Factor Analysis

Qi Zhao

August 23, 2018

1. Introduction

As for the mixtures of Gaussians [2] and the t-distribution [3], it is possible to view the factor analysis model as a marginalization [1] of a joint distribution between the observed data x and a K-dimensional hidden variable h. It defines in Equation 1.

$$
\begin{aligned}
Pr(x|h) &= Norm_x[\mu + \Phi h, \Sigma] \\
Pr(h) &= Norm_h[O, I]
\end{aligned} \tag{1}
$$

where I represents the identity matrix.

Expressing factor analysis as a marginalization reveals a simple method to draw samples from the distribution. It first draws a hidden variable h from the normal prior. It then draws the sample x from a normal distribution [4] with mean μ + Φh and diagonal covariance Σ.

2. Conclusions

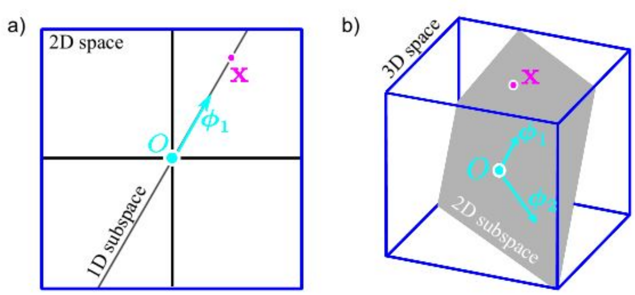

Figure 1. Linear subspaces a) A one dimensional subspace (a line through the origin, O) is embedded in a two dimensional space. Any point x in the subspace can be reached by weighting the single basis vector ϕ_1 appropriately. b) A two dimensional subspace (a plane through the origin, O) is embedded in a three dimensional space. Any point x in the subspace can be reached using a linear combination x = $\alpha\phi_1$ + $\beta\phi_2$ of the two basis functions ϕ_1, ϕ_2 that describe the subspace. In general a K-dimensional subspace can be described using K basis functions.

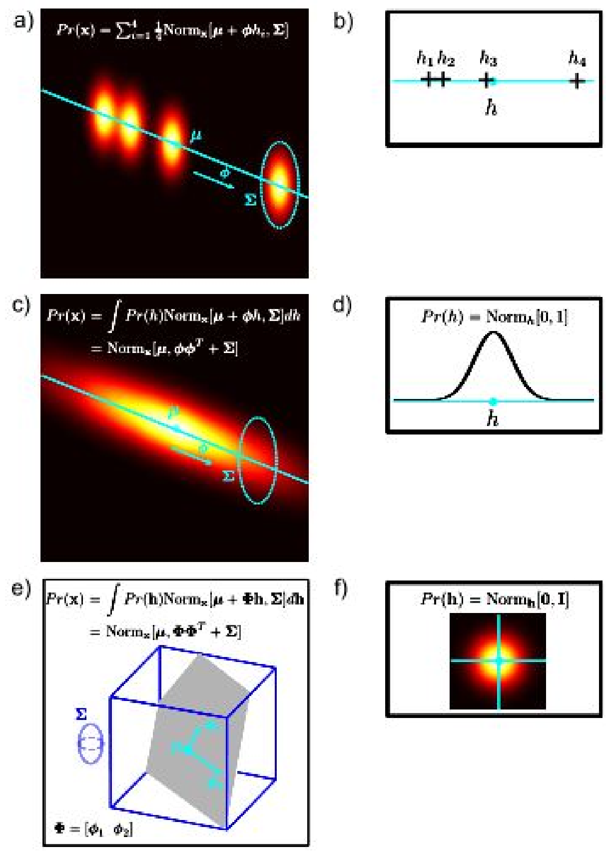

Figure 2. Relationship between factor analysis and mixtures of Gaussians (MoG). a) Consider a MoG model where each component has identical diagonal covariance Σ. We could describe variation in a particular direction ϕ by parameterizing the mean of each Gaussian as μ_i = μ+ϕh_i. b) Different values of the scalar hidden variable h i determine different positions along direction ϕ. c) Now we replace the MoG with an infinite sum (integral) over a continuous family of Gaussians, each of which is determined by a certain value of h. d) If we choose the prior over the hidden variable to be normal, then this integral has a closed form solution and is a factor analyzer. e) More generally we want to describe variance in a set of directions Φ = [ϕ_1, ϕ_2,..., ϕ_k] in a high dimensional space. f) To this end we use a K-dimensional hidden variable h and an associated normal prior Pr(h).

This leads to a simple interpretation of the hidden variable h: each element h_k weights the associated basis function φ_k in the matrix ϕ and hence defines a point on the subspace (Figure 1). The final density is hence an infinite weighted sum of normal distributions with the same diagonal covariance σ and means $\mu + \phi$h that are distributed over the subspace. The relationship between mixture models and factor analysis is explored further in Figure 2.

References

[1] E. J. Calabrese and L. A. Baldwin. The marginalization of hormesis. *Toxicologic Pathology*, 27(2):187, 1999. 1

[2] S. Dasgupta. Learning mixture of Gaussians. In *IEEE Symposium on Foundations of Computer Science*, pages 634–644, 1999. 1

[3] H. O. Hartley and E. S. Pearson. Table of the probability integral of the t-distribution. *Biometrika*, 37(1):168–172, 1950. 1

[4] W. Hoeffding. A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics*, 19(3):293–325, 1948. 1