Mineria de Datos

Estudiante UNED

August 2019

Ejercicio 1

Calcule la divergencia de Kullback-Leibler entre dos gaussianas $p(x) = \mathcal{N}(x|\mu, \Sigma)$ y $q(x) = \mathcal{N}(x|m, L)$

1.30

Evaluate the Kullback-Leibler divergence (1.113) between two Gaussians $p(x) = \mathcal{N}(x|\mu, \Sigma)$ and $q(x) = \mathcal{N}(x|m, L)$

$$KL(p||q) = -\int p(x) \ln \frac{q(x)}{p(x)} dx$$

$$\ln \frac{p(x)}{q(x)} = \ln \frac{\mathcal{N}(x|m, L)}{\mathcal{N}(x|\mu, \Sigma)} = \ln \left\{ \frac{\frac{1}{(2\pi)^{D/2}} \frac{1}{|L|^{1/2}} \exp\left\{-\frac{1}{2}(x-m)^T L^{-1}(x-m)\right\}}{\frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right\}} \right\} =$$

$$\ln\left\{ \left(\frac{|\Sigma|}{|L|}\right)^{\frac{1}{2}} \exp\left\{\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu) - \frac{1}{2}(x-m)^T L^{-1}(x-m)\right\}\right\} =$$

$$\frac{1}{2}\ln\left(\frac{|\Sigma|}{|L|}\right) + \frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu) - \frac{1}{2}(x-m)^T L^{-1}(x-m) =$$

$$-\frac{1}{2}\left\{\ln\left(\frac{|\Sigma|}{|L|}\right) + x^T\Sigma^{-1}(x-\mu) - \mu^T\Sigma^{-1}(x-\mu) - x^T L^{-1}(x-m) + m^T L^{-1}(x-m)\right\} =$$

$$-\frac{1}{2}\left\{\ln\left(\frac{|\Sigma|}{|L|}\right) + x^T\Sigma^{-1}x \underbrace{-x^T\Sigma^{-1}\mu - \mu^T\Sigma^{-1}x}_{x^T\Sigma^{-1}\mu \equiv \mu^T\Sigma^{-1}x} + \mu^T\Sigma^{-1}\mu - x^T L^{-1}x \underbrace{+x^T L^{-1}m + m^T L^{-1}x}_{x^T L^{-1}m \equiv m^T L^{-1}x} - m^T L^{-1}m\right\} =$$

$$-\frac{1}{2}\left\{\ln\left(\frac{|\Sigma|}{|L|}\right) + x^T\left(\Sigma^{-1} - L^{-1}\right)x - 2\mu^T\Sigma^{-1}x + \mu^T\Sigma^{-1}\mu + 2m^T L^{-1}x - m^T L^{-1}m\right\} =$$

$$-\frac{1}{2}\left\{\ln\left(\frac{|\Sigma|}{|L|}\right) + x^T\left(\Sigma^{-1} + L^{-1}\right)x - 2\left(\mu^T\Sigma^{-1} - m^T L^{-1}\right)x + \mu^T\Sigma^{-1}\mu - m^T L^{-1}m\right\}$$

$$KL(p||q) = -\int p(x) \ln \frac{q(x)}{p(x)} dx =$$

$$-\int \mathcal{N}(x|\mu, \Sigma) \ln \frac{\mathcal{N}(x|m, L)}{\mathcal{N}(x|\mu, \Sigma)} dx = -\int \mathcal{N}(x|\mu, \Sigma)\, dx \left(-\frac{1}{2}\right)\left\{\ln\left(\frac{|\Sigma|}{|L|}\right) + x^T\left(\Sigma^{-1} + L^{-1}\right)x - 2\left(\mu^T\Sigma^{-1} - m^T L^{-1}\right)x + \mu^T\Sigma^{-1}\mu -$$

$$\frac{1}{2}\left\{\underbrace{\int \ln\left(\frac{|\Sigma|}{|L|}\right)\mathcal{N}(x|\mu, \Sigma)\, dx}_{\int \mathcal{N}(x|\mu,\Sigma)dx=1} + \int x^T\left(\Sigma^{-1} + L^{-1}\right)x\mathcal{N}(x|\mu, \Sigma)\, dx + \int -2\left(\mu^T\Sigma^{-1} - m^T L^{-1}\right)x\mathcal{N}(x|\mu, \Sigma)\, dx + \underbrace{\int \left(\mu^T\Sigma^{-1}\mu - m^T L\right.}_{\int \mathcal{N}(x|\mu}\right.$$

$$\frac{1}{2}\left\{\ln\left(\frac{|\Sigma|}{|L|}\right)+\int x^T\left(\Sigma^{-1}+L^{-1}\right)x\mathcal{N}\left(x|\mu,\Sigma\right)dx-2\left(\mu^T\Sigma^{-1}-m^TL^{-1}\right)\underbrace{\int x\mathcal{N}\left(x|\mu,\Sigma\right)dx}_{\mu}+\left(\mu^T\Sigma^{-1}\mu-m^TL^{-1}m\right)\right\}=$$
$$\frac{1}{2}\left\{\ln\left(\frac{|\Sigma|}{|L|}\right)+\int x^T\left(\Sigma^{-1}+L^{-1}\right)x\mathcal{N}\left(x|\mu,\Sigma\right)dx-2\left(\mu^T\Sigma^{-1}-m^TL^{-1}\right)\mu+\left(\mu^T\Sigma^{-1}\mu-m^TL^{-1}m\right)\right\}=$$
$$\frac{1}{2}\left\{\ln\left(\frac{|\Sigma|}{|L|}\right)+\int x^T\left(\Sigma^{-1}+L^{-1}\right)x\mathcal{N}\left(x|\mu,\Sigma\right)dx+\left(2m^TL^{-1}\mu-\mu^T\Sigma^{-1}\mu-m^TL^{-1}m\right)\right\}$$

Ejercicio 1

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

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

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

2.8

Consider two variables x and y with joint distribution p(x, y). Prove the following two results

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

Here $E_x\left[x|y\right]$ denotes the expectation of x under the conditional distribution $p\left(x|y\right)$, with a similar notation for the conditional variance.

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

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

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

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

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

$$E\left[x^2\right]-E\left[x\right]^2=\mathbf{var}\left[\mathbf{x}\right]$$

Ejercicio 1

Sabiendo que en un problema de clasificación con dos clases

$$p\left(C_1|x\right)=\frac{1}{1+\exp\left(-\ln\frac{p(x|c_1)p(c_1)}{p(x|c_2)p(c_2)}\right)}=\frac{1}{1+\exp\left(-a\right)}=\sigma\left(a\right)$$

y suponiendo un modelo generativo en el que las verosimilitudes (likelihoods) de las dos clases vienen dadas por dos gaussianas de medias μ_1 y μ_2 , pero la misma varianza Σ , demuestre que

$$p\left(c_1|x\right)=\sigma\left(w^Tx+w_0\right)$$

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

$$w_0=-\frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1+\frac{1}{2}\mu_2^T\Sigma^{-1}\mu_2+\ln\frac{c_1}{c_2}$$

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.

$$p\left(c_1|x\right)=\frac{p\left(x|c_1\right)p\left(c_1\right)}{p\left(x|c_1\right)p\left(c_1\right)+p\left(x|c_2\right)p\left(c_2\right)}=\frac{1}{1+\exp\left(-a\right)}=\sigma\left(a\right)$$

$$a=\ln\frac{p\left(x|c_1\right)p\left(c_1\right)}{p\left(x|c_2\right)p\left(c_2\right)}$$

$$p\left(c_1|x\right)=\sigma\left(w^Tx+w_0\right)$$

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

$$w_0=-\frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1+\frac{1}{2}\mu_2^T\Sigma^{-1}\mu_2+\ln\frac{c_1}{c_2}$$

Entonces

$$\ln\frac{p\left(x|c_1\right)p\left(c_1\right)}{p\left(x|c_2\right)p\left(c_2\right)}=w^Tx+w_0$$

Teniendo en cuenta que:

$$p\left(x|c_k\right)=\frac{1}{\left(2\pi\right)^{D/2}}\frac{1}{\left|\Sigma\right|^{1/2}}\exp\left\{-\frac{1}{2}\left(x-\mu_k\right)^T\Sigma^{-1}\left(x-\mu_k\right)\right\}$$

$$\ln\frac{p\left(x|c_1\right)p\left(c_1\right)}{p\left(x|c_2\right)p\left(c_2\right)}=$$

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

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

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

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

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

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

$$\ln\frac{p(c_1)}{p(c_2)}=$$

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

$$w^Tx+w_0$$

Ejercicio 1

Considere un modelo generativo de clasificación de K clases definido por K probabilidades a priori $p\left(C_k\right) = \pi_k$ y densidades de probabilidad del vector de características de entrada Φ condicionadas a la clase $p\left(\Phi|C_k\right)$ dadas por distribuciones normales multi-variantes con la misma covarianza:

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

Suponga que se nos proporciona un conjunto de entrenamiento Φ_m, t_n donde el subíndice n tomca valores $n = 1, ..., N$ y t_n es un vector binario de longitud K que utiliza la codificación uno-de-K (es decir, que sus componentes son $t_{n,j} = I_{j,k}$ si el patrón t_n pertenece a la clase C_k). Si asumimos que el conjunto de entrenamiento constituye una muestra independiente de datos de este modelo, entonces el estimador máximo-verosímil de las probabilidades a priori viene dado por

$$\pi_k = \frac{N_k}{N}$$

donde N_k es el número de patrones asignados ala clase C_k.
Demuestre que el estimador máximo-versímil de la media de la distribucipn de la clase C_k viene dado por

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

y de la matriz de covarianza, viene dado por

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

con

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

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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