Mulitvariate analysis 364-2-1121

Spring 2019

Assignment 3 (Submission until 16.6.19

1. Assume that 200 observations were sampled from a Multi-Normal Distribution, with $p = 3$.
   The correlation matrix R was computed from the sample and it equals
   $$R = \begin{pmatrix} 1 & .34 & .26 \\ \cdot & 1 & .12 \\ \cdot & \cdot & 1 \end{pmatrix}$$
   (a) Estiamte the loads and the specific variances using PFA method
   (b) Estimate the loads and the specific variances using MLE method

   Note: You have to apply the procedure yourselves, not to use build in packeges.

2. Consider the following 5 points in the plain
   $$\begin{pmatrix} 1.9 \\ 0.64 \end{pmatrix}, \begin{pmatrix} 0.87 \\ -1.2 \end{pmatrix}, \begin{pmatrix} 0.3 \\ 0.09 \end{pmatrix}, \begin{pmatrix} -2.08 \\ -0.45 \end{pmatrix}, \begin{pmatrix} 0.85 \\ 0.46 \end{pmatrix}$$
   We would like to cluster these 5 points such that the measure $aSSW + bSSB + ck$ is minimized. For computing, use the distance fucntion $d(x, y) = (x - y)^T (x - y)$.

There are 52 partitions. Of course, we don't want to check all of them, and hence we will work stepwise. That is, we start with the partition of each point is a cluster and then try to merge, with the distance between clusters is define is deistance between their means.

3. Consider a sample $X_1, X_2, ...X_n \sim Gamma(3, \theta)$. That is,

$$f_Y(y) = \frac{y^2 e^{-\theta y} \theta^3}{2}, \; F(y) = 1 - \left(1 + \theta y + \frac{\theta^2 y^2}{2}\right) e^{-\theta y}, \quad y > 0$$

We assume the method of censored data. That is, there are known constants $t_1, ...t_n$, such that we observe only $Y_i = \min\{X_i, t_i\}$. In the excel file there are two columns. The first is the observed value Y and second equals 0 is the observation was censored and equals 1 otherwise. Also, Sort the data such that the uncensored observations are the first K observations.

(a) Write the partial likelihood (the one that relies on Y_i) that we saw in class. Use a numerical maximization method or a solver to maximize this likelihood.

(b) Show that

$$\mathrm{E}(Y|Y > t) = \frac{\int\limits_T^\infty x f(x) dx}{1 - F(x)} = \frac{1}{\theta} \frac{6 + 6\theta t + 3(\theta t)^2 + (\theta t)^3}{2 + 2\theta t + (\theta t)^2}$$

(c) Write one iteration of the EM algorithm

(d) Execute the EM algorithm and compare your result with one you got in the first item