Factor Analysis

Qi Zhao

August 21, 2018

1. Introduction

Now address the final problem with the normal distribution [2]. Visual data are often very high dimensional; in the face detection [4] task the data comes in the form of 60 X 60 RGB images and is hence characterized as a 60 X 60 X 3 = 10800 dimensional vector. To model this data with the full multivariate normal distribution, it requires a covariance matrix [1] of dimensions 10800 X 10800: it would need a very large number of training examples to get good estimates of all of these parameters in the absence of prior information. Furthermore, to store the covariance matrix it will need a large amount of memory, and there remains the problem of inverting this large matrix when it evaluates the normal likelihood.

2. Descriptions

If it just use the diagonal form of the covariance matrix which contains only 10800 parameters. However, this is too great a simplification: for face images, each dimension of the data is not independent. For example, in the cheek region, the RGB values of neighboring pixels covary very closely. A good model should capture this information. Factor analysis provides a compromise in which the covariance matrix is structured so that it contains fewer unknown parameters than the full matrix but more than the diagonal form. One way to think about the covariance of a factor analyzer is that it models part of the high-dimensional space with a full model and mops up remaining variation with a diagonal model.

More precisely, the factor analyzer describes a linear subspace with a full covariance model. A linear subspace is a subset of a high dimensional space that can be reached by taking linear combinations (weighted sums) of a fixed set of basis functions (Figure 1). So, a line through the origin is a subspace in two dimensions as it can reach any point on it by weighting a single basis vector. A line through the origin is also a subspace in three dimensions, but so is a plane through the origin: it can reach any point on the plane by taking linear combinations of two basis vectors. In general, a D-dimensional space contains subspaces of dimensions 1,2,...,D - 1.
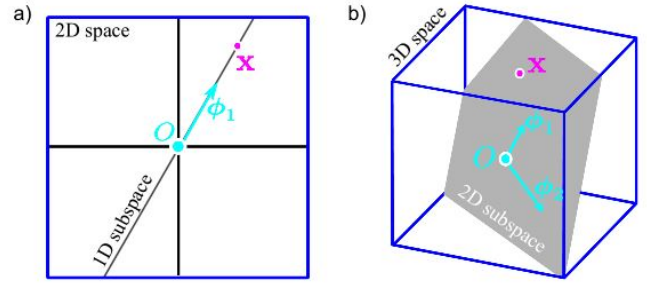

Figure 1. Linear subspaces a) A one dimensional subspace (a line through the origin, O) is embedded in a two dimensional space. Any point x in the subspace can be reached by weighting the single basis vector ϕ_1 appropriately. b) A two dimensional subspace (a plane through the origin, O) is embedded 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.

The probability density function of a factor analyzer is given in Equation 1.

$$Pr(x) = Norm_x[\mu, \Phi\Phi^T + \varepsilon] \tag{1}$$

where the covariance matrix $\Phi\Phi^T + \varepsilon$ contains a sum of two terms. The first term $\Phi\Phi^T$ describes a full covariance model over the subspace: the K columns of the portrait[1] rectangular matrix $\phi = [\phi_1, \phi_1,..., \phi_k]$ are termed factors. The factors are basis vectors that determine the subspace modeled. When it fits the model to data the factors will span the set of directions where the data covary the most. The second term σ is a diagonal matrix that accounts for all remaining variation.

Notice that this model has K X D parameters to describe Φ and another D parameters to describe the diagonal matrix Σ. If the number of factors K is much less than the dimensionality of the data D then this model has fewer parameters than a normal with full covariance and hence can be learned from fewer training examples.

When Σ is a constant multiple of the identity matrix (i.e., models spherical covariance) the model is called probabilistic principal component analysis. This simpler model has slightly fewer parameters and can be fit in closed form (i.e., without the need for the EM algorithm [3]), but otherwise it has no advantages over factor analysis.

References

[1] M. R. Ahmad. On testing sphericity and identity of a covariance matrix with large dimensions. *Mathematical Methods of Statistics*, 25(2):121–132, 2016. 1

[2] A. Azzalini and A. Capitanio. Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society*, 61(3):579–602, 2010. 1

[3] DebashisKushary. The EM algorithm and extensions. *Technometrics*, 40(3):260–266, 1997. 2

[4] E. Osuna, R. Freund, and F. Girosi. Training support vector machines: an application to face detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 130–136, 2002. 1