

CS28010 Homework 2

Yajing Chen

November 1, 2017

1 Minimizing error pt2

1.1 Principal component analysis

Suppose we have N data points $\mathbf{x}_i, i = 1, \dots, N$, where each \mathbf{x}_i is a d -dimensional vector $\mathbf{x}_i = [x_{i,1}, x_{i,2}, \dots, x_{i,d}]^T$. You are required find a single line that best represents these N points. We assume $\mu = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i = \mathbf{0}$. And we use \mathbf{w} to denote the direction of the line and $\|\mathbf{w}\| = 1$. Please find \mathbf{w} using the data. (Knowledge about eigen vector, matrix derivatives might be required to finish this problem)

2 Factor analysis

2.1 Distribution of observed data

Suppose we observe N data points $\mathbf{x}_i, i = 1, \dots, N$, where each \mathbf{x}_i is a d -dimensional vector $\mathbf{x}_i = [x_{i,1}, x_{i,2}, \dots, x_{i,d}]^T$. In order to explain the inner relationship of these data, we specify some factors $\mathbf{y}_i, i = 1, \dots, N$, where each \mathbf{y}_i is a m -dimensional vector $\mathbf{y}_i = [y_{i,1}, y_{i,2}, \dots, y_{i,m}]^T$. We denote the linear relationship between \mathbf{x}_i and \mathbf{y}_i to be $\mathbf{x}_i = \mathbf{A}\mathbf{y}_i + \epsilon_i$, where $\mathbf{A}_{d \times m}$ is a matrix and ϵ_i is the error term. Suppose $\mathbf{y} \sim \mathcal{N}(\mu, \mathbf{\Lambda})$, $\epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$, $E(\mathbf{y}\epsilon^T) = \mathbf{0}$. Please compute the marginal distribution of observed data $q(\mathbf{x})$ and the conditional distribution $q(\mathbf{x}|\mathbf{y})$.

3 Optional summary work

Note: this is an optional homework. Please use your words to explain in the settings above, how many matrix \mathbf{A} satisfy the condition $\mathbf{x} = \mathbf{A}\mathbf{y} + \epsilon$.

4 Homework submission

Deadline is **Nov 8st, 2017 24:00**

Please submit your homework in *pdf* format. Rename the file as *id_name_hw2.pdf*. For example, 015033910032_chenyajing_hw2.pdf. Upload the pdf file to the following ftp:

ftp://public.sjtu.edu.cn

username: cyj907

password: public