

# Combining Models

Qi Zhao

August 25, 2018

There are some reasons it can get the mixture of Gaussians [3], t-distribution [5], and factor analysis [1] models are constructed similarly: each is a weighted sum or integral of a set of constituent normal distributions [2]. Mixture of Gaussian models consist of a weighted sum of  $K$  normal distributions with different means and variances. The t-distribution consists of an infinite weighted sum of normal distributions with the same mean, but different covariances. Factor analysis models consist of an infinite weighted sum of normal distributions with different means, but the same diagonal covariance. Figure 1 shows the parameters of a factor analysis model fitted to the face data using ten iterations of the EM algorithm [4]. The different factors encode different modes of variation of the data set which often have real-world interpretations such as changes in pose or lighting.

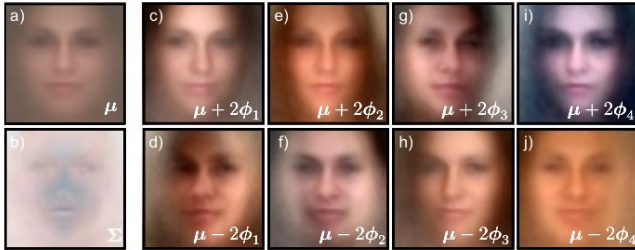


Figure 1. Factor analyzer with ten factors (four shown) for face classes. a) Mean  $\mu$  for face model. b) Diagonal covariance component  $\Sigma$  for face model. To visualize the effect of the first factor  $\phi_1$  we add (c) or subtract (d) a multiple of it from the mean: move along one axis of the 10D subspace that seems to encode mainly the mean intensity. Other factors (e-j) encode changes in the hue and the pose of the face.

There is a weighted sum of factor analyzers, each of which has a different mean and allocates high probability density to a different subspace. So it combines mixture models and factor analyzers, we get a mixture of factor analyzers (MoFA) model [6]. Similarly, combining mixture models and t-distributions results in a mixture of t-distributions or robust mixture model. Combining t-distributions and factor analyzers, it can construct a robust

subspace model, which models data that lie primarily in a subspace but is tolerant to outliers. The associated density function is in Equation 1.

$$Pr(x) = \sum_{k=1}^K \lambda_k Stud_x[\mu_k, \Phi_k \Phi_k^T + \Sigma_k, \nu_k] \quad (1)$$

where  $\mu_k$ ,  $\Phi_k$  and  $\Sigma_k$  represent the mean, factors, and diagonal covariance matrix belonging to the  $k^{th}$  component,  $\lambda_k$  represents the weighting of the  $k^{th}$  component and  $\nu_k$  represents the degrees of freedom of the  $k^{th}$  component.

## References

- [1] H. Akaike. Factor analysis and AIC. *Psychometrika*, 52(3):317–332, 1987. 1
- [2] D. F. Andrews and C. L. Mallows. Scale mixtures of normal distributions. *Journal of the Royal Statistical Society*, 36(1):99–102, 1974. 1
- [3] S. Dasgupta. Learning mixture of Gaussians. In *IEEE Symposium on Foundations of Computer Science*, pages 634–644, 1999. 1
- [4] DebashisKushary. The EM algorithm and extensions. *Technometrics*, 40(3):260–266, 1997. 1
- [5] M. C. Jones and M. J. Faddy. A skew extension of the t-distribution, with applications. *Journal of the Royal Statistical Society*, 65(1):159–174, 2010. 1
- [6] G. J. McLachlan, D. Peel, and R. W. Bean. Modelling high-dimensional data by mixtures of factor analyzers. *Computational Statistics & Data Analysis*, 41(3):379–388, 2003. 1