CS28010 Homework 2

Guoxin SUI

November 2, 2017

1 Minimizing error pt2

1.1 Principal component analysis

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

The length of the projection of x_i onto the line is given by $x_i^T w$. The line that best repsesents these N points maximizes the variance of the projections.

So, we look for the w to maximize :

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}({x_i}^T w)^2 &= \frac{1}{N}\sum_{i=1}^{N}(w^T x_i {x_i}^T w) \\ &= w^T \left(\frac{1}{N}\sum_{i=1}^{N}(x_i {x_i}^T)\right) w\end{aligned}$$

Here we define $\Sigma = \frac{1}{N}\sum_{i=1}^{N}(x_i {x_i}^T)$, whici is just the covariance matrix of the data since $\frac{1}{N}\sum_{i=1}^{N} x_i = 0$. Intuitively Σ is symmetric. Then we can describe the problem as

$$max_{w \subset R^N} w^T \Sigma w \quad \text{s.t.} \quad ||w|| = 1$$

We form the Lagrangian as :

$$\mathcal{L}(w, \lambda) = w^T \Sigma w - \lambda w^T w$$

At the optimal point, the gradient of the Lagrangian has to be zero :

$$\nabla_w \mathcal{L}(w, \lambda) = \nabla_w (w^T \Sigma w - \lambda w^T w) = 2\Sigma^T x - 2\lambda x = 0$$

This gives $\Sigma x = \lambda x$, which shows that w is a eigenvector of Σ. To reach the maximazation, w should be the principal eigenvector of Σ, which corresponds to the biggest eigenvalue.

2 Factor analysis

2.1 Distribution of observed data

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

From $\begin{cases} x_i = Ay_i + \epsilon_i \\ y \sim \mathcal{N}(\mu, \Lambda) \\ \epsilon \sim \mathcal{N}(0, \sigma^2 I) \end{cases}$ we know that the conditional distribution $x|y \sim \mathcal{N}(\mu y, \sigma^2 I)$.

The random variables x and y have a joint Gaussian distribution as

$$\begin{bmatrix} y \\ x \end{bmatrix} \sim \mathcal{N}(\mu_{yx}, \Sigma)$$

We have

$$E(y) = \mu$$

and

$$\begin{aligned} E(x) &= E(Ay + \epsilon) \\ &= AE(y) + E(\epsilon) \\ &= A\mu \end{aligned}$$

Then we get

$$\mu_{yx} = \begin{bmatrix} \mu \\ A\mu \end{bmatrix}$$

Define $\Sigma = \begin{bmatrix} \Sigma_{yy}, \Sigma_{yx} \\ \Sigma_{xy}, \Sigma_{xx} \end{bmatrix}$, then we have

$$\begin{aligned}
\Sigma_{yy} &= \Lambda \\
\Sigma_{yx} &= E[(y-\mu)(Ay+\epsilon-A\mu)^T] \\
&= E(yy^TA^T + y\epsilon^T - y\mu^TA^T - \mu y^TA^T - \mu\epsilon^T + \mu\mu^TA^T) \\
&= E(yy^TA^T) + E(y\epsilon^T) - E(y)\mu^TA^T - \mu E(y^T)A^T - \mu E(\epsilon^T) + \mu\mu^TA^T \\
&= (\mu\mu^T + \sigma^2 I)A^T + 0 - \mu\mu^TA^T - \mu\mu^TA^T - 0 + \mu\mu^TA^T \\
&= \sigma^2 A^T \\
\Sigma_{xy} &= \Sigma_{yx}^T \\
&= \sigma^2 A \\
\Sigma_{xx} &= E[(Ay+\epsilon-A\mu)(Ay+\epsilon-A\mu)^T] \\
&= E(Ayy^TA^T + Ay\epsilon^T - Ay\mu^TA^T + \epsilon y^TA^T + \epsilon\epsilon^T - \epsilon\mu^TA^T - A\mu y^TA^T - A\mu\epsilon^T + A\mu\mu^TA^T) \\
&= AE(yy^T)A^T + AE(y\epsilon^T) - AE(y)\mu^TA^T + E(\epsilon y^T)A^T + E(\epsilon\epsilon^T) - E(\epsilon)\mu^TA^T - A\mu E(y^T)A^T - A\mu E(\epsilon^T) + A\mu\mu^T A \\
&= A(\mu\mu^T + \sigma^2 I)A^T + 0 + 0 + \sigma^2 I + 0 - A\mu\mu^TA^T - A\mu\mu^TA^T - 0 + A\mu\mu^TA^T \\
&= \sigma^2 AA^T + \sigma^2 I
\end{aligned}$$

Putting everything together we have :

$$\begin{bmatrix} y \\ x \end{bmatrix} \sim \mathcal{N}(\begin{bmatrix} \mu \\ A\mu \end{bmatrix}, \begin{bmatrix} \Lambda, \sigma^2 A^T \\ \sigma^2 A, \sigma^2 AA^T + \sigma^2 I \end{bmatrix})$$

Hence, the marginal distribution of x is given by:

$$x \sim \mathcal{N}(A\mu, \sigma^2 AA^T + \sigma^2 I)$$

3 Optional summary work

Please use your words to explain in the settings above, how many matrix A satisfy the condition $x = Ay + \epsilon$.
Since y and ϵ are actually latent variables, there are infinite possibilites for the value of y and ϵ. So we have infinite possiblites of matrix A that satisty the condition.