Mineria de Datos

Estudiante UNED

August 2019

Ejercicio 1

Considere dos variables x e y con distribución de probabilidad conjunta $p(x,y)$. Demuestre que:

- $E[x] = E_y[E_x[x|y]]$
- $var[x] = E_y[var_x[x|y]] + var_y[E_x[x|y]]$

donde $E_x[x|y]$ representa el valor esperado de x asumiendo la distribución de probabilidad condicionada $p(x|y)$, y una notación equivalente se utiliza para la varianza condicional.

2.8

$$\mathbf{E_y}[\mathbf{E_x}[\mathbf{x}|\mathbf{y}]] = \int_y \left(\int_x x\,p(x|y)\,dx\right) p(y)\,dy = \int_y \int_x x\,p(x|y)\,p(y)\,dy\,dx = \int_y \int_x x\,p(x,y)\,dy\,dx = \int_x x\left(\int_y p(x,y)\,dy\right)dx = \int_x p(x)\,dx = \mathbf{E}[\mathbf{x}]$$

$$\mathbf{E_y}[\mathbf{var_x}[\mathbf{x}|\mathbf{y}]] + \mathbf{var_y}[\mathbf{E_x}[\mathbf{x}|\mathbf{y}]] =$$

$$\int_y \left(\int_x (x - E_x[x|y])^2\, p(x|y)\,dx\right) p(y)\,dy + \int_y \left(\underbrace{E_y[E_x[x|y]]}_{E[x]} - E_x[x|y]\right)^2 p(y)\,dy =$$

$$\int_y \int_x \left(x^2 - 2xE_x[x|y] + E[x|y]^2\right) p(x|y)\,p(y)\,dy\,dx + \int_y \left(E[x]^2 - 2E[x]\,E_x[x|y] + E_x[x|y]^2\right) p(y)\,dy =$$

$$\int_y \int_x x^2 \underbrace{p(x|y)\,p(y)}_{p(x,y)=p(x|y)p(y)}\,dy\,dx - 2\int_y \int_x xE_x[x|y]\,p(x|y)\,p(y)\,dy\,dx + \int_y \int_x E_x[x|y]^2 \underbrace{p(x|y)\,p(y)}_{p(x,y)=p(x|y)p(y)}\,dy\,dx + \int_y E[x]^2\,p(y)\,dy - 2\int_y E[x]\,E_x[x|y]\,p(y)\,dy + \int_y E_x[x|y]^2\,p(y)\,dy =$$

$$\int_x x^2 \underbrace{\left(\int_y p(x,y)\,dy\right)}_{p(x)} dx - 2\int_y E_x[x|y] \underbrace{\left(\int_x x\,p(x|y)\,dx\right)}_{E_x[x|y]} p(y)\,dy + \int_y E_x[x|y]^2 \underbrace{\left(\int_x p(x,y)\,dx\right)}_{p(y)} dy + E[x]^2 \int_y p(y)\,dy - 2E[x] \underbrace{\int_y E_x[x|y]\,p(y)\,dy}_{E_y[E_x[x|y]]} + \int_y E_x[x|y]^2\,p(y)\,dy =$$

$$\int_x x^2 p(x)\,dx - 2\int_y E_x[x|y]\,E_x[x|y]\,p(y)\,dy + \int_y E_x[x|y]^2\,p(y)\,dy + E[x]^2 \int_y p(y)\,dy - 2E[x] \underbrace{E_y[E_x[x|y]]}_{E[x]} + \int_y E_x[x|y]^2\,p(y)\,dy =$$

$E\left[x^2\right] - 2\int_y \cancel{E_x\left[x|y\right]^2 p\left(y\right)dy} + \int_y \cancel{E_x\left[x|y\right]^2 p\left(y\right)dy} + E\left[x\right]^2 - 2E\left[x\right]E\left[x\right] + \int_y \cancel{E_x\left[x|y\right]^2 p\left(y\right)dy} =$
$E\left[x^2\right] - E\left[x\right]^2 = \mathbf{var}\left[\mathbf{x}\right]$

Ejercicio 1

4.8

Using (4.57) and (4.58), derive the result (4.65) for the posterior class probability in the two-class generative model with Gaussian densities, and verify the results (4.66) and (4.67) for the parameters w and w0.

$p\left(c_1|x\right) = \dfrac{p\left(x|c_1\right)p\left(c_1\right)}{p\left(x|c_1\right)p\left(c_1\right) + p\left(x|c_2\right)p\left(c_2\right)} = \dfrac{1}{1+\exp\left(-a\right)} = \sigma\left(a\right)$
$a = \ln\dfrac{p\left(x|c_1\right)p\left(c_1\right)}{p\left(x|c_2\right)p\left(c_2\right)}$
$p\left(c_1|x\right) = \sigma\left(w^T x + w_0\right)$
$w = \Sigma^{-1}\left(\mu_1 - \mu_2\right)$
$w_0 = -\dfrac{1}{2}\mu_1^T\Sigma^{-1}\mu_1 + \dfrac{1}{2}\mu_2^T\Sigma^{-1}\mu_2 + \ln\dfrac{c_1}{c_2}$

Entonces
$\ln\dfrac{p\left(x|c_1\right)p\left(c_1\right)}{p\left(x|c_2\right)p\left(c_2\right)} = w^T x + w_0$
Teniendo en cuenta que:
$p\left(x|c_k\right) = \dfrac{1}{\left(2\pi\right)^{D/2}}\dfrac{1}{\left|\Sigma\right|^{1/2}}\exp\left\{-\dfrac{1}{2}\left(x-\mu_k\right)^T\Sigma^{-1}\left(x-\mu_k\right)\right\}$
$\ln\dfrac{p\left(x|c_1\right)p\left(c_1\right)}{p\left(x|c_2\right)p\left(c_2\right)} =$
$\ln\left\{\dfrac{\frac{1}{\left(2\pi\right)^{D/2}}\frac{1}{\left|\Sigma\right|^{1/2}}\exp\left\{-\frac{1}{2}\left(x-\mu_1\right)^T\Sigma^{-1}\left(x-\mu_1\right)\right\}}{\frac{1}{\left(2\pi\right)^{D/2}}\frac{1}{\left|\Sigma\right|^{1/2}}\exp\left\{-\frac{1}{2}\left(x-\mu_2\right)^T\Sigma^{-1}\left(x-\mu_2\right)\right\}}\right\} + \ln\dfrac{p\left(c_1\right)}{p\left(c_2\right)} =$
$\ln\left\{\exp\left\{-\dfrac{1}{2}\left(x-\mu_1\right)^T\Sigma^{-1}\left(x-\mu_1\right) + \dfrac{1}{2}\left(x-\mu_2\right)^T\Sigma^{-1}\left(x-\mu_2\right)\right\}\right\} + \ln\dfrac{p\left(c_1\right)}{p\left(c_2\right)} =$
$-\dfrac{1}{2}\left\{x^T\Sigma^{-1}\left(x-\mu_1\right) - {\mu_1}^T\Sigma^{-1}\left(x-\mu_1\right) - x^T\Sigma^{-1}\left(x-\mu_2\right) + {\mu_2}^T\Sigma^{-1}\left(x-\mu_2\right)\right\} + \ln\dfrac{p\left(c_1\right)}{p\left(c_2\right)} =$
$-\dfrac{1}{2}\left\{x^T\Sigma^{-1}x - x^T\Sigma^{-1}\mu_1 - {\mu_1}^T\Sigma^{-1}x + {\mu_1}^T\Sigma^{-1}\mu_1 - x^T\Sigma^{-1}x + x^T\Sigma^{-1}\mu_2 + {\mu_2}^T\Sigma^{-1}x - {\mu_2}^T\Sigma^{-1}\mu_2\right\} + \ln\dfrac{p\left(c_1\right)}{p\left(c_2\right)} =$
$-\dfrac{1}{2}\left\{-x^T\Sigma^{-1}\left(\mu_1-\mu_2\right) - \left(\mu_1-\mu_2\right)^T\Sigma^{-1}x + {\mu_1}^T\Sigma^{-1}\mu_1 - {\mu_2}^T\Sigma^{-1}\mu_2\right\} + \ln\dfrac{p\left(c_1\right)}{p\left(c_2\right)} =$
$-\dfrac{1}{2}\left\{-2\left(\mu_1-\mu_2\right)^T\Sigma^{-1}x + {\mu_1}^T\Sigma^{-1}\mu_1 - {\mu_2}^T\Sigma^{-1}\mu_2\right\} + \ln\dfrac{p\left(c_1\right)}{p\left(c_2\right)} = \left(\mu_1-\mu_2\right)^T\Sigma^{-1}x - \dfrac{1}{2}{\mu_1}^T\Sigma^{-1}\mu_1 + \dfrac{1}{2}{\mu_2}^T\Sigma^{-1}\mu_2 +$
$\ln\dfrac{p\left(c_1\right)}{p\left(c_2\right)} =$
$\underbrace{\left(\Sigma^{-1}\left(\mu_1-\mu_2\right)\right)^T}_{w^T} x \underbrace{-\dfrac{1}{2}{\mu_1}^T\Sigma^{-1}\mu_1 + \dfrac{1}{2}{\mu_2}^T\Sigma^{-1}\mu_2 + \ln\dfrac{p\left(c_1\right)}{p\left(c_2\right)}}_{w_0} =$
$w^T x + w_0$

4.10

Consider the classification model of Exercise 4.9 and now suppose that the class-conditional densities are given by Gaussian distributions with a shared covariance matrix, so that

$$p\left(\phi_n|C_k\right) = \mathcal{N}\left(\phi_n|\mu,\Sigma\right)$$

Show that the maximum likelihood solution for the mean of the Gaussian distribution for class Ck is given by

$$\mu_k \ = \frac{1}{N_k}\ \sum_{n=1}^{N}\ t_{n,k}\phi_n$$

which represents the mean of those feature vectors assigned to class Ck. Similarly, show that the maximum likelihood solution for the shared covariance matrix is given by

$$\Sigma = \sum_{k=1}^{K}\frac{N_k}{N}\ S_k$$

where

$$S_k = \frac{1}{N_k}\sum_{n=1}^{N}\ t_{n,k}\left(\phi_n-\mu_k\right)\left(\phi_n-\mu_k\right)^T$$

Thus Σ is given by a weighted average of the covariances of the data associated with each class, in which the weighting coefficients are given by the prior probabilities of the classes.

$$p\left(\{\phi_n,t_n\}\,|\,\{\pi_k\}\right) = \prod_{n=1}^{N}\prod_{k=1}^{K}\left\{p\left(\phi_n|C_k\right)p\left(C_k\right)\right\}^{t_{n,k}}$$

$$p\left(C_k\right) = \pi_k$$
$$p\left(\phi_n|C_k\right) = \mathcal{N}\left(\phi_n|\mu,\Sigma\right)$$

$$\ln p\left(\{\phi_n,t_n\}\,|\,\{\pi_k\}\right) = \ln\prod_{n=1}^{N}\prod_{k=1}^{K}\left\{\mathcal{N}\left(\phi_n|\mu,\Sigma\right)\ \pi_k\right\}^{t_{n,k}}$$

$$\ln\prod_{m=1}^{M}a_m = \ln\sum_{m=1}^{M}a_m$$
$$\ln a^n = n\ln a$$

$$\ln p\left(\{\phi_n,t_n\}\,|\,\{\pi_k\}\right) = \sum_{n=1}^{N}\sum_{k=1}^{K}t_{n,k}\ln\left\{\mathcal{N}\left(\phi_n|\mu,\Sigma\right)\ \pi_k\right\}$$

$$\ln\mathcal{N}\left(\phi_n|\mu,\Sigma\right) = \ln\left(2\pi\left|\Sigma\right|\right)^{-1/2} - \frac{1}{2}\left(\phi_n-\mu_k\right)^T\Sigma^{-1}\left(\phi_n-\mu_k\right)$$

$$\ln p\left(\{\phi_n,t_n\}\,|\,\{\pi_k\}\right) = \sum_{n=1}^{N}\sum_{k=1}^{K}t_{n,k}\left\{-\frac{1}{2}\ln\left(\left|\Sigma\right|\right) - \frac{1}{2}\left(\phi_n-\mu_k\right)^T\Sigma^{-1}\left(\phi_n-\mu_k\right) - \frac{1}{2}\ln\left(2\pi\right) + \ln\pi_k\right\}$$

$$\frac{\partial}{\partial\mu_m}\left\{\sum_{m=1}^{m}F\left(x_m\right)\right\} = \frac{\partial F\left(x_m\right)}{\partial\mu_m}$$

$$\frac{\partial}{\partial \mu_k} \left\{ \ln p \left(\{\phi_n, t_n\} \mid \{\pi_k\} \right) \right\} = \frac{\partial}{\partial \mu_k} \left\{ \sum_{n=1}^{N} t_{n,k} \left\{ -\frac{1}{2} \left(\phi_n - \mu_k\right)^T \Sigma^{-1} \left(\phi_n - \mu_k\right) \right\} \right\} = 0$$

$$\frac{\partial}{\partial s} \left(x - s\right)^T W \left(x - s\right) = 2W \left(x - s\right)$$

$$\sum_{n=1}^{N} t_{n,k} \ \left\{ -\Sigma^{-1} \left(\phi_n - \mu_k\right) \right\} = 0 \quad \Longrightarrow \quad -\sum_{n=1}^{N} \ t_{n,k} \Sigma^{-1} \phi_n - t_{n,k} \Sigma^{-1} \mu_k = 0$$

$$\sum a - b = \sum a - \sum b$$

$$\sum_{n=1}^{N} \ t_{n,k} \Sigma^{-1} \mu_k - \sum_{n=1}^{N} \ t_{n,k} \Sigma^{-1} \phi_n = 0 \quad \Longrightarrow \quad \sum_{n=1}^{N} \ t_{n,k} \Sigma^{-1} \mu_k = \sum_{n=1}^{N} \ t_{n,k} \Sigma^{-1} \phi_n$$

$$\Sigma \cdot \Sigma^{-1} = I$$

$$\Sigma^{-1} \mu_k \sum_{n=1}^{N} \ t_{n,k} = \Sigma^{-1} \sum_{n=1}^{N} \ t_{n,k} \phi_n \ \Longrightarrow \ \Sigma \left\{ \Sigma^{-1} \mu_k \sum_{n=1}^{N} \ t_{n,k} \right\} = \Sigma \left\{ \Sigma^{-1} \sum_{n=1}^{N} \ t_{n,k} \phi_n \right\}$$

$$N_k = \sum_{n=1}^{N} \ t_{n,k}$$

$$\mu_k \ N_k = \sum_{n=1}^{N} \ t_{n,k} \phi_n \quad \Longrightarrow \quad \mu_k \ = \frac{1}{N_k} \ \sum_{n=1}^{N} \ t_{n,k} \phi_n$$

$$\frac{\partial}{\partial \Sigma} \left\{ \ln p \left(\{\phi_n, t_n\} \mid \{\pi_k\} \right) \right\} = \frac{\partial}{\partial \Sigma} \left\{ \sum_{n=1}^{N} \sum_{k=1}^{K} t_{n,k} \ \left\{ -\frac{1}{2} \ln \left(|\Sigma|\right) - \frac{1}{2} \left(\phi_n - \mu_k\right)^T \Sigma^{-1} \left(\phi_n - \mu_k\right) \right\} \right\}$$

$$\sum a + b = \sum a + \sum b$$
$$D \left(a + b\right) = Da + Db$$

$$-\frac{1}{2} \frac{\partial}{\partial \Sigma} \left\{ \sum_{n=1}^{N} \sum_{k=1}^{K} t_{n,k} \ \ln \left(|\Sigma|\right) \right\} - \frac{1}{2} \frac{\partial}{\partial \Sigma} \left\{ \sum_{n=1}^{N} \sum_{k=1}^{K} t_{n,k} \ \left(\phi_n - \mu_k\right)^T \Sigma^{-1} \left(\phi_n - \mu_k\right) \right\} = 0$$

$$S_k = \frac{1}{N_k} \sum_{n=1}^{N} \ t_{n,k} \left(\phi_n - \mu_k\right) \left(\phi_n - \mu_k\right)^T$$

$$-\frac{1}{2} \left\{ \sum_{n=1}^{N} \sum_{k=1}^{K} t_{n,k} \frac{\partial \ln \left(|\Sigma|\right)}{\partial \Sigma} \ \right\} = \frac{1}{2} \left\{ \sum_{k=1}^{K} \frac{\partial}{\partial \Sigma} \underbrace{\sum_{n=1}^{N} t_{n,k} \left(\phi_n - \mu_k\right)^T \Sigma^{-1} \left(\phi_n - \mu_k\right)}_{N_k \ \ Tr(\Sigma^{-1} S_k)} \right\}$$

$$\frac{\partial l \ n(|A|)}{\partial \Sigma} = \left(A^{-1}\right)^T$$

$$-\sum_{n=1}^{N}\sum_{k=1}^{K} t_{n,k}\underbrace{\left(\Sigma^{-1}\right)^T}_{\Sigma^{-1}} = \sum_{k=1}^{K}\frac{\partial\left\{N_k\ Tr\left(\Sigma^{-1}S_k\right)\right\}}{\partial\Sigma}$$

$$\frac{\partial\left\{\ Tr\left(AX^{-1}B\right)\right\}}{\partial X} = \left(X^{-1}B\ A\ X^{-1}\right)^T$$

$$-\sum_{n=1}^{N}\sum_{k=1}^{K} t_{n,k}\Sigma^{-1} = -\sum_{k=1}^{K} N_k\left(\Sigma^{-1}S_k\Sigma^{-1}\right)^T$$

$$(A\ B\ C)^T = C^T B^T A^T\ N = \sum_{n=1}^{N}\sum_{k=1}^{K} t_{n,k}$$

$$\Sigma^{-1}\underbrace{\sum_{n=1}^{N}\sum_{k=1}^{K} t_{n,k}}_{N} = \sum_{k=1}^{K} N_k\left(\Sigma^{-1}S_k\Sigma^{-1}\right)^T$$

Multiplicando por Σ tanto por la izquierda como por la derecha ambos miembros de la ecuación

$$\Sigma\left(\Sigma^{-1}N\right)\Sigma = \Sigma\left(\sum_{k=1}^{K} N_k\Sigma^{-1}S_k\Sigma^{-1}\right)\Sigma\ N\ \Sigma = \sum_{k=1}^{K} N_k S_k \Longrightarrow \quad \Sigma = \sum_{k=1}^{K}\frac{N_k}{N}\ S_k$$