

CS28010 Homework 3

Yajing Chen

November 8, 2017

1 Factor analysis

1.1 Linear factor analysis

We denote the observed data as x , the latent factor as y and the error as ϵ . Suppose $y \sim \mathcal{N}(y|\mu, \Lambda)$, $\epsilon \sim \mathcal{N}(\epsilon|0, \Sigma)$ and $x = Ay + \epsilon$, where A is an $n \times m$ matrix, n is the dimension of x , m is the dimension of y and $m < n$. Please explain why there is more than one solution that satisfy $E(xx^T) = A\Lambda A^T + \Sigma$. When Σ is not a general positive definite matrix, but a diagonal matrix, how many solution exists? And if $\Sigma = \sigma^2 \mathcal{I}$, how many solution exists?

1.2 Binary factor analysis

If y is a latent factor where each dimension is an independent variable that subjects to a different Bernoulli distribution, what are the answers to the above three questions?

2 Projection

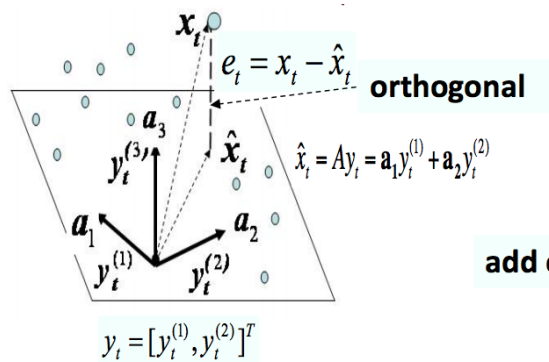
2.1 Orthogonal projection

Suppose we have a hyperplane whose orthogonal basis are $\alpha_1, \alpha_2, \dots, \alpha_k$, $k < n$. Now we have a n -dimensional vector x and we want to apply an orthogonal projection on the hyperplane. Please compute the corresponding projection matrix P .

3 Clustering

3.1 Comparison between Gaussian mixture model and k-means

Please add constraints to Gaussian mixture model so that it degenerates into k-means algorithm.



4 Optional summary work

Note: this is an optional homework. Please compare PCA, FA and ICA.

5 Homework submission

Deadline is **Nov 15th, 2017 24:00**

Please submit your homework in *pdf* format. Rename the file as *id_name_hw3.pdf*. For example, 015033910032_chenyajing_hw3.pdf. Upload the pdf file to the following ftp:

ftp://public.sjtu.edu.cn

username: cyj907

password: public