

# Probabilistic Models Based on the $\Pi$ -Sigmoid Distribution

Anastasios Alivanoglou and Aristidis Likas

Department of Computer Science, University of Ioannina  
GR 45110, Ioannina, Greece  
{aalivano, arly}@cs.uoi.gr

**Abstract.** Mixture models constitute a popular type of probabilistic neural networks which model the density of a dataset using a convex combination of statistical distributions, with the Gaussian distribution being the one most commonly used. In this work we propose a new probability density function, called the  $\Pi$ -sigmoid, from its ability to form the shape of the letter “ $\Pi$ ” by appropriately combining two sigmoid functions. We demonstrate its modeling properties and the different shapes that can take for particular values of its parameters. We then present the  $\Pi$ -sigmoid mixture model and propose a maximum likelihood estimation method to estimate the parameters of such a mixture model using the Generalized Expectation Maximization algorithm. We assess the performance of the proposed method using synthetic datasets and also on image segmentation and illustrate its superiority over Gaussian mixture models.

**Keywords:** Probabilistic neural networks, mixture models,  $\Pi$ -sigmoid distribution, orthogonal clustering, image segmentation.

## 1 Introduction

Gaussian mixture models (GMM) are a valuable statistical tool for modeling densities. They are flexible enough to approximate any given density with high accuracy, and in addition they can be interpreted as a soft clustering solution. Thus, they have been widely used in both supervised and unsupervised learning, and have been extensively studied, e.g. [3]. They can be trained through a convenient EM procedure [4] that yields maximum likelihood estimates for the parameters of the mixture. However, it exhibits some weaknesses, the most notable being its lack of interpretability, since it provides spherical (or ellipsoidal in the most general case) shaped clusters that are inherently hard to be understood by humans. It is widely acknowledged that humans prefer solutions in the form of *rectangular shaped* clusters which are directly interpretable. Another weakness of the GMM approach is that it is not efficient when used to model data that are uniformly distributed in some regions.

With the aim to adequately treat the above issues, in this work we propose a new probability distribution called the  $\Pi$ -sigmoid ( $\Pi$ s) distribution. This distribution is obtained as the difference of two translated sigmoid functions and is flexible enough

to approximate data distributions ranging from Gaussian to uniform depending on the slope of the sigmoids. We also propose a mixture model with  $\Pi$ -sigmoid distributions called  $\Pi$ -sigmoid mixture model ( $\Pi$ sMM) and show that it is capable of providing probabilistic clustering solutions for the case of rectangular-shaped clusters that are straightforward to transform into an interpretable set of rules. We present a maximum likelihood technique to estimate the parameters of  $\Pi$ sMM using the Generalized EM algorithm [3,4]. As experimental results indicate, due to its flexibility to approximate both the rectangular uniform and bell-shaped distributions, the  $\Pi$ sMM provides superior solutions compared to GMMs when the data are not Gaussian.

## 2 The $\Pi$ -Sigmoid Distribution

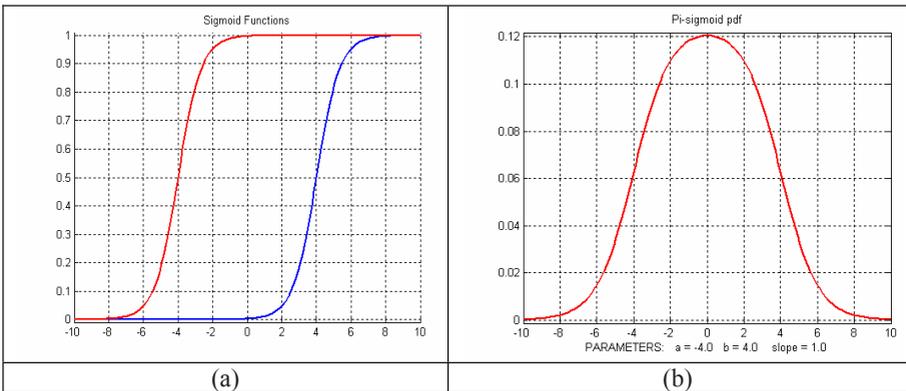
The one-dimensional  $\Pi$ -sigmoid distribution is computed as the difference between two logistic sigmoid functions with the same slope. The logistic sigmoid with slope  $\lambda$  is given by:

$$\sigma(x) = \frac{1}{1 + e^{-\lambda x}} \tag{1}$$

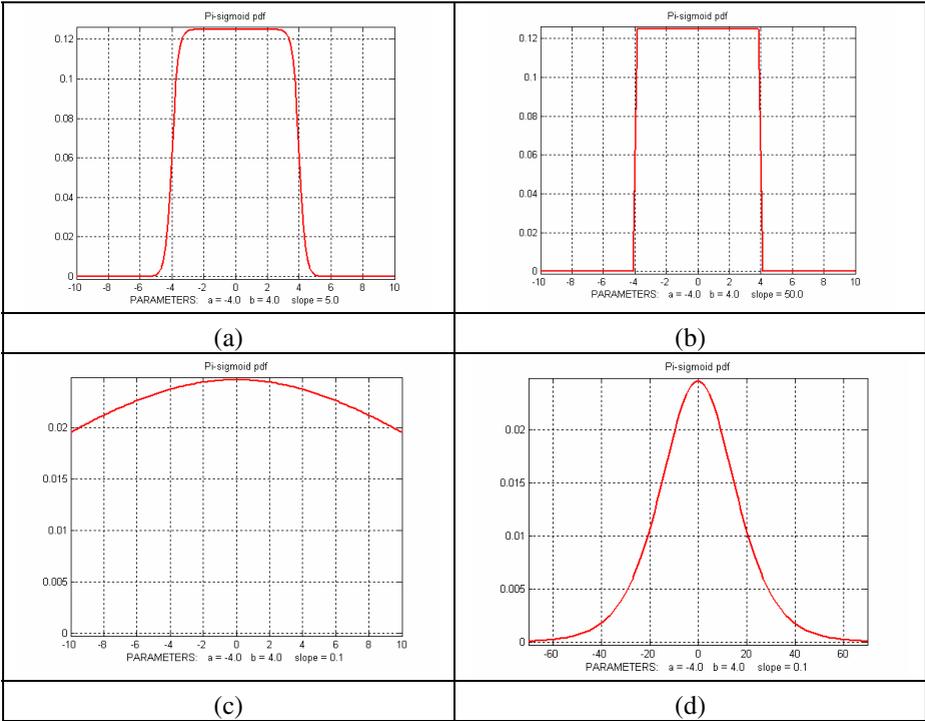
The  $\Pi$ -sigmoid pdf with parameters  $a, b, \lambda$  (with  $b > a$ ) is defined by subtracting two translated sigmoids:

$$\Pi_s(x) = \left( \frac{1}{b - a} \right) \left[ \frac{1}{1 + e^{-\lambda(x-a)}} - \frac{1}{1 + e^{-\lambda(x-b)}} \right], \quad b > a, \quad \lambda > 0 \tag{2}$$

The term  $1/(b-a)$  is the normalization constant to ensure that the integral of  $\Pi_s(x)$  is unit. It is interesting to note that the integral  $\int [\sigma(x-a) - \sigma(x-b)] dx = b - a$  independently of  $\lambda$ . Figure 1 describes two translated sigmoids and the resulting  $\Pi$ -sigmoid distribution.



**Fig. 1.** The one-dimensional  $\Pi$ -sigmoid distribution (b) obtained as the difference of two translated sigmoid functions (a)



**Fig. 2.** Several shapes of the  $\Pi$ -sigmoid distribution by varying the values of its parameters

In Fig.2 we present several shapes of the  $\Pi$ -sigmoid distribution by varying the values of its parameters. It is clear that both uniform (Fig. 2b) and bell-shaped distributed (Fig. 2d) can be adequately approximated.

The *multidimensional  $\Pi$ -sigmoid distribution* is obtained under the assumption of independence along each dimension. More specifically, for a vector  $x=(x_1, x_2, \dots, x_D)^T$ :

$$\Pi_S(x) = \prod_{d=1}^D \Pi_S(x_d) = \prod_{d=1}^D \frac{1}{1 + e^{-\lambda_d(x_d - a_d)}} \frac{1}{1 + e^{-\lambda_d(x_d - b_d)}} \frac{1}{b_d - a_d} \tag{3}$$

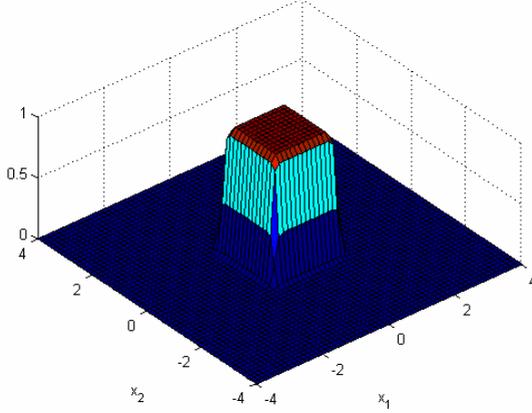
with  $b_d > a_d$  and  $\lambda_d > 0$ . Fig. 3 illustrates how a two-dimensional  $\Pi$ -sigmoid distribution that approximates a uniform distribution on rectangular domain.

**2.1 Maximum Likelihood Estimation**

Suppose we are given a dataset  $X = \{x^1, \dots, x^N\}$ ,  $x^i \in R^D$ , to be modeled by a  $\Pi$ -sigmoid distribution. The parameters of the distribution can be estimated by maximizing the likelihood of the dataset  $X$ :

$$L(X; A, B, \Lambda) = \sum_{i=1}^N \log \Pi s(x^i) = \sum_{i=1}^N \sum_{d=1}^D \log \left( \frac{\frac{1}{1 + e^{-\lambda_d(x_d^i - a_d)}} - \frac{1}{1 + e^{-\lambda_d(x_d^i - b_d)}}}{b_d - a_d} \right) \quad (4)$$

with respect to the set of parameters  $A = \{a_d\}, B = \{b_d\}, \Lambda = \{\lambda_d\}, d=1, \dots, D$ .



**Fig. 3.** A two-dimensional  $\Pi$ -sigmoid probability density function

The maximum likelihood solution cannot be obtained in closed form. However, since the gradient of the likelihood with respect to the parameters can be computed, gradient-based maximization methods can be employed (for example the simple gradient ascent or the more sophisticated quasi-Newton methods such as the BFGS).

### 3 The $\Pi$ -Sigmoid Mixture Model

Using the proposed  $\Pi$ -sigmoid distribution, a mixture model can be defined called  $\Pi$ sMM ( $\Pi$ -sigmoid Mixture Model) can be defined as follows

$$p(x) = \sum_{k=1}^K \pi_k \Pi s(x; A_k, B_k, \Lambda_k) \quad (5)$$

where  $K$  is the number of  $\Pi$ -sigmoid components,  $A_k = \{a_{kd}\}, B_k = \{b_{kd}\}, \Lambda_k = \{\lambda_{kd}\}$  are the parameters of  $k$ -th component and the mixing weights  $\pi_k$  satisfy the constraints:

$$\pi_k \geq 0, \sum_{k=1}^K \pi_k = 1.$$

Given a dataset  $X = \{x^1, \dots, x^N\}$ ,  $x^i \in R^D$  the parameters of the  $\Pi$ sMM can be estimated through maximum likelihood using the EM algorithm as is the also the case with GMMs. The EM algorithm is an iterative approach involving two steps at each iteration. The *E-step* is the same in all mixture models and computes the posterior probability that  $x^i$  belongs to component  $k$ :

$$P(k | x^i) = \frac{\pi_k \Pi s(x^i; A_k, B_k, \Lambda_k)}{\sum_{j=1}^K \pi_j \Pi s(x^i; A_j, B_j, \Lambda_j)} \quad (6)$$

The *M-step* requires the maximization of the complete likelihood  $L_c$  (eq. (7)) with respect to the parameters of the  $\Pi$ sMM model.

$$L_c = \sum_{i=1}^N \sum_{k=1}^K P(k | x^i) \log[\pi_k \Pi s(x^i; A_k, B_k, \Lambda_k)] \quad (7)$$

For the parameters  $\pi_k$  the update equation is the same for all mixture models:

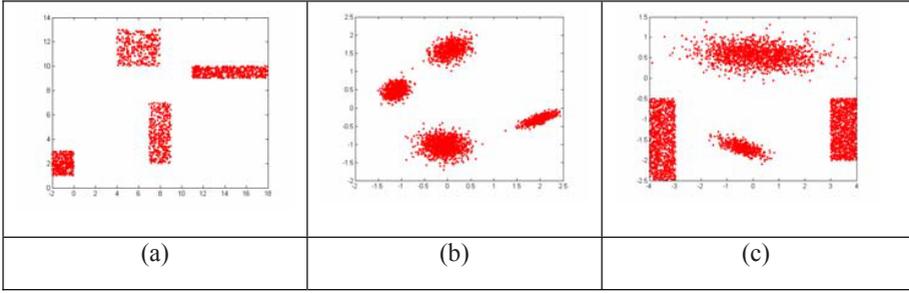
$$\pi_k = \frac{1}{N} \sum_{i=1}^N P(k | x^i) \quad (8)$$

In contrast to the GMM case, the *M-step* does not lead to closed form update equations for the parameters  $A_k, B_k, \Lambda_k$  of the  $\Pi$ -sigmoid components. Thus we resort to the GEM (generalized EM) algorithm [3,4], which suggests to update the model parameters so that obtain higher (not necessarily maximum) values of the complete likelihood are obtained. In this work a few updates of each parameter  $\theta$  along the direction of the gradient  $\partial L_c / \partial \theta$  are computed. The GEM algorithm ensures that, at each iteration, the parameters  $\theta$  are updated so that the likelihood increases, until a local maximum is reached.

It is well known that the EM algorithm is very sensitive to the initialization of the model parameters. To deal with this issue, we first apply the Greedy-EM algorithm [2] to fit a GMM with  $K$  components (and diagonal covariance matrices) on the dataset. Then the parameters of the  $k$ -th  $\Pi$ sMM component are initialized using the obtained parameters (mean and variance) of the  $k$ -th GMM component. The values of  $\lambda$  are always initialized to 1.5, although the performance is not sensitive to this choice.

## 4 Experimental Results

We have compared the modeling capabilities of  $\Pi$ sMM against the widely used GMM. First artificial datasets were considered with a) only uniform rectangular, (Fig. 4a) b) only Gaussian (Fig. 4b) and c) mixed (uniform and Gaussian) clusters (Fig. 4c). We considered the case  $D=2$ ,  $K=4$  (as in Fig. 4) and  $D=5$ ,  $K=4$ . Each dataset contained 5000 data points used for training and 5000 used for testing. It must be emphasized that we are not interested in data clustering, but in building accurate models of the density of the given datasets. In this spirit, the obtained mixture models



**Fig. 4.** Three types of artificial datasets: (a) rectangular uniform, (b) Gaussian, (c) mixed

**Table 1.** Test set likelihood values for three types of datasets (Gaussian, uniform, mixed) for the cases with  $D=2, K=4$  and  $D=5, K=4$

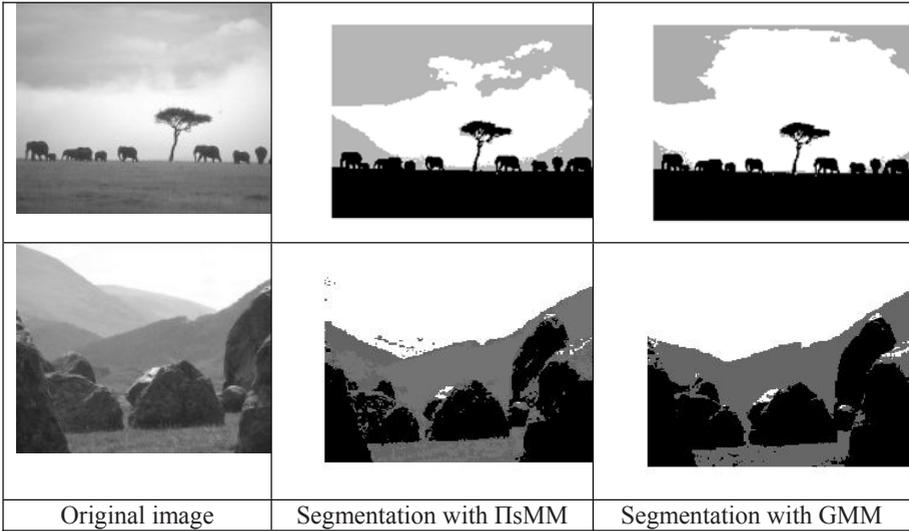
	D=2, K=4			D=5, K=4		
	Uniform	Gaussian	Mixed	Uniform	Gaussian	Mixed
$\Pi$ sMM	-8572	3802	-11558	-10167	-8168	-46616
GMM	-11118	3836	-12653	-8856	-8352	-49044

were compared in terms of the likelihood on the test set which constitutes the most reliable measure to compare statistical models.

The results are summarized in Table 1. In the case of uniform clusters the  $\Pi$ sMM was highly superior to GMM, in the case of Gaussian clusters the GMM was only slightly superior, while in the case of mixed clusters the performance of  $\Pi$ sMM was much better. The inability of GMMs to efficiently approximate the uniform distribution is an old and well-known problem and the proposed  $\Pi$ sMM model provides an efficient solution to modeling uniformly distributed data, while at the same time is able to adequately model Gaussian data.

Since GMMs have been successfully employed for image modeling and segmentation using pixel intensity information, we tested the performance of the proposed  $\Pi$ sMM model on the image segmentation task and compare against GMM. We considered  $256 \times 256$  grey level images and for each image the intensities of 5000 randomly selected pixels were used as the training set to fit a GMM and a  $\Pi$ sMM with the same number of components ( $K=5$ ). After training, the remaining pixels were assigned to the component (cluster) with the maximum posterior probability (eq (8)). The segmentation results for two example images are shown in Fig. 5.

It must be noted that the likelihood values on the large set of pixels not used for training (test set) of the  $\Pi$ sMM and GMM were -28229 and -30185, respectively, for the top image and -23487 and -26576 for the bottom image of Fig. 5, also indicating that the  $\Pi$ sMM provides superior statistical models of the images compared to GMM. This superiority is also confirmed from visual inspection of the segmented images (which is more clear in the top row images).



**Fig. 5.** Segmentation results for two natural images (first column) using  $\Pi$ sMM (second column) and GMM (third column)

## 5 Conclusions

We have proposed a new probability density function ( $\Pi$ -sigmoid) defined as the difference of two translated logistic sigmoids. Depending on the slope value of the sigmoids, the shape of the distribution may vary from bell-shaped to uniform allowing the flexibility to model a variety of datasets from Gaussian to uniform. We have also presented the  $\Pi$ -sigmoid mixture model ( $\Pi$ sMM) and show how to estimate its parameters under the maximum likelihood framework using the Generalized EM algorithm. Experimental comparison with the Gaussian Mixture Models indicate that  $\Pi$ sMM is more flexible than GMM providing solutions of higher likelihood. Also a notable characteristic of  $\Pi$ sMM is its ability to accurately identify rectangular shaped clusters, which constitutes a well-known weakness of GMMs.

It must be noted that another probabilistic model capable of identifying rectangular shaped clusters has been proposed in [1], called mixture of rectangles. In that model a component distribution is a uniform distribution with a Gaussian tail and it is difficult to train, due to the inability to define the gradient of the likelihood with respect to the parameters. Thus one has to resort to line search optimization methods to perform training of the model.

Our current work is focused on extending the  $\Pi$ -sigmoid distribution with the aim to describe rotated rectangles. The main issue to be addressed is how to develop an efficient training algorithm to adjust the additional parameters defining the rotation matrix. Another important issue is to develop a methodology to estimate the number of mixture components in  $\Pi$ sMM, based on recent methods developed in the context of GMMs.

*Acknowledgement.* Information dissemination of this work was supported by the European Union in the framework of the project “Support of Computer Science Studies in the University of Ioannina” of the “Operational Program for Education and Initial Vocational Training” of the 3rd Community Support Framework of the Hellenic Ministry of Education, funded by national sources and by the European Social Fund (ESF).

## References

- [1] Pelleg, D., Moore, A.: Mixture of rectangles: Interpretable soft clustering. In: Proc. ICML (2001)
- [2] Vlassis, N., Likas, A.: A greedy EM algorithm for Gaussian mixture learning. *Neural Processing Letters* 15, 77–87 (2002)
- [3] McLachlan, J.G., Krishnan, T.: *Finite Mixture Models*. Wiley, Chichester (2000)
- [4] McLachlan, J.G., Krishnan, T.: *The EM algorithm and extensions*. Marcel Dekker, New York (1997)