 'feder', 'fedex', 'fee', 'feed', 'feedback', 'feel', 'feet', 'fei',
'fel', 'feldman', 'felix', 'fell', 'fellow', 'fellrath', 'felt', 'femal',
'fenc', 'fenway', 'ferment', 'fermi', 'fermilab', 'fernandez', 'fernando',
'fester', 'fever', 'fewer', 'ff', 'ff1', 'fg', 'fh', 'fi', 'fiasco', 'fiber',
'fiction', 'fido', 'fidonet', 'field', 'fielder', 'fierkelab', 'fifth', 'fifti',
'fight', 'fighter', 'figur', 'fij', 'fil', 'file', 'filenam', 'film', 'final',
'financ', 'financi', 'fine', 'finger', 'finish', 'finley', 'finnal', 'firm',
'firststep', 'fiscal', 'fischer', 'fish', 'fisk', 'fission', 'fit', 'fix', 'fj',
'fk', 'fl', 'flame', 'flanagan', 'flare', 'flash', 'flat', 'flatter', 'flaw',
'flb', 'fleet', 'flew', 'flexibl', 'fli', 'flight', 'flint', 'flip', 'float',
'flood', 'floor', 'floppi', 'florida', 'flow', 'flown', 'fls', 'fluid', 'fluke',
'fly', 'flybi', 'fm', 'fn', 'fnal', 'fnalf', 'fng', 'fo', 'foci', 'focus',

'foil', 'fold', 'folk', 'follow', 'followup', 'font', 'food', 'fool', 'foolish',
'foot', 'footbal', 'forc', 'forecast', 'foreign', 'foresight', 'forev',
'forget', 'forgiv', 'forgot', 'forgotten', 'form', 'formal', 'format',
'formula', 'forrest', 'fort', 'forth', 'fortran', 'fortun', 'forum', 'forward',
'foster', 'foul', 'foundat', 'fourth', 'fox', 'foxx', 'fp', 'fq', 'fql', 'fr',
'fraction', 'fraer', 'frame', 'framework', 'franchis', 'francisco', 'frank',
'franklin', 'fraser', 'fre', 'freak', 'fred', 'frederick', 'free', 'freebairn',
'freed', 'freedom', 'freeli', 'freeman', 'freenet', 'freewar', 'freez',
'fregosi', 'french', 'frequenc', 'frequent', 'fresh', 'fri', 'friday', 'friend',
'fring', 'frog', 'frontier', 'frungi', 'frustrat', 'fryman', 'fs', 'fs7', 'fsu',
'fsx', 'ft', 'ftcollinsco', 'ftp', 'ftpmail', 'ftz', 'fu', 'fuel', 'fuess',
'fulfil', 'fulli', 'fun', 'function', 'fund', 'fundament', 'funni',
'furthermor', 'fuselag', 'fusion', 'futur', 'fuzzi', 'fv', 'fvmk', 'fw', 'fx',
'fy', 'fyl', 'fyn', 'fyw', 'fz', 'fz0', 'g0', 'g1', 'g2', 'g215a', 'g3', 'g4',
'g4p', 'g5', 'g6', 'g7', 'g8', 'g9', 'g9p', 'g9u', 'g9v', 'g_', 'ga', 'gabriel',
'gaetti', 'gagn', 'gain', 'gaj', 'gajarski', 'gal', 'galact', 'galarraga',
'galaxi', 'galaxy', 'galileo', 'gam', 'game', 'gamma', 'gang', 'gant', 'gap',
'garbag', 'gardner', 'gari', 'garnet', 'garrison', 'garvey', 'gas', 'gate',
'gatech', 'gateway', 'gather', 'gave', 'gb', 'gc', 'gd', 'ge', 'gee', 'geez',
'gehrel', 'gehrels3', 'gehrig', 'gem', 'gemini', 'gen', 'gene', 'gener',
'general', 'generat', 'generic', 'generous', 'genetic', 'genius', 'genral',
'geo', 'geochem', 'geod', 'geoffrey', 'geolog', 'geophys', 'georg',
'georgetown', 'georgia', 'geostationari', 'geosynchron', 'gerald', 'german',
'germani', 'gerri', 'gestur', 'gf', 'gfk39017', 'gg', 'gg5u', 'gh', 'ghhwang',
'ghj', 'ghjn', 'ghost', 'ghostscript', 'gi', 'giant', 'gibson', 'gif', 'gilkey',
'girl', 'given', 'givinh', 'giz', 'gizw', 'gizwt', 'gj', 'gk', 'gl', 'glad',
'glavin', 'glenn', 'glitch', 'global', 'globe', 'glori', 'glove', 'gm', 'gm2',
'gmt', 'gmu', 'gmuvax2', 'gn', 'gnb', 'gnp', 'goal', 'god', 'goddard', 'goe',
'gold', 'golden', 'goldin', 'golf', 'goltz', 'gomez', 'gon', 'gone', 'gonzal',
'gonzalez', 'good', 'gord', 'gordon', 'gorman', 'gosh', 'got', 'gothamcity',
'gotten', 'gov', 'govern', 'government', 'govt', 'gp', 'gpf', 'gps', 'gq',
'gq32', 'gr', 'grab', 'grace', 'grad', 'grade', 'graduat', 'grafitti', 'grand',
'grant', 'graph', 'graphic', 'graphit', 'grappl', 'grate', 'gravit', 'graviti',
'gray', 'graydon', 'great', 'greater', 'greatest', 'green', 'greenbelt',
'greenberg', 'greet', 'greg', 'gregg', 'gregori', 'grey', 'grief', 'griffey',
'griffin', 'gross', 'groucho', 'ground', 'grounder', 'group', 'grove', 'grow',
'growth', 'grp', 'gruber', 'gs', 'gsfc', 'gspira', 'gsx', 'gt', 'gt0523e',
'gt7469a', 'gtonwu', 'gu', 'guarante', 'guess', 'guest', 'gui', 'guid',
'guidanc', 'guidelin', 'guilti', 'gun', 'gup', 'gupta', 'gurus', 'gut', 'guvax',
'guy', 'guzman', 'gv', 'gw', 'gwu', 'gwynn', 'gx', 'gy', 'gyro', 'h0', 'h1',
'h2', 'h3', 'h4', 'h5', 'h6', 'h7', 'h8', 'h8z', 'h9', 'h_', 'ha', 'haapanen',
'habit', 'habitat', 'hack', 'hacker', 'hain', 'hair', 'hal', 'half', 'halifax',
'hall', 'ham', 'hamilton', 'hammer', 'hancock', 'hand', 'handbook', 'hander',
'handi', 'handl', 'hang', 'hank', 'hannib', 'hansen', 'hanson', 'happen',
'happi', 'hard', 'harder', 'hardwar', 'hardy', 'harkey', 'harmon', 'harmoni',
'harold', 'harper', 'harri', 'harvard', 'harvey', 'hasn', 'hassl', 'hat',
'hatcher', 'hate', 'haven', 'having', 'hawaii', 'hawk', 'hay', 'hayesj',
'hazard', 'hb', 'hc', 'hcf', 'hd', 'head', 'header', 'headlin', 'headquart',

'health', 'healthi', 'heap', 'hear', 'heard', 'hearsay', 'heart', 'heat',
'heaven', 'heavi', 'heavier', 'heavili', 'heck', 'height', 'heinlein', 'hela',
'held', 'helios', 'helium', 'hell', 'hello', 'help', 'helsinki', 'henc',
'henderson', 'henri', 'hep', 'hepnet', 'herc', 'hercul', 'herm', 'hermes',
'hernandez', 'hero', 'hesketh', 'hewlett', 'hex', 'hey', 'hf', 'hg', 'hga',
'hh', 'hh8', 'hi', 'hic', 'hidden', 'hideous', 'hierarchi', 'higgin', 'high',
'higher', 'highest', 'highlight', 'highway', 'hill', 'hint', 'hip', 'hire',
'histor', 'histori', 'hit', 'hite', 'hiten', 'hitt', 'hitter', 'hj', 'hk', 'hl',
'hlp', 'hlv', 'hlvs', 'hm', 'hm9', 'hmm', 'hmmm', 'hn', 'ho', 'hobbi',
'hoboken', 'hobokenit', 'hockey', 'hof', 'hoff', 'hog', 'hoil', 'hojo', 'hold',
'holder', 'hole', 'holiday', 'holli', 'hollin', 'hom', 'home', 'homer',
'homerun', 'homewood', 'honest', 'honor', 'hood', 'hook', 'hope', 'hopkin',
'horizon', 'horizont', 'hormon', 'horner', 'horribl', 'hors', 'hose', 'host',
'hot', 'hotel', 'hound', 'hour', 'hous', 'houston', 'hov', 'hover', 'howabout',
'howard', 'howard_wong', 'howe', 'howev', 'hp', 'hp4', 'hpcc01', 'hplab', 'hq',
'hr', 'hrs', 'hs', 'hsinchu', 'hst', 'hsvaic', 'ht', 'hte', 'hu', 'hua',
'hubbl', 'huckabay', 'hug', 'huge', 'hugh', 'huh', 'hulman', 'hum', 'human',
'humbl', 'humili', 'humor', 'hundr', 'hung', 'hunt', 'hunter', 'huntsvill',
'hurri', 'hurt', 'husband', 'hut', 'hv', 'hw', 'hx', 'hy', 'hydra', 'hydrogen',
'hype', 'hyperspac', 'hz', 'hzri', 'i0', 'i01', 'i0mf', 'i0mf9l3', 'i0mfq',
'i0o', 'i0w', 'i1', 'i2', 'i24', 'i3', 'i4', 'i5', 'i6', 'i7', 'i8', 'i9', 'i_',
'ia', 'iacs3650', 'iai', 'ial', 'ian', 'iastate', 'iau', 'ib', 'ibm', 'ic',
'icbm', 'ice', 'icon', 'ics', 'id', 'id9', 'ida', 'ide', 'idea', 'ideal',
'idealist', 'ident', 'identifi', 'idiot', 'idk', 'idl', 'ifa', 'iglesia',
'ignor', 'ih', 'ihlpm', 'ii', 'iii', 'ij', 'ik', 'ikesgl', 'il', 'iler', 'ill',
'illeg', 'illinoi', 'illumin', 'illustr', 'im', 'imag', 'imagin', 'imaginit',
'imho', 'imi', 'immatur', 'immedi', 'immens', 'immort', 'imo', 'impact',
'impair', 'implement', 'implent', 'impli', 'implod', 'import', 'imposs',
'impress', 'improv', 'impuls', 'incaviglia', 'incent', 'inch', 'incid',
'inclin', 'includ', 'incom', 'incompat', 'inconst', 'incorpor', 'incorrect',
'increas', 'incred', 'increment', 'incrib', 'inde', 'independ', 'index',
'indian', 'indiana', 'indic', 'individu', 'industri', 'indyvax', 'inevit',
'inexpens', 'infam', 'infer', 'inferior', 'infield', 'infin', 'infinitt',
'inflat', 'influenc', 'info', 'inform', 'informatica', 'informix', 'infoserv',
'infoworld', 'infrar', 'infrastructur', 'ing', 'ingr', 'inher', 'ini', 'initi',
'inject', 'injur', 'injuri', 'ink', 'inland', 'inner', 'inning', 'innov',
'input', 'inqmind', 'inquir', 'inquiri', 'ins', 'insert', 'insid', 'insight',
'insist', 'inspect', 'inspector', 'inspir', 'inst', 'instabl', 'instal',
'instanc', 'instant', 'instead', 'institut', 'instruct', 'instrument', 'insul',
'insult', 'insur', 'int', 'intact', 'integr', 'intel', 'intellect',
'intellectu', 'intellig', 'intend', 'intens', 'intent', 'inter', 'interact',
'intercept', 'interf', 'interfac', 'interfer', 'intergr', 'intern', 'internet',
'interplanetari', 'interpret', 'interrupt', 'interstellar', 'interv',
'interview', 'intrigu', 'introduc', 'introduct', 'invalid', 'invent',
'inventor', 'invest', 'investig', 'investor', 'invit', 'invok', 'involv', 'io',
'iowa', 'ioy', 'ip', 'ipx', 'iq', 'iqi', 'iqr', 'iqvg', 'ir', 'ira', 'ireland',
'iron', 'irq', 'irq5', 'irrelev', 'isa', 'isbn', 'isc', 'isi', 'isis',
'iskandar', 'island', 'isn', 'iso', 'isol', 'isp', 'issu', 'ist', 'isu',

'isunet', 'isv', 'itch', 'item', 'ithaca', 'iti', 'iu', 'iubac', 'iupui', 'iv',
'ivi', 'ix', 'iy', 'iz', 'j0', 'j1', 'j1d', 'j2', 'j3', 'j4', 'j5', 'j6', 'j7',
'j8', 'j9', 'j_', 'ja', 'jack', 'jackhammering', 'jacki', 'jackson', 'jacob',
'jacobson', 'jail', 'jame', 'jami', 'jan', 'jane', 'januari', 'japan',
'japanes', 'jason', 'javier', 'jay', 'jayson', 'jb', 'jbh55289', 'jc', 'jd',
'jdnicol', 'je', 'jeff', 'jefferi', 'jeffrey', 'jemison', 'jenk', 'jeopardi',
'jeremi', 'jerk', 'jerri', 'jersey', 'jess', 'jet', 'jewelri', 'jewish', 'jf',
'jg', 'jgarland', 'jgreen', 'jh', 'jhu', 'jhuapl', 'jhunix', 'jhuvms', 'ji',
'jiann', 'jim', 'jimmi', 'jj', 'jja2h', 'jjj', 'jk', 'jking', 'jl', 'jm',
'jmcocker', 'jn', 'jo', 'job', 'jockey', 'jodi', 'joe', 'joel', 'john',
'johnni', 'johnson', 'join', 'joint', 'joke', 'jolla', 'jon', 'jonathan',
'jone', 'jordan', 'jose', 'joseph', 'josh', 'journal', 'journalism', 'journey',
'jove', 'jovian', 'joy', 'jp', 'jp2', 'jpaparel', 'jpl', 'jpv', 'jpw', 'jq',
'jr', 'jrl', 'jrogoff', 'js', 'js1', 'jsc', 'jt', 'jtc', 'jtcent', 'jtchern',
'ju', 'juan', 'judg', 'judy', 'jule', 'juli', 'jump', 'jumpi', 'jun', 'june',
'junk', 'jupit', 'jupiter', 'jus', 'just', 'justic', 'justifi', 'justin', 'jv',
'jw', 'jx', 'jz', 'k0', 'k1', 'k4', 'k5', 'k6', 'k7', 'k8', 'k80', 'k86', 'k9',
'k_', 'k_mullin', 'ka', 'kansa', 'karl', 'karla', 'karlin', 'kasajian', 'kathi',
'kayman', 'kb', 'kc', 'kd', 'ke', 'ke4zv', 'kean', 'kearney', 'keele', 'keith',
'keithley', 'kelli', 'kelso', 'kelvin', 'ken', 'kennedi', 'kenneth', 'kent',
'kentil', 'keplerian', 'kept', 'kermit', 'kernel', 'kerr', 'kevin', 'key',
'keyboard', 'keynes', 'keyword', 'kf', 'kg', 'kh', 'khf', 'khz', 'ki', 'kick',
'kid', 'kill', 'kilmer', 'kilogram', 'kilomet', 'kim', 'kimbal', 'kimbark',
'kin', 'kind', 'kinda', 'kiner', 'king', 'kingdom', 'kingman', 'kingoz',
'kingston', 'kipl', 'kippur', 'kir', 'kirbi', 'kirk', 'kirsch', 'kit', 'kj',
'kjenk', 'kjr', 'kjb', 'kjznb', 'kk', 'kkzp', 'kl', 'klesko', 'klopfen',
'klopfenstein', 'km', 'kmembri', 'kn', 'knee', 'knew', 'knight', 'knoblauch',
'knock', 'know', 'knowledg', 'known', 'ko', 'kong', 'kotfr', 'koufax', 'kourou',
'kozierok', 'kp', 'kp3', 'kq', 'kqv', 'kr', 'kretzschmar', 'krill', 'krueger',
'kruk', 'krumin', 'ks', 'ksc', 'ksr', 'kstar', 'ksu', 'kt', 'ku', 'kudo',
'kuip', 'kuiper', 'kuleuven', 'kuo', 'kurt', 'kv', 'kvp', 'kw', 'kx', 'ky',
'kz', 'l0', 'l2', 'l3', 'l4', 'l45u', 'l5', 'l6', 'l7', 'l8', 'l9', 'la', 'lab',
'label', 'laboratori', 'lack', 'ladi', 'lafayette', 'lafibm', 'lag', 'lah',
'laid', 'lake', 'laker', 'lame', 'lamont', 'lan', 'lanc', 'land', 'lander',
'landmann', 'landsat', 'landscap', 'lane', 'langley', 'languag', 'lankford',
'lankform', 'lans', 'lanza', 'larc', 'larg', 'larger', 'largest', 'larkin',
'larri', 'larrison', 'larussa', 'laser', 'laserjet', 'lasorda', 'lat', 'late',
'later', 'latest', 'latitud', 'laudicina', 'laugh', 'launch', 'launcher',
'laurentian', 'lavallier', 'law', 'lawrenc', 'lawson', 'layer', 'layn', 'lazi',
'lb', 'lbs', 'lc', 'lc4', 'lcs', 'ld', 'le', 'lead', 'leader', 'leadership',
'leadoff', 'leagu', 'leaguer', 'leak', 'lean', 'leap', 'learn', 'leas', 'leav',
'lectur', 'led', 'lee', 'leebr', 'leech', 'lefebr', 'lefebvr', 'left', 'lefti',
'lefty', 'leg', 'legal', 'legend', 'legisl', 'legitim', 'lehigh', 'lei',
'leigh', 'leius', 'leland', 'lemk', 'len', 'length', 'lenni', 'leo', 'lerc',
'lesson', 'lest', 'let', 'letter', 'level', 'levi', 'levin', 'lew', 'lewi',
'lewisvill', 'leyland', 'lf', 'lg', 'lga', 'lgo', 'lh', 'lho', 'lhz', 'li',
'liabil', 'lib', 'liber', 'librari', 'licens', 'lick', 'lie', 'lieft', 'liek',
'life', 'lifetim', 'lift', 'liftoff', 'light', 'lighten', 'lightn', 'liguori',

'lii', 'lik', 'like', 'lim', 'limit', 'limv', 'lin', 'lind', 'linda',
'lindbergh', 'linear', 'lineup', 'link', 'linsenbigl', 'linux', 'lion', 'lip',
'liquid', 'liris', 'lisa', 'lisbon', 'list', 'listen', 'lite', 'liter',
'literatur', 'littl', 'liu', 'liv', 'live', 'livermor', 'lj', 'lj1', 'lk',
'lk8', 'lkjz', 'll', 'llnl', 'lloyd', 'lm', 'ln', 'lo', 'load', 'loan', 'lobbi',
'local', 'locat', 'lock', 'lockhe', 'locus', 'log', 'logic', 'login',
'logitech', 'logo', 'london', 'long', 'longer', 'longest', 'longitud', 'lonni',
'loo', 'look', 'loop', 'loos', 'lopez', 'lord', 'los', 'lose', 'loser', 'loss',
'lost', 'lot', 'lotus', 'lou', 'loud', 'loui', 'louisiana', 'louisvill',
'louisville', 'louray', 'lousi', 'love', 'lover', 'low', 'lowel', 'lowenstein',
'lower', 'lowest', 'lox', 'lp', 'lpt1', 'lq', 'lr', 'lru', 'ls', 'lsl', 'lt',
'ltt', 'lu', 'luck', 'lucki', 'luddington', 'lui', 'luna', 'lunar', 'lunat',
'lunch', 'lundi', 'lung', 'lupica', 'luri', 'luriem', 'lustig', 'lv', 'lw',
'lwp', 'lx', 'ly', 'lyford', 'lz', 'lzt1260', 'm0', 'm0d', 'm0dtv5p', 'm1',
'm1t', 'm2', 'm24', 'm24e', 'm27', 'm2j', 'm3', 'm34', 'm34u', 'm3l', 'm4',
'm45', 'm4pa', 'm4u', 'm5', 'm5u', 'm6', 'm7', 'm7ex', 'm7ey', 'm7ez', 'm7klj',
'm7kn', 'm8', 'm86', 'm9', 'm9f8', 'm9f9fq', 'm9l0qax', 'm9l3', 'm9p', 'm9v',
'm_', 'm_klein', 'ma', 'ma4', 'ma7', 'mac', 'macadam', 'macdonald', 'mach',
'machin', 'macinterest', 'macintosh', 'mack', 'macro', 'mad', 'madam', 'maddux',
'madison', 'mae', 'mag', 'magadan', 'magazin', 'magellan', 'magic', 'magnet',
'magnetomet', 'magnitud', 'magowan', 'mahler', 'mai', 'mail', 'mailbox',
'mailer', 'mailhost', 'main', 'maine', 'mainfram', 'maintain', 'mainten', 'maj',
'major', 'make', 'maker', 'mal', 'malama', 'maldonado', 'male', 'malfunct',
'mam', 'mamatha', 'man', 'manag', 'mancus', 'maneu', 'mani', 'manifest',
'manipul', 'mankind', 'manner', 'mantl', 'manual', 'manufactur', 'manwar',
'map', 'mar', 'marc', 'march', 'marcus', 'margin', 'mari', 'marin', 'marinara',
'mario', 'mark', 'market', 'marlin', 'marquett', 'mars', 'marshal', 'martian',
'martin', 'martinez', 'marvel', 'marvin', 'marxist', 'maryland', 'mas', 'mass',
'massachusetts', 'massiv', 'master', 'mat', 'match', 'materi', 'math', 'mathcad',
'mathemat', 'maths', 'matloff', 'matt', 'matter', 'matthew', 'matur', 'max',
'maxim', 'maximum', 'maximus', 'maxtor', 'mayan', 'mayb', 'maynard', 'mb',
'mb4q', 'mb8e', 'mb8f', 'mb8g', 'mbn', 'mbp', 'mbs0t', 'mbs0tq', 'mbs3',
'mbxlt', 'mbxn', 'mbxom', 'mc', 'mc8', 'mcaloon', 'mccal', 'mccall', 'mccovey',
'mcdonald', 'mcdonnel', 'mcdowel', 'mcelwain', 'mcgee', 'mcgill', 'mcgraw',
'mcgriff', 'mcgwire', 'mckinley', 'mckissock', 'mcrae', 'mcwilliam', 'md',
'md1', 'md2', 'md2_', 'md8', 'md9', 'mdc', 'mdi', 'mead', 'meaddata', 'mean',
'meaning', 'meaningless', 'meant', 'meanwhil', 'measur', 'mechan', 'med',
'media', 'medic', 'medicin', 'mediocr', 'mediot', 'medium', 'mee', 'meet',
'meg', 'mei', 'mel', 'melbourn', 'melido', 'mellon', 'mem', 'member', 'membri',
'memor', 'memori', 'men', 'mention', 'mentorg', 'menu', 'menus', 'merchandis',
'mercuri', 'mere', 'merg', 'meridian', 'merit', 'merl', 'mess', 'messag', 'met',
'metal', 'meteor', 'meter', 'methink', 'method', 'methodolog', 'metric',
'metrics', 'mets', 'mexico', 'meyer', 'mf', 'mf0', 'mf3t', 'mf9d', 'mf9f9',
'mfc', 'mfyl', 'mg', 'mg9', 'mg9v', 'mh', 'mh8', 'mhd', 'mhs', 'mhz', 'mi',
'mi0l', 'mi0mfq', 'mi4', 'miami', 'mice', 'michael', 'michigan', 'mickey',
'micro', 'micrograv', 'microsoft', 'microwav', 'mid', 'middl', 'midi',
'midnight', 'midway', 'midwest', 'mig', 'mighti', 'mike', 'mil', 'mile',
'mileag', 'militari', 'militello', 'miller', 'million', 'millitello',

'milwaukee', 'mimi', 'min', 'mind', 'mindlink', 'miner', 'minim', 'minimum',
'minneapolis', 'minnesota', 'minor', 'minut', 'mip', 'mir', 'miracl', 'mirror',
'mis', 'misc', 'miscellan', 'misplac', 'miss', 'missil', 'mission',
'mississippi', 'missl', 'missouri', 'mistak', 'mistaken', 'misunderstand',
'mit', 'mitch', 'mitchel', 'mitsubishi', 'mix', 'mixtur', 'mizar', 'mize', 'mj',
'mjone', 'mk', 'mksol', 'ml', 'ml6', 'mlb', 'mle', 'mlinsenb', 'mlk', 'mm',
'mmk', 'mn', 'mnemosyne', 'mnk', 'mnuy', 'mo', 'mobil', 'mod', 'mode', 'model',
'modem', 'moder', 'modern', 'modest', 'modif', 'modifi', 'modul', 'modular',
'moe', 'mohney', 'moin', 'mojo', 'molecular', 'molitor', 'moment', 'momentum',
'mon', 'monash', 'monday', 'money', 'monica', 'monitor', 'monsoon', 'monster',
'montana', 'month', 'montreal', 'monu6', 'moon', 'moonbas', 'moop', 'moor',
'mop', 'moral', 'moratorium', 'morgan', 'morn', 'morri', 'mors', 'moscow',
'mot', 'mother', 'motherboard', 'motion', 'motiv', 'motor', 'motorola', 'mound',
'mount', 'mountain', 'mous', 'mouse', 'mouth', 'movement', 'movi', 'mp', 'mpce',
'mq', 'mq6', 'mqi', 'mqq', 'mqr', 'mqvg', 'mr', 'mr47', 'mr8f', 'mrfw', 'mrmc8',
'ms', 'msb', 'msdos', 'msfc', 'mss', 'msstate', 'msu', 'msuinfo', 'mt', 'mt7',
'mtg', 'mtm', 'mtt', 'mu', 'muenchen', 'mule', 'mulli', 'multi', 'multimedia',
'multipl', 'multitask', 'multithread', 'multius', 'mun', 'mundstock', 'muniz',
'munizb', 'murdoch', 'murmur', 'murphi', 'murray', 'muscl', 'museum', 'music',
'mussina', 'mv', 'mveraart', 'mvp', 'mw', 'mwa', 'mwt', 'mwwhj', 'mwwiz', 'mx',
'mx8', 'myer', 'mylar', 'mym', 'mysteri', 'myth', 'mz', 'n0', 'n1', 'n109',
'n2', 'n3', 'n5', 'n6', 'n8', 'n9', 'na', 'nag', 'nahf', 'nai', 'nail', 'najz',
'nam', 'nanderso', 'narrow', 'nasa', 'nasda', 'nashvill', 'nasm', 'nasp',
'nasti', 'nathan', 'nation', 'nativ', 'natur', 'naval', 'navi', 'navig',
'navstar', 'navy', 'nax', 'nb', 'nb8', 'nc', 'ncar', 'ncc', 'ncr', 'ncsu',
'nctu', 'nd', 'ndis', 'ndk', 'ndw', 'ne', 'ne1', 'neit', 'neal', 'near',
'nearbi', 'nearest', 'neat', 'nebraska', 'nec', 'necessari', 'necessarili',
'neck', 'need', 'needl', 'neff123', 'negat', 'neglig', 'negoti', 'nei', 'nei4',
'neil', 'nelson', 'neptun', 'net', 'netbeui', 'netcom', 'netherland', 'netnew',
'netnews', 'netwar', 'network', 'netx', 'neu', 'neutron', 'nevada', 'new',
'newbi', 'newer', 'newest', 'newkirk', 'news', 'newsgroup', 'newshost',
'newshub', 'newslett', 'newspap', 'newsread', 'newton', 'newtout', 'nextwork',
'nf', 'ng', 'ngu', 'nh', 'ni', 'nic', 'nice', 'nicho', 'nichol', 'nick',
'nickh', 'nicol', 'nie', 'niederstrass', 'niev', 'nifti', 'night', 'niguma',
'nih', 'niland', 'nimast', 'ninth', 'nixon', 'nj', 'njin', 'nk', 'nkh', 'nkhf',
'nki', 'nkjr', 'nkjz', 'nkz', 'nl', 'nlcs', 'nlm', 'nls', 'nlu', 'nm', 'nn',
'nntp', 'noaa', 'nobodi', 'node', 'nodecg', 'nodomain', 'nois', 'noke', 'nolan',
'nome', 'nomin', 'non', 'nonsens', 'noon', 'nope', 'normal', 'norman', 'north',
'northern', 'northwestern', 'norton', 'norway', 'nose', 'nostra', 'nosubdomain',
'note', 'notebook', 'notepad', 'noth', 'notic', 'notr', 'nova', 'novel',
'novemb', 'novic', 'nowher', 'nozzl', 'np', 'npo', 'nq', 'nr', 'nrhj', 'nriz',
'nrizw', 'nrizwt', 'ns', 'ns1', 'nsmca', 'nss', 'nssdca', 'nsw', 'nswc', 'nt',
'ntaib', 'nthu', 'ntis', 'ntreal', 'nu', 'nuclear', 'nucleus', 'nuetron',
'number', 'numer', 'nuntius', 'nut', 'nux', 'nuy', 'nv', 'nw', 'nwi', 'nwk',
'nwu', 'nx', 'ny', 'nyx', 'nz', 'o0', 'o0i', 'o1', 'o2', 'o3', 'o4', 'o4u',
'o5i', 'o5ih', 'o6', 'o7', 'o8', 'oa', 'oai', 'oak', 'oakland', 'oasys', 'ob',
'oba', 'obispo', 'object', 'oblig', 'obnox', 'oboe', 'obp', 'observ',
'observatori', 'observatory', 'obsolet', 'obtain', 'obvious', 'oc', 'occas',

'occasion', 'occult', 'occur', 'ocean', 'ocf', 'ocs', 'oct', 'octob', 'od',
'od9', 'odd', 'odi', 'odin', 'oe', 'ofa123', 'offend', 'offens', 'offer',
'offerman', 'offhand', 'offic', 'offici', 'offlin', 'offset', 'oft', 'og', 'oh',
'ohio', 'ohsu', 'oi', 'oil', 'oit', 'oj', 'ok', 'okay', 'okg', 'oklahoma',
'okz', 'ol', 'old', 'older', 'oldest', 'ole', 'olerud', 'oliv', 'oliveira',
'olivetti', 'olson', 'olymp', 'olympus', 'om', 'om4', 'omar', 'oml', 'onboard',
'ongo', 'onion', 'onlin', 'ontario', 'oo', 'ooo', 'oop', 'op', 'open', 'oper',
'opinion', 'oppon', 'opportun', 'oppos', 'opposit', 'opt', 'optic', 'optim',
'optimist', 'option', 'optiplan', 'oq', 'oqi', 'oracl', 'oracle', 'orang',
'orbit', 'ord', 'order', 'ordinari', 'oregon', 'org', 'organis', 'organiz',
'ori', 'orient', 'origin', 'original', 'orin', 'oriol', 'oriolefan', 'orion',
'orteig', 'orthogon', 'orvill', 'os', 'os2', 'os4l', 'os7', 'oscar', 'oscs',
'osf', 'oswego', 'ot', 'otago', 'otc', 'otehr', 'otherwis', 'ott', 'ottawa',
'otto', 'otv', 'ou', 'ouellett', 'ought', 'outcom', 'outer', 'outfield',
'outing', 'outlin', 'output', 'outrag', 'outsid', 'outstand', 'outta',
'outweigh', 'ov', 'ovei', 'overal', 'overburden', 'overcom', 'overhead',
'overlap', 'overr', 'overrun', 'overstress', 'overview', 'overweight',
'overwrit', 'ovg', 'ow', 'owen', 'owner', 'ox', 'oxid', 'oxygen', 'oy', 'oz',
'ozon', 'ozzi', 'p0', 'p1', 'p2', 'p3', 'p3o', 'p4', 'p4u', 'p5', 'p7', 'p8',
'p9', 'p_', 'pa', 'pa881a', 'pablo', 'pace', 'pacif', 'pack', 'packag',
'packard', 'packet', 'pad', 'padi', 'padr', 'pag', 'page', 'pagliarulo', 'paid',
'pain', 'paint', 'paladin', 'palmer', 'pam', 'pan', 'panason', 'panayiotaki',
'panel', 'panix', 'paparella', 'paper', 'paperback', 'par', 'parachut',
'paradis', 'paradox', 'paragraph', 'parallel', 'paramet', 'parent', 'pari',
'pariti', 'park', 'parker', 'parti', 'partial', 'particip', 'particl',
'particular', 'partner', 'pasadena', 'pascal', 'pass', 'passag', 'passeng',
'password', 'past', 'pat', 'patch', 'path', 'pathet', 'patienc', 'patrick',
'pattern', 'paul', 'paula', 'paus', 'pavo', 'pax', 'pay', 'payload', 'payment',
'pb', 'pbd', 'pc', 'pcf', 'pcplus', 'pct', 'pd', 'pdt', 'pdx', 'pe', 'peac',
'peach', 'peak', 'peanut', 'pedersen', 'pegasus', 'pen', 'pena', 'penalti',
'pendleton', 'penn', 'pennant', 'pennsylvania', 'pentium', 'peopel', 'peopl',
'percent', 'percentag', 'percept', 'perez', 'perfect', 'perform', 'perhap',
'peri', 'perige', 'perihelion', 'perijov', 'period', 'peripher', 'perman',
'permana', 'permanet', 'permiss', 'permit', 'perot', 'perpetu', 'perri',
'persist', 'person', 'personnel', 'perth', 'perturb', 'pet', 'pete', 'peter',
'pf', 'pfaffenwaldr', 'pg', 'pgf', 'pgt', 'ph', 'phantom', 'phase', 'phenomena',
'phi', 'phil', 'philadelphia', 'philip', 'philips', 'phill', 'philli',
'phillip', 'philosophi', 'phobo', 'phoenix', 'phone', 'photo', 'photograph',
'photomet', 'phrase', 'phys', 'physic', 'physicist', 'physics', 'pi', 'piazza',
'pick', 'pictur', 'piec', 'pif', 'pilot', 'pin', 'pinch', 'pioneer', 'pip',
'pipelin', 'pirat', 'piss', 'piston', 'pit', 'pitch', 'pitcher', 'pitt',
'pittsburgh', 'pixel', 'pj', 'pjtier01', 'pk', 'pl', 'pl4', 'pl5', 'pl6', 'pl8',
'pl9', 'place', 'plain', 'plan', 'plane', 'planet', 'planetari', 'planetarium',
'plant', 'plantier', 'plasma', 'plastic', 'plate', 'platform', 'platoon',
'plausibl', 'play', 'player', 'playoff', 'plaza', 'pleas', 'pleasant', 'plenti',
'plier', 'plot', 'plug', 'plus', 'pluto', 'plymouth', 'pm', 'pmf', 'pmf9f',
'pmf9f9f9l3', 'pmf9f9fq', 'pmf9l0q', 'pmf9l3', 'pmfq', 'pn', 'pne', 'pne1',
'pne1t', 'pnei', 'pnei4', 'po', 'pocket', 'point', 'pointer', 'poison', 'polar',

'pole', 'poli', 'polici', 'polish', 'polit', 'pollut', 'polymorph', 'polytechn',
'pomona', 'ponder', 'pool', 'poor', 'pop', 'popul', 'popular', 'port',
'portabl', 'portal', 'portion', 'portuges', 'pose', 'poseidon', 'posit',
'possibl', 'post', 'postscript', 'poster', 'posting', 'postion', 'postscript',
'potenti', 'pound', 'powel', 'power', 'pp', 'ppd', 'pq', 'pr', 'practic',
'prado', 'prael', 'prais', 'prb', 'pre', 'preced', 'precis', 'predict',
'prefer', 'prefix', 'preliminari', 'prepar', 'preseason', 'presenc', 'present',
'preserv', 'presid', 'press', 'pressur', 'prestig', 'presum', 'pretend',
'pretti', 'prevent', 'previous', 'price', 'pricetag', 'primari', 'primarili',
'prime', 'primit', 'princ', 'princeton', 'princip', 'principi', 'print',
'printer', 'prior', 'prioriti', 'prism', 'prison', 'privat', 'prize', 'prl',
'pro', 'prob', 'probabl', 'probe', 'problem', 'procedur', 'proceed', 'process',
'processor', 'procomm', 'prodigi', 'prodigy', 'produc', 'product', 'prof',
'profession', 'professor', 'profil', 'profit', 'progman', 'program', 'programm',
'progress', 'prohibit', 'project', 'promis', 'promot', 'prompt', 'prone',
'proof', 'propel', 'proper', 'properti', 'proport', 'propos', 'propuls',
'prospect', 'protect', 'protest', 'protocol', 'proton', 'prototyp', 'proud',
'prove', 'proven', 'provid', 'provo', 'prowess', 'ps', 'psc', 'psu', 'psuvm',
'psych', 'psycholog', 'pt', 'ptm', 'pu', 'pu1', 'pub', 'public', 'publish',
'puckett', 'puckey', 'pull', 'puls', 'pulsar', 'pump', 'punishments', 'purchas',
'purdu', 'purdue', 'pure', 'purpos', 'pursu', 'push', 'puzzl', 'pv', 'pw',
'pwe', 'pwiseman', 'pws', 'pwu', 'px', 'py', 'pyron', 'q0', 'q1', 'q145', 'q2',
'q3', 'q30t', 'q30tbxn', 'q32', 'q4', 'q5', 'q6', 'q7', 'q8', 'q9', 'q_', 'qa',
'qax', 'qb', 'qc', 'qd', 'qd9', 'qe', 'qg', 'qgq', 'qh', 'qhl', 'qi', 'qj',
'qk', 'ql', 'qm', 'qn', 'qo', 'qp', 'qq', 'qr', 'qr_', 'qrs', 'qs', 'qs6_',
'qs7', 'qt', 'qtct', 'qtcv9', 'qtm', 'qu', 'qualif', 'qualifi', 'qualit',
'qualiti', 'quantiti', 'quantum', 'quarter', 'qucdn', 'queen', 'queensu',
'queri', 'quest', 'question', 'quick', 'quicken', 'quicker', 'quiet', 'quit',
'quot', 'quotat', 'qv', 'qvei', 'qvfi', 'qvg', 'qw', 'qx', 'qy', 'qz', 'r0',
'r0d', 'r1', 'r186', 'r1865', 'r2', 'r3', 'r4', 'r5', 'r6', 'r7', 'r8', 'r8e',
'r8f', 'r_', 'ra', 'race', 'rack', 'radar', 'radiat', 'radic', 'radii', 'radio',
'radius', 'rafia', 'raider', 'rail', 'rain', 'rais', 'rak', 'ralli', 'ralph',
'ram', 'ramsey', 'ran', 'rand', 'randal', 'randi', 'random', 'rang', 'ranger',
'rank', 'rap', 'raphael', 'rapid', 'rare', 'rat', 'rate', 'ratio', 'rational',
'ratnam', 'rattl', 'rave', 'ravel', 'raw', 'ray', 'rb', 'rbd', 'rbi', 'rbis',
'rborden', 'rc', 'rck', 'rcs', 'rd', 'rdetweil', 'reach', 'reaction', 'reactor',
'read', 'readabl', 'reader', 'readi', 'readili', 'real', 'reali', 'realis',
'realiti', 'realiz', 'realli', 'really', 'reap', 'reardon', 'reason', 'reboot',
'rec', 'recal', 'receiv', 'recent', 'recogn', 'recommend', 'record', 'recov',
'recoveri', 'red', 'redesign', 'redirect', 'reduc', 'reduct', 'redwood', 'reed',
'reentri', 'refer', 'refin', 'reflect', 'refresh', 'refus', 'regain', 'regal',
'regard', 'regardless', 'regedit', 'reggi', 'region', 'regist', 'registr',
'regul', 'regular', 'rehabilit', 'reiken', 'reimert', 'reinstal', 'reject',
'rel', 'relat', 'relationship', 'relay', 'releas', 'relev', 'reli', 'reliabl',
'relief', 'reliev', 'reload', 'rem', 'remain', 'remaind', 'remark', 'remeb',
'rememb', 'remind', 'remot', 'remov', 'renam', 'render', 'rene', 'rent', 'rep',
'repair', 'repeat', 'replac', 'replay', 'repli', 'reply', 'repo', 'report',
'repost', 'repres', 'reproduct', 'republ', 'reput', 'request', 'requir',

'rescu', 'research', 'resembl', 'reserv', 'reset', 'resid', 'residu', 'resist',
'resiz', 'resolut', 'resolv', 'resolve', 'resort', 'resourc', 'respect',
'respond', 'respons', 'rest', 'restart', 'reston', 'restor', 'restrict',
'restructur', 'result', 'resuppli', 'retain', 'retir', 'retriev', 'return',
'reulbach', 'reusabl', 'reveal', 'revenu', 'revers', 'review', 'revis',
'revolt', 'revolut', 'reward', 'rewrit', 'rex', 'reynold', 'rf', 'rg', 'rh',
'rhea', 'rheolog', 'rhode', 'ri', 'rich', 'richard', 'richardson', 'richmond',
'rick', 'rickc', 'rickert', 'rickey', 'rid', 'riddl', 'ride', 'ridicul',
'rigel', 'right', 'righti', 'ring', 'rins', 'rintintin', 'rip', 'ripken',
'risc', 'rise', 'risk', 'riski', 'river', 'rivera', 'rj', 'rjn', 'rk', 'rl',
'rl9', 'rle', 'rlg', 'rlh', 'rli', 'rlk', 'rlk8', 'rls', 'rm', 'rma', 'rmohn',
'rn', 'rnichol', 'ro', 'road', 'rob', 'robb', 'robbi', 'robbin', 'robert',
'roberto', 'robin', 'robinson', 'robot', 'robust', 'rochest', 'rochester',
'rock', 'rocket', 'rocketri', 'rocki', 'rockwel', 'rockwell', 'rod', 'rodan',
'rodriguez', 'roger', 'rogoff', 'role', 'roll', 'roller', 'rollout', 'rom',
'roman', 'romant', 'ron', 'roof', 'rooki', 'room', 'root', 'rope', 'rose',
'rosen', 'rosie', 'ross', 'roster', 'rotat', 'roth', 'rotten', 'rough', 'round',
'roush', 'routin', 'rover', 'row', 'roy', 'royal', 'rp', 'rpi', 'rq', 'rr',
'rs', 'rsb', 'rsf', 'rt', 'rtf', 'rtg', 'rubber', 'rudi', 'ruin', 'rule',
'rumor', 'rumour', 'run', 'runner', 'runyon', 'rush', 'russ', 'russel',
'russia', 'russian', 'rutger', 'rutgers', 'ruth', 'rv', 'rvesterm', 'rw',
'rwang', 'rwtms2', 'rx', 'ry', 'ryan', 'ryn', 'ryne', 'ryukoku', 'rz', 's0',
's1', 's2', 's3', 's4', 's5', 's6', 's8', 's8v', 's9', 's_', 'sa', 'sabo',
'sabr', 'sac', 'sad', 'safe', 'safeti', 'saic', 'said', 'sail', 'saint', 'sake',
'sal', 'salari', 'sale', 'salmon', 'salt', 'salyut', 'sam', 'samba', 'sampl',
'san', 'sand', 'sandberg', 'sander', 'sandi', 'sandman', 'santa', 'santiago',
'santo', 'saratoga', 'sarcasm', 'sas', 'sat', 'satellit', 'satisfi', 'saturday',
'saturn', 'saucer', 'saundrsg', 'save', 'saw', 'sawyer', 'sax', 'say', 'sb',
'sbp002', 'sc', 'scale', 'scan', 'scanner', 'scare', 'scari', 'scatter', 'scd',
'scenario', 'sceneri', 'schaefer', 'schaumburg', 'schedul', 'scheme',
'schiewer', 'schlatter', 'schmidt', 'school', 'schwarzenegg', 'sci', 'scicom',
'scienc', 'science', 'scientif', 'scientist', 'scope', 'scor', 'score',
'scoreboard', 'scoreless', 'scorn', 'scott', 'scout', 'scrambl', 'scratch',
'screen', 'screw', 'script', 'scroll', 'scsi', 'scuglia', 'sd', 'sdcn', 'sdi',
'sdio', 'sdk', 'se', 'se1', 'sea', 'seal', 'sean', 'search', 'seas', 'season',
'seat', 'seattl', 'seaver', 'sec', 'second', 'secondari', 'secret', 'section',
'sector', 'secur', 'seek', 'seen', 'seg', 'segment', 'sei', 'select', 'self',
'sell', 'semi', 'senat', 'senator', 'send', 'sender', 'senior', 'sens',
'sensit', 'sensor', 'sent', 'sentenc', 'sep', 'separ', 'seper', 'sepinw',
'sepinwal', 'sept', 'septemb', 'sequenc', 'seri', 'serial', 'serv', 'server',
'servic', 'session', 'set', 'seti', 'settl', 'setup', 'seven', 'seventh',
'seventi', 'sever', 'sexual', 'sez', 'sf', 'sfasu', 'sfsu', 'sfsuvax1', 'sfu',
'sg', 'sgi', 'sh', 'shadow', 'shafer', 'shakala', 'shake', 'shakespear',
'shaki', 'shall', 'shame', 'shannon', 'shape', 'share', 'sharewar', 'sharp',
'shawn', 'shea', 'sheet', 'sheffield', 'shel', 'sheldev', 'shelf', 'shell',
'shelley', 'shelter', 'sherri', 'sherzer', 'shield', 'shift', 'shim', 'ship',
'shit', 'shiva', 'sho', 'shock', 'shoemaker', 'shoot', 'shop', 'shopper',
'shore', 'short', 'shorter', 'shortstop', 'shot', 'shoulder', 'showalt',

'shower', 'shown', 'shut', 'shutout', 'shuttl', 'shuttle', 'shz', 'si', 'sic',
'sick', 'sid', 'sierra', 'sig', 'sigh', 'sight', 'sigma', 'sign', 'signal',
'signatur', 'signific', 'sigurdsson', 'sii', 'silicon', 'silli', 'silver',
'sim', 'similar', 'simmon', 'simon', 'simpl', 'simpli', 'simplic', 'simplifi',
'simul', 'simultan', 'simultaneuo', 'sin', 'sinc', 'sincer', 'singer', 'singl',
'single', 'sink', 'sir', 'sit', 'site', 'situat', 'sixth', 'size', 'sj', 'sjl',
'sk', 'sk8', 'skeptic', 'ski', 'skidmor', 'skidmore', 'skill', 'skip', 'skndiv',
'sky', 'skylab', 'sl', 'slam', 'slaught', 'slc', 'sleep', 'slg', 'slgsun',
'slice', 'slick', 'slid', 'slide', 'slider', 'slight', 'slip', 'slot', 'slow',
'slower', 'slowli', 'slug', 'slugger', 'slump', 'sm', 'small', 'smaller',
'smart', 'smartdriv', 'smartdrv', 'smile', 'smiley', 'smith', 'smithsonian',
'smoke', 'smoltz', 'smooth', 'smsu', 'sn', 'snichol', 'snow', 'snydefj',
'snyder', 'soak', 'soc', 'social', 'societi', 'socket', 'soda', 'soft',
'softquad', 'softwar', 'soil', 'sol', 'solar', 'sold', 'sole', 'solid', 'solo',
'solomon', 'solut', 'solv', 'somebodi', 'someday', 'someon', 'someth',
'sometim', 'somewhat', 'somewher', 'son', 'song', 'soon', 'sooner', 'sophist',
'sorri', 'sort', 'sought', 'sound', 'soundblast', 'sourc', 'south', 'southern',
'southwestern', 'soviet', 'sox', 'soyuz', 'sp', 'spac', 'space', 'spacecraft',
'spaceflight', 'spacehab', 'spacelift', 'spacelink', 'spaceship', 'spacewalk',
'spacewatch', 'spacifi', 'span', 'spanish', 'spanki', 'spare', 'sparki',
'spatial', 'spe', 'speak', 'speaker', 'spec', 'special', 'specialist', 'specif',
'specifi', 'spectromet', 'spectrum', 'specul', 'speech', 'speed', 'speedisk',
'speedstar', 'spell', 'spencer', 'spend', 'spent', 'sphere', 'spheric', 'spi',
'spike', 'spin', 'spinoff', 'spira', 'spirit', 'spite', 'spitz', 'split',
'spoke', 'sponsor', 'sport', 'sportscast', 'sportswrit', 'sportwrit', 'spot',
'spread', 'spring', 'spss', 'sq', 'sq9', 'sqk', 'sql', 'sqm', 'sqmv', 'suar',
'sr', 'srb', 'srbs', 'srg', 'srl02', 'srl03', 'srm', 'ss', 'ss24x', 'ss8',
'ssa', 'sscl', 'ssd', 'ssf', 'ssi', 'ssme', 'ssrbs', 'ssrt', 'ssto', 'st',
'st902415', 'stabil', 'stabl', 'stack', 'stacker', 'stadium', 'staff', 'stag',
'stage', 'staion', 'stake', 'stan', 'stand', 'standard', 'stanford', 'stanki',
'stankiewicz', 'stankowitz', 'stanley', 'star', 'starfleet', 'starflight',
'stark', 'stars', 'start', 'starter', 'startl', 'startup', 'stat', 'state',
'statement', 'station', 'statist', 'status', 'staub', 'stay', 'std', 'stdvax',
'steal', 'stealth', 'steam', 'steel', 'steer', 'stefan', 'stein', 'steinly',
'steinman', 'steinn', 'stellar', 'stengel', 'step', 'steph', 'stephen',
'stephenson', 'steve', 'steven', 'stewart', 'stick', 'stiff', 'stock', 'stole',
'stolen', 'stomach', 'stone', 'stop', 'storag', 'store', 'stori', 'storm',
'str', 'straight', 'strang', 'strap', 'strategi', 'stratospher', 'strawberri',
'streak', 'stream', 'street', 'strength', 'stress', 'stretch', 'strict',
'stride', 'strike', 'strikeout', 'string', 'strip', 'strobe', 'stroke',
'strong', 'stronger', 'structur', 'struggl', 'sts', 'stsci', 'stuck', 'student',
'studi', 'stuff', 'stupid', 'stupidest', 'stuttgart', 'style', 'su', 'sub',
'subdirector', 'submiss', 'submit', 'suborbit', 'subscrib', 'subsequ',
'subset', 'subsidiari', 'substanc', 'substanti', 'substitut', 'subsystem',
'succeed', 'success', 'suck', 'sudden', 'sue', 'suffer', 'suffici', 'suggest',
'suit', 'suitabl', 'sum', 'summar', 'summari', 'summer', 'sun', 'sun121',
'sunb', 'sunday', 'suni', 'sunlight', 'sunnyval', 'suno', 'sunris', 'sunrise',
'sunset', 'super', 'superb', 'superior', 'supernova', 'supersed', 'superson',

'superstar', 'suppli', 'supplier', 'support', 'suppos', 'sure', 'surfac',
'surfeit', 'surgeri', 'surpass', 'surpris', 'surround', 'survey', 'surviv',
'susanna', 'susect', 'suspect', 'suspici', 'sustain', 'sutcliff', 'sv', 'svein',
'svga', 'svr3', 'svr4', 'sw', 'sw2', 'swap', 'swartz', 'swartzendrüb', 'swear',
'sweden', 'sweep', 'sweet', 'sweetpea', 'swell', 'swept', 'swetski', 'swh',
'swing', 'switch', 'swyatt', 'sx', 'sy', 'symantec', 'symbol', 'symptom', 'sys',
'sysedit', 'sysmgr', 'sz', 't0', 't1', 't2', 't3', 't3qs', 't3s', 't4', 't44',
't5', 't6', 't7', 't8', 't9', 't_', 'ta', 'tab', 'tabl', 'tag', 'tahiti',
'taib', 'tail', 'taiwan', 'taken', 'taker', 'tal', 'talent', 'talk', 'tall',
'tamu', 'tandem', 'tank', 'tanstaafl', 'tape', 'target', 'tartabul', 'task',
'tast', 'tate', 'tater', 'tattam', 'taurus', 'tax', 'taxpay', 'taylor', 'tb',
'tc', 'tcd', 'tcp', 'tct', 'td', 'td4i', 'te', 'teach', 'teacher', 'teal',
'team', 'teammät', 'tech', 'technic', 'techniqu', 'technolog', 'ted', 'tedward',
'teel', 'teh', 'tek2q', 'tel', 'telecom', 'telecommun', 'telecomwa',
'telemetri', 'telephon', 'telescop', 'telesoft', 'televis', 'television', 'tell',
'telnet', 'telo', 'temp', 'temperatur', 'temporari', 'temporarili', 'tend',
'tens', 'tent', 'term', 'termin', 'terrestri', 'terri', 'terriabl', 'terrifi',
'terror', 'tesla', 'test', 'testb', 'tether', 'tettleton', 'tew', 'texa',
'texan', 'text', 'tf', 'tffreeba', 'tg', 'th', 'thank', 'thanx', 'theme',
'theoret', 'theori', 'theporch', 'therefor', 'thermal', 'thf2', 'thier',
'thing', 'think', 'thiokol', 'thirti', 'thoma', 'thomas', 'thompson', 'thor',
'thorn', 'thorough', 'thought', 'thousand', 'thread', 'threat', 'threaten',
'threw', 'throw', 'thrown', 'thru', 'thrust', 'thruster', 'thu', 'thumb',
'thursday', 'thx', 'ti', 'ticket', 'tie', 'tierney', 'tiger', 'tight', 'til',
'till', 'tilt', 'tim', 'time', 'timer', 'times', 'timothi', 'tin', 'tini',
'tinman', 'tio2', 'tip', 'tire', 'titan', 'titl', 'tj', 'tk', 'tl', 'tm',
'tm0006', 'tn', 'tno', 'tobia', 'today', 'todd', 'togeth', 'token', 'told',
'toler', 'tom', 'tomassi', 'tomh', 'tomlin', 'tommi', 'tomorrow', 'ton',
'toner', 'toni', 'tonight', 'took', 'tool', 'toolkit', 'topaz', 'topic',
'toronto', 'torr', 'toss', 'total', 'touch', 'tough', 'tour', 'tout', 'tower',
'town', 'tp', 'tq', 'tr', 'track', 'trade', 'tradit', 'traffic', 'trail',
'train', 'trajectori', 'transfer', 'transform', 'transit', 'translat',
'transmiss', 'transmit', 'transmitt', 'transpond', 'transport', 'trap', 'trash',
'travel', 'traven', 'treat', 'tree', 'trek', 'tremend', 'trench', 'trend',
'tri', 'tribe', 'tribun', 'trick', 'trident', 'triniti', 'trintex', 'trip',
'tripl', 'tripoli', 'triumph', 'trivia', 'trivial', 'troop', 'troubl', 'true',
'truety', 'truli', 'trumpet', 'trust', 'truth', 'trw', 'ts', 'tsd', 'tsing',
'tss', 'tt', 'ttacs1', 'ttfont', 'ttq6', 'ttu', 'tu', 'tub', 'tuba', 'tube',
'tucson', 'tue', 'tuesday', 'tug', 'tune', 'tung', 'turbo', 'turn',
'turnaround', 'tv', 'tw', 'twg', 'twice', 'twilight', 'twin', 'twist', 'tx',
'txc', 'txt', 'ty', 'typ', 'type', 'typic', 'typo', 'tz', 'tzx', 'u0', 'u1',
'u2', 'u3', 'u34', 'u34u', 'u3l', 'u4', 'u45', 'u5', 'u6', 'u7', 'u8', 'u9',
'ua', 'uar', 'uart', 'uax', 'ub', 'ubc', 'uc', 'ucar', 'ucdavis', 'ucflvm',
'uchicago', 'uci', 'uco', 'ucs', 'ucsc', 'ucsd', 'ucsu', 'ud', 'udel', 'ue9',
'uf', 'ufo', 'ug', 'uga', 'ugli', 'ugliest', 'ugly', 'uh', 'ui', 'uic',
'uihepa', 'uit', 'uiuc', 'uj', 'ujq', 'uk', 'uky', 'ul', 'ulkyvx', 'ulowell',
'ultim', 'ultra', 'ultrastor', 'ultraviolet', 'ulyss', 'um', 'umass', 'umb',
'umbc', 'umd', 'ummm', 'umn', 'ump', 'umpir', 'unabl', 'unauthent', 'unawar',

'unbeliev', 'unc', 'uncertain', 'uncertainti', 'uncl', 'uncompress', 'und',
'underdog', 'underestim', 'underneth', 'understand', 'understood', 'unexpected',
'unfortun', 'uni', 'uniform', 'uninstal', 'union', 'uniqu', 'unit', 'unity',
'univ', 'univers', 'universiteit', 'unix', 'unixg', 'unknown', 'unless',
'unlik', 'unlimit', 'unlv', 'unm', 'unman', 'unreason', 'unstabl', 'unt',
'unus', 'unusu', 'uo', 'uoknor', 'upcom', 'updat', 'updatedir', 'upenn',
'upgrad', 'upload', 'upper', 'upset', 'uq', 'ur', 'uranium', 'uranus', 'urbana',
'urepli', 'usa', 'usaf', 'usag', 'usc', 'usd', 'use', 'useless', 'usenet',
'user', 'userag', 'usg', 'usl', 'usr', 'ussr', 'usual', 'usui', 'uswest', 'ut',
'utah', 'utexas', 'util', 'utoronto', 'utter', 'utzoo', 'uu', 'uucp', 'uudecod',
'uuencod', 'uug', 'uunet', 'uv', 'uvic', 'uw', 'uwaterloo', 'uwe', 'uwm', 'uwo',
'uwyo', 'ux', 'ux4', 'uxa', 'uy', 'uz', 'v0', 'v1', 'v2', 'v3', 'v4', 'v5',
'v6', 'v7', 'v8', 'v9', 'v9f', 'v9f9f', 'v9f9l3', 'v9fq', 'v_', 'va',
'vacation', 'vacuum', 'vagu', 'val', 'valentin', 'valid', 'valley', 'valu',
'valuabl', 'van', 'vanc', 'vancouv', 'vandal', 'vandenber', 'vanish', 'vapor',
'vari', 'variabl', 'variat', 'varieti', 'various', 'varmit', 'vast', 'vaughn',
'vax', 'vax1', 'vaxc', 'vb', 'vb30', 'vc', 'vcd', 'vcu', 'vd', 've', 'vega',
'veget', 'vehicl', 'vei', 'veloc', 'venari', 'vendor', 'vent', 'ventur',
'ventura', 'venus', 'ver', 'veraart', 'veri', 'verifi', 'vern', 'vers', 'versa',
'version', 'versus', 'vertic', 'vesa', 'vesterman', 'vet', 'veteran', 'vf',
'vf1', 'vg', 'vga', 'vgalogo', 'vh', 'vhf', 'vi', 'viamar', 'vice', 'vicin',
'victor', 'victori', 'video', 'view', 'viewer', 'vigil', 'vike', 'villag',
'villanueva', 'vinc', 'vincent', 'viola', 'violat', 'violent', 'violet',
'virginia', 'virtual', 'virtually', 'virus', 'visa', 'visibl', 'vision',
'visit', 'visual', 'vital', 'vizcaino', 'vizquel', 'vj', 'vk', 'vl', 'vladimir',
'vlb', 'vlbi', 'vm', 'vma', 'vmk', 'vms', 'vn', 'vnet', 'vnew', 'vo', 'voic',
'void', 'vol', 'volt', 'volum', 'volunt', 'von', 'vote', 'voyag', 'vp', 'vpnet',
'vq', 'vr', 'vram', 'vs', 'vt', 'vtz', 'vu', 'vulcan', 'vulner', 'vum', 'vv',
'vw', 'vw8', 'vx', 'vy', 'vz', 'w0', 'w1', 'w11', 'wit', 'w2', 'w3', 'w3q',
'w4', 'w44', 'w45', 'w47', 'w4w', 'w4wg', 'w5', 'w6', 'w7', 'w8', 'w9', 'w9v',
'w_', 'wa', 'wa4', 'wade', 'wager', 'wagner', 'wait', 'waiver', 'wale', 'wales',
'walford', 'walk', 'walker', 'wall', 'wallach', 'wallop', 'wallpap', 'walt',
'walter', 'wan', 'wang', 'want', 'war', 'ward', 'warfar', 'warm', 'warn',
'warner', 'warp', 'warren', 'warrent', 'wash', 'washington', 'wasn', 'wast',
'waste', 'wat', 'watch', 'water', 'waterloo', 'watson', 'watt', 'wav', 'wave',
'wavelength', 'wax', 'way', 'wayn', 'wb', 'wc', 'wd', 'wdwell', 'weak',
'weapon', 'wear', 'weather', 'weber', 'wed', 'wednesday', 'week', 'weekend',
'weigh', 'weight', 'weird', 'weiss', 'welcom', 'wellington', 'went', 'wes',
'west', 'western', 'wet', 'wetstein', 'wetteland', 'wf', 'wfan', 'wfw', 'wfw2',
'wfwg', 'wg', 'wh', 'whatev', 'whatsoev', 'whenev', 'wherea', 'whine', 'whitak',
'white', 'whiten', 'whoosh', 'wi', 'wickman', 'wide', 'width', 'wierd', 'wife',
'wig', 'wil', 'wilbur', 'wild', 'wildstrom', 'willi', 'william', 'wilson',
'wim', 'wimsey', 'wimvh', 'win', 'win16', 'win3', 'win31', 'win32', 'winadv',
'winbench', 'wind', 'window', 'windows', 'winfax', 'winfield', 'wing',
'wingless', 'wingo', 'winmark', 'winner', 'winqvt', 'winqvtnet', 'winsock',
'winter', 'wintrumpet', 'winword', 'wip', 'wipe', 'wir', 'wire', 'wisc',
'wisconsin', 'wisdom', 'wise', 'wish', 'wiskund', 'wit', 'wizard', 'wj', 'wk',
'wl', 'wlieftin', 'wm', 'wm4', 'wm4u', 'wmbxlt', 'wmbxn', 'wmw', 'wn',

```
'wnkretz', 'wo', 'woman', 'women', 'won', 'wonder', 'wong', 'woo', 'wood',
'woodruff', 'woof', 'worcest', 'word', 'worden', 'wordperfect', 'wordprocessor',
'work', 'workabl', 'workaround', 'worker', 'workgroup', 'workplac', 'workshift',
'workshop', 'workspac', 'workstat', 'world', 'worldwid', 'worri', 'wors',
'worst', 'worth', 'worthless', 'wouldn', 'wound', 'wouter', 'wow', 'wp', 'wpi',
'wq', 'wrap', 'wrench', 'wri', 'wright', 'wrigley', 'wrist', 'write', 'writer',
'written', 'wrj', 'wrksft16', 'wrong', 'wrote', 'ws', 'wspace', 'wspdpsf', 'wt',
'wu', 'wuarhive', 'wustl', 'wv', 'ww', 'ww2', 'ww7', 'wwhj', 'wwhjn',
'wwhjnux', 'wwhjnuy', 'wwiz', 'wwizw', 'wx', 'wy', 'wyatt', 'wz', 'x0', 'x1',
'x11', 'x2', 'x3', 'x4', 'x5', 'x6', 'x7', 'x8', 'x8z', 'x9', 'x9i', 'x_', 'xa',
'xb', 'xc', 'xd', 'xe', 'xenon', 'xg', 'xh', 'xi', 'xj', 'xk', 'xl', 'xm', 'xn',
'xp', 'xq', 'xr', 'xrcjd', 'xs', 'xssr', 'xt', 'xu', 'xv', 'xw', 'xx', 'xxdate',
'xxi', 'xxmessage', 'xy', 'y0', 'y1', 'y2', 'y24', 'y3', 'y4', 'y5', 'y5v',
'y6', 'y7', 'y8', 'y9', 'y9s', 'y_', 'ya', 'yale', 'yamauchi', 'yank', 'yanke',
'yankee', 'yard', 'yb', 'yc', 'yd', 'yeah', 'year', 'yee', 'yell', 'yellow',
'yep', 'yer', 'yes', 'yesterday', 'yf', 'yf9', 'yf9f3t', 'yf9f9d', 'yf9f9f9',
'yg', 'yh', 'yi', 'yield', 'yj', 'yl', 'ym', 'yn', 'yo', 'yogi', 'yom', 'york',
'yorku', 'young', 'younger', 'youngster', 'yount', 'yp', 'yq', 'yr', 'ys', 'yt',
'yu', 'yuggoth', 'yup', 'yv', 'yw', 'yx', 'yy', 'yz', 'z0', 'z1', 'z2', 'z3',
'z4', 'z5', 'z6', 'z6e', 'z6e1', 'z6e1t', 'z6ei', 'z6ei4', 'z7', 'z72', 'z8',
'z9', 'z_', 'za', 'zai', 'zane', 'zb', 'zbh', 'zc', 'zd', 'zd9', 'zealand',
'zeil', 'zero', 'zeta', 'zeus', 'zf', 'zg', 'zgb0', 'zh', 'zi', 'zikzak', 'zip',
'zj', 'zl', 'zm', 'zn', 'znb', 'zng', 'znh', 'znk', 'znkjz', 'znl', 'zo',
'zone', 'zoo', 'zoolog', 'zp', 'zq', 'zr', 'zrck', 'zri', 'zrlk', 'zs', 'zt',
'zu', 'zv', 'zw', 'zx', 'zy', 'zz']
```

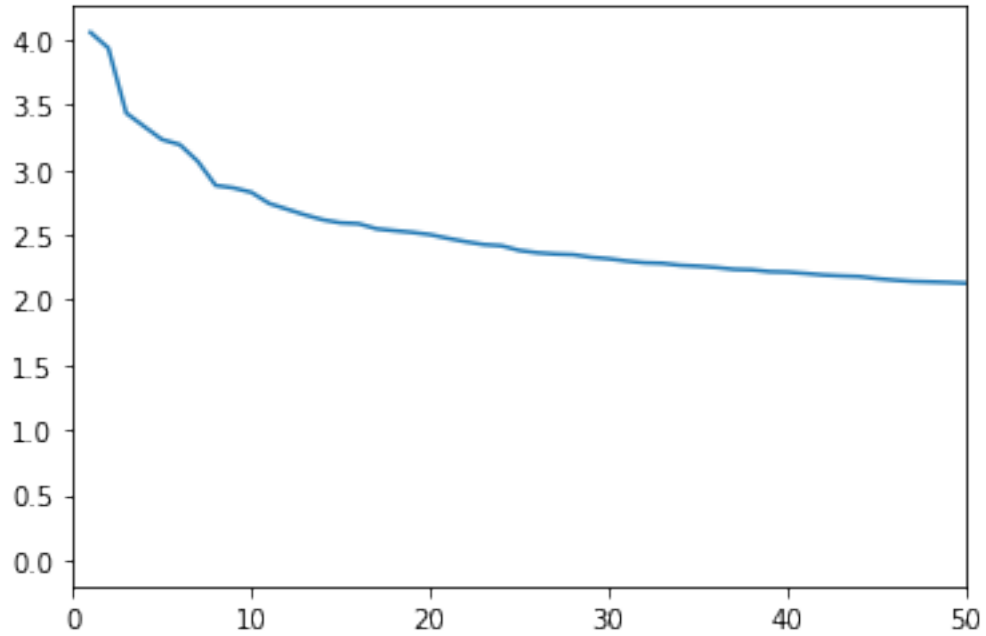
```
[98]: dtm_dense = dtm.todense()
centered_dtm = dtm_dense - np.mean(dtm_dense, axis=0)
np.sum(centered_dtm,axis=0)[:,:10]
```

```
[98]: matrix([[ 9.68843061e-16, -5.26054894e-15,  2.83258660e-15,
 5.48389459e-16,  4.95209342e-15,  2.94935517e-15,
 1.01882478e-15, -6.34345007e-15,  1.35613092e-14,
 6.73457600e-15]])
```

```
[99]: u, s, vt = np.linalg.svd(centered_dtm)
```

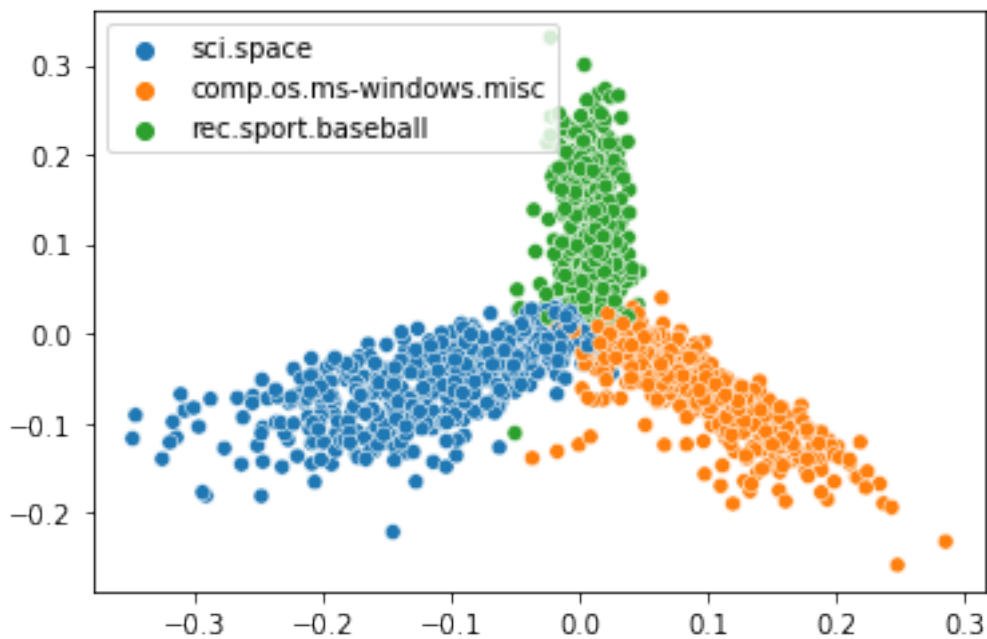
```
[100]: plt.xlim([0,50])
plt.plot(range(1,len(s)+1),s)
```

```
[100]: [<matplotlib.lines.Line2D at 0x16e3713c310>]
```



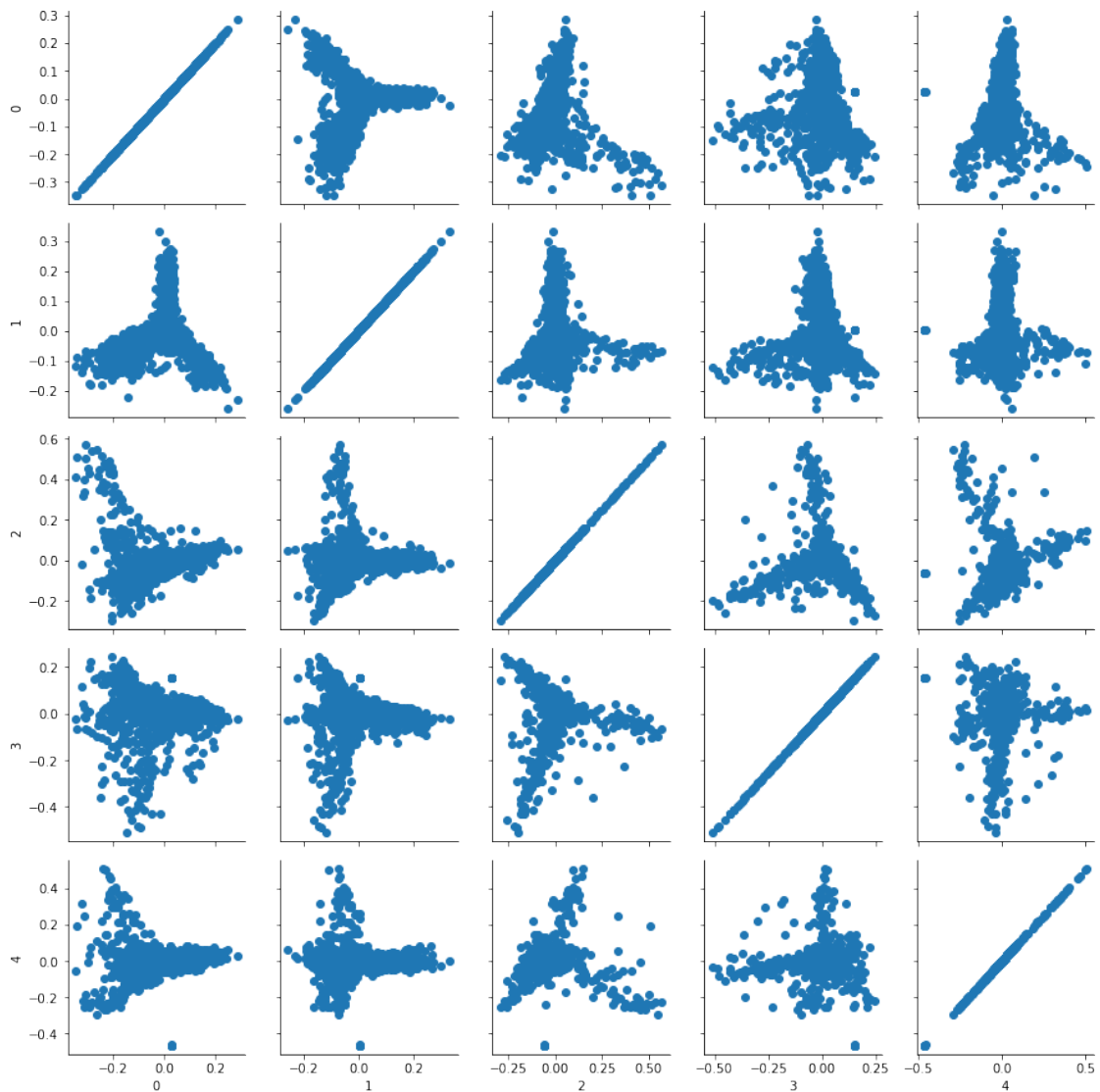
```
[101]: k=2
vectorsk = np.array(u[:, :k] @ np.diag(s[:k]))
labels = [news_data.target_names[i] for i in news_data.target]
sns.scatterplot(x=vectorsk[:,0], y=vectorsk[:, 1], hue=labels)
```

[101]: <AxesSubplot:>



```
[102]: import seaborn as sns
k = 5
Xk = u[:, :k] @ np.diag(s[:k])
X_df = pd.DataFrame(Xk)
g = sns.PairGrid(X_df)
g.map(plt.scatter)
```

[102]: <seaborn.axisgrid.PairGrid at 0x16e370a9d60>



```
[104]: terms = vectorizer.get_feature_names()
for i in range(6):
    top = np.argsort(vt[i])
    topterms = [terms[top[0,f]] for f in range(12)]
    print (i, topterms)

0 ['space', 'henry', 'nasa', 'alaska', 'toronto', 'gov', 'moon', 'zoo',
'aurora', 'spencer', 'nsmca', 'pat']
1 ['windows', 'space', 'access', 'nasa', 'file', 'dos', 'gov', 'digex', 'files',
'use', 'henry', 'program']
2 ['access', 'nasa', 'digex', 'gov', 'pat', 'jpl', 'space', 'baalke', '___',
'com', 'kelvin', '__']
3 ['access', 'digex', 'pat', 'com', 'prb', 'net', 'express', 'online',
'communications', 'dseg', 'usa', 'ti']
4 ['ax', 'henry', 'toronto', 'zoo', 'spencer', 'zoology', 'jpl', 'baalke',
'___', 'com', 'gov', 'kelvin']
5 ['nasa', 'gov', 'henry', 'jpl', 'baalke', 'toronto', '___', 'windows',
'kelvin', '__', 'ca', 'team']
```

```
[ ]:
```