Mineria de Datos

Estudiante UNED

August 2019

2.8

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

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

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.

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

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

$$\ln\left\{\frac{\frac{1}{(2\pi)^{D/2}}\frac{1}{|\Sigma|^{1/2}}\exp\left\{-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)\right\}}{\frac{1}{(2\pi)^{D/2}}\frac{1}{|\Sigma|^{1/2}}\exp\left\{-\frac{1}{2}(x-\mu_2)^T\Sigma^{-1}(x-\mu_2)\right\}}\right\}+\ln\frac{p(c_1)}{p(c_2)}=$$

$$\ln\left\{\exp\left\{-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)+\frac{1}{2}(x-\mu_2)^T\Sigma^{-1}(x-\mu_2)\right\}\right\}+\ln\frac{p(c_1)}{p(c_2)}=$$

$$-\frac{1}{2}\left\{x^T\Sigma^{-1}(x-\mu_1)-{\mu_1}^T\Sigma^{-1}(x-\mu_1)-x^T\Sigma^{-1}(x-\mu_2)+{\mu_2}^T\Sigma^{-1}(x-\mu_2)\right\}+\ln\frac{p(c_1)}{p(c_2)}=$$

$$-\frac{1}{2}\left\{x^T\Sigma^{-1}x-x^T\Sigma^{-1}\mu_1-{\mu_1}^T\Sigma^{-1}x+{\mu_1}^T\Sigma^{-1}\mu_1-x^T\Sigma^{-1}x+x^T\Sigma^{-1}\mu_2+{\mu_2}^T\Sigma^{-1}x-{\mu_2}^T\Sigma^{-1}\mu_2\right\}+\ln\frac{p(c_1)}{p(c_2)}=$$

$$-\frac{1}{2}\left\{-x^T\Sigma^{-1}(\mu_1-\mu_2)-(\mu_1-\mu_2)^T\Sigma^{-1}x+{\mu_1}^T\Sigma^{-1}\mu_1-{\mu_2}^T\Sigma^{-1}\mu_2\right\}+\ln\frac{p(c_1)}{p(c_2)}=$$

$$-\frac{1}{2}\left\{-2(\mu_1-\mu_2)^T\Sigma^{-1}x+{\mu_1}^T\Sigma^{-1}\mu_1-{\mu_2}^T\Sigma^{-1}\mu_2\right\}+\ln\frac{p(c_1)}{p(c_2)}=(\mu_1-\mu_2)^T\Sigma^{-1}x-\frac{1}{2}{\mu_1}^T\Sigma^{-1}\mu_1+\frac{1}{2}{\mu_2}^T\Sigma^{-1}\mu_2+\ln\frac{p(c_1)}{p(c_2)}=$$

$$\underbrace{\left(\Sigma^{-1}(\mu_1-\mu_2)\right)^T}_{w^T}x\underbrace{-\frac{1}{2}{\mu_1}^T\Sigma^{-1}\mu_1+\frac{1}{2}{\mu_2}^T\Sigma^{-1}\mu_2+\ln\frac{p(c_1)}{p(c_2)}}_{w_0}=$$

$$w^Tx+w_0$$

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

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

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

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

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.

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

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

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

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

$$\ln a^n = n \ln a$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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