Deficiencies of the Multivariate Normal Model [4]

Qi Zhao

August 3, 2018

1. Introduction

There is a generative approach to face detection; it will model the probability of the data x and parameterize this by the world state w. It will describe the data with a multivariate normal distribution so that or treating the two possible values of the state w separately.

Unfortunately, this model does not detect faces reliably. While this model achieves above-chance performance, it doesn't come close to producing a state-of-the-art result. This is hardly surprising: the success of this classifier hinges [1] on fitting the data with a normal distribution [5]. Unfortunately, this fit is poor for three reasons (Figure 1).
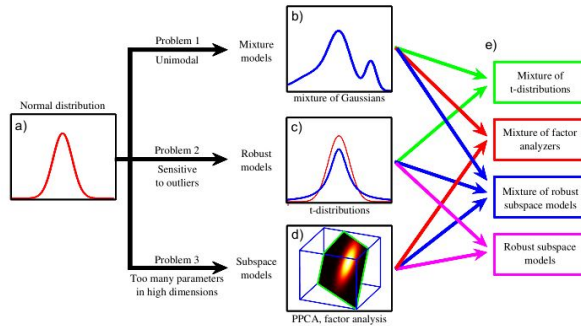

Figure 1. a) Problems with the multivariate normal density. b) Normal models are unimodal, but mixtures of Gaussians can model multi-modal distributions. c) Normal distributions are not robust to outliers, but t-distributions can cope with unusual observations. d) Normal models need many parameters in high dimensions but subspace models reduce this requirement. e) These solutions can be combined to form hybrid models addressing several of these problems at once.

1.The normal distribution is unimodal; neither faces nor background regionscare well represented by a pdf with a single peak.

2.The normal distribution is not robust; a single outlier can dramatically affect the estimates of the mean and covariance.

3.The normal distribution has too many parameters; here the data have D = 10800 dimensions. The full covariance matrix contains $\frac{D(D+1)}{2}$ parameters. With only 1000 training examples, it cannot even specify these parameters uniquely, so it was forced to use the diagonal form.

2. Conclusions

There are some ways to tackling these problems. To make the density multi-modal, it introduces mixture models [3]. To make the density robust, it replaces the normal with the t-distribution [2]. To cope with parameter estimation in high dimensions, it introduces subspace models [6].

The new models have much in common with each other. In each case it introduces a hidden or latent variable h_i associated with each observed data point x_i. The hidden variable induces the more complex properties of the resulting pdf.

References

[1] S. S. Chen and F. W. Wheaton. Oyster hinge line detection using image processing. *Aquacultural Engineering*, 8(5):307–327, 1989. 1

[2] E. A. Coornish. The multivariate t-distribution associated with a set of normal sample deviates. *Australian Journal of Physics*, 7(4):531–542, 1954. 1

[3] M. A. T. Figueiredo and A. K. Jain. Unsupervised learning of finite mixture models. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 24(3):381–396, 2000. 1

[4] L. T. Skovgaard. A riemannian geometry of the multivariate normal model. *Scandinavian Journal of Statistics*, 11(4):211–223, 1984. 1

[5] C. M. Stein. Estimation of the mean of a multivariate normal distribution. *Annals of Statistics*, 9(6):1135–1151, 1981. 1

[6] M. Verhaegen and D. Westwick. Identifying MIMO wiener systems using subspace model identification methods. In *IEEE Conference on Decision and Control*, pages 4206–4211, 1995. 1