Self-Assessment Homework 03

r02b48003 Ko, Kuei-Yueh

April 17, 2017

1 Question 01

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ x_{d1} & x_{d2} & x_{d3} & \dots & x_{dn} \end{bmatrix}$$

$$T = XW + E$$

$$T = \begin{bmatrix} t_1^T \\ t_2^T \\ \vdots \\ t_n^T \end{bmatrix} = \begin{bmatrix} t'_1 \dots t'_n \end{bmatrix}$$

$$W = \begin{bmatrix} w_1 \dots w_n \end{bmatrix}$$

$$E = \begin{bmatrix} e_1 \dots e_n \end{bmatrix}$$

$E_D(w)$
$= \frac{1}{2} \operatorname{Tr} \left[(XW - T)^T (XW - T) \right]$
$= \frac{1}{2} \operatorname{Tr} \left[E^T E \right]$
$= \frac{1}{2} \sum_{i=1}^{k} e_i^T e_i$
$= \frac{1}{2} \sum_{i=1}^{k} (Xw_i - t'_i)^T (Xw_i - t'_i)$

$\frac{\partial E_D(w)}{\partial w^T}$
$= \left[\frac{\partial E_D(w)}{\partial w_1} \dots \frac{\partial E_D(w)}{\partial w_k} \right]$
$= \left[\frac{\partial (Xw_1 - t'_1)^T (Xw_1 - t'_1)}{\partial w_1} \dots \frac{\partial (Xw_k - t'_k)^T (Xw_k - t'_k)}{\partial w_k} \right]$

Each item is a partial derivative of each residual sum of square in linear regression. By letting $\frac{\partial E_D(w)}{\partial w^T} = 0$,

$$\left[\frac{\partial (Xw_1 - t'_1)^T (Xw_1 - t'_1)}{\partial w_1} = 0 \dots \frac{\partial (Xw_k - t'_k)^T (Xw_k - t'_k)}{\partial w_k} = 0 \right]$$

we can solve for each residual, which is exactly the same for simple linear regression.

$$\left[w_1 = (X^T X)^{-1} X^T t_1' \dots w_k = (X^T X)^{-1} X^T t_k'\right]$$

By combining the results, we can acquire the formula as follow

$$W = \left[w_1 \dots w_k\right] = (X^T X)^{-1} X^T T$$

2 Question 02

Using (4.57) and (4.58), derive the result (4.65) for the posterior class probability in teh two-class generative model with Gaussian densities, and verify the results (4.66) and (4.67) for the parmetres w and w_0

- 4.57 $p(C_1|x) = \frac{p(x|C_1)p(C_1)}{p(x|C_1)p(C_1)+p(x|C_2)p(C_2)} = \frac{1}{1+exp(-a)} = \sigma(a)$
- 4.58 a = $\ln \frac{p(x|C_1)p(C_1)}{p(x|C_2)p(C_2)}$
- 4.65 $p(C_1|x) = \sigma(w^T x + w_0)$
- 4.66 $w = \Sigma^{-1}(\mu_1 - \mu_2)$
- 4.67 $w_0 = -\frac{1}{2}\mu_1^T \Sigma^{-1} \mu_1^T + \frac{1}{2}\mu_2^T \Sigma^{-1} \mu_2^T + \ln \frac{p(C_1)}{p(C_2)}$

If a, b are vectors and A is a matrix, we know that $a^T A b$ is a scalar and therefore $a^T A b = b^T A a$

Based on the Gaussian density function

$$p(x|C_k) = (2\pi)^{-\frac{D}{2}} |\Sigma|^{-1} \exp(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu))$$

we can derive the following equations:
$p(x|C_1) = (2\pi)^{-\frac{D}{2}} |\Sigma|^{-1} \exp(-\frac{1}{2}(x-\mu_1)^T \Sigma^{-1}(x-\mu_1))$
$p(x|C_1) = (2\pi)^{-\frac{D}{2}} |\Sigma|^{-1} \exp(-\frac{1}{2}(x-\mu_2)^T \Sigma^{-1}(x-\mu_2))$

To calculate $p(C_1|x)$, we will first derive a
$p(C_1|x) = \sigma(a)$ where $a = \ln \frac{p(x|C_1)p(C_1)}{p(x|C_2)p(C_2)}$
$\ln \frac{p(x|C_1)p(C_1)}{p(x|C_2)p(C_2)} = \ln \frac{p(x|C_1)}{p(x|C_2)} + \ln \frac{p(C_1)}{p(C_2)}$
$\ln \frac{p(x|C_1)}{p(x|C_2)}$
$= \ln \frac{(2\pi)^{-\frac{D}{2}} |\Sigma|^{-1} \exp(-\frac{1}{2}(x-\mu_1)^T \Sigma^{-1}(x-\mu_1))}{(2\pi)^{-\frac{D}{2}} |\Sigma|^{-1} \exp(-\frac{1}{2}(x-\mu_2)^T \Sigma^{-1}(x-\mu_2))}$
$= \ln \frac{\exp(-\frac{1}{2}(x-\mu_1)^T \Sigma^{-1}(x-\mu_1))}{\exp(-\frac{1}{2}(x-\mu_2)^T \Sigma^{-1}(x-\mu_2))}$
$= -\frac{1}{2}(x-\mu_1)^T \Sigma^{-1}(x-\mu_1) + \frac{1}{2}(x-\mu_2)^T \Sigma^{-1}(x-\mu_2)$

$= \frac{1}{2}(2x^T\Sigma^{-1}(\mu_1 - \mu_2) - \mu_1^T\Sigma^{-1}\mu_1 + \mu_2^T\Sigma^{-1}\mu_2)$
$= x^T\Sigma^{-1}(\mu_1 - \mu_2) - \mu_1^T\Sigma^{-1}\mu_1 + \mu_2^T\Sigma^{-1}\mu_2$
$= (\Sigma^{-1}(\mu_1 - \mu_2))^T x - \mu_1^T\Sigma^{-1}\mu_1 + \mu_2^T\Sigma^{-1}\mu_2$

From the equation above,

$$p(x|C_1) = \sigma(a)$$

where

$$a = w^T x + w_0 = (\Sigma^{-1}(\mu_1 - \mu_2))^T x - \mu_1^T\Sigma^{-1}\mu_1 + \mu_2^T\Sigma^{-1}\mu_2 + \ln\frac{p(C_1)}{p(C_2)}$$

Therefore,

$$w = \Sigma^{-1}(\mu_1 - \mu_2)$$

$$w_0 = -\frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1^T + \frac{1}{2}\mu_2^T\Sigma^{-1}\mu_2^T + \ln\frac{p(C_1)}{p(C_2)}$$

3 Question 03

4 Question 04