TEMA 2.2 (Estimadores ML)

**Ejercicio 1**

Suponga que se dispone de un conjunto de $N$ vectores de características $\mathbf{x}_i$, $i$=1,…,$N$ conjuntamente Gaussianas, cuya funcion de densidad de probabilidad es:

$$f_{\mathbf{x}}(\mathbf{x}\,|\,\boldsymbol{\theta}) = \frac{1}{(2\pi)^{d/2}|\mathbf{C}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu}\ \ )\right\}$$

Si los vectores observados son estadísticamente independientes, y la matriz de covarianza $\mathbf{C}$ es conocida, demuestre que el estimador ML de la media $\boldsymbol{\mu}_i$, es:

$$\hat{\boldsymbol{\mu}}_{ML} = \frac{1}{N}\sum_{k=1}^{N}\mathbf{x}_k$$

**Ejercicio 2**

Bajo las mismas hipótesis del ejercicio 1, demuestre que los estimadores de ML del vector de media y de la matriz de covarianza vienen dados por:

$$\hat{\boldsymbol{\mu}}_{ML} = \frac{1}{N}\sum_{k=1}^{N}\mathbf{x}_k \qquad\qquad \hat{\mathbf{C}}_{ML} = \frac{1}{N}\sum_{k=1}^{N}(\mathbf{x}_k-\hat{\boldsymbol{\mu}}_{ML})(\mathbf{x}_k-\hat{\boldsymbol{\mu}}_{ML})^T$$

para lo cual use las relaciones de análisis matricial que aparecen en el ANEXO, así como la propiedad $\mathbf{a}^T\mathbf{R}\mathbf{a} = traza\left(\mathbf{R}\mathbf{a}\mathbf{a}^T\right)$.

**Ejercicio 3**

Suponga que se dispone de un conjunto de $N$ vectores de características $\mathbf{x}_i$, $i$=1,…,$N$ . Las $d$ componentes del vector son datos binarios {0,1} como probabilidades de aparición $p_k$, $k$=1,…, $d$, de forma que la funcion de densidad de probabilidad del vector de observaciones es:

$$f_{\mathbf{x}}(\mathbf{x}|\mathbf{p}) = \prod_{k=1}^{d} p_k^{\ x_k}(1-p_k)^{1-x_k}$$
$$\mathbf{p} = [\,p_1,\ldots,p_d\,]$$

Tanto los vectores observados como las características son independientes entre si. Obtenga la estimación ML de las probabilidades.

**Ejercicio 4**

Sean $\mathbf{S}_B$ y $\mathbf{S}_C$ dos matrices reales simétricas de dimensiones $d$x$d$. Se definen los autovectores generalizados como aquellos que verifican $\mathbf{S}_B\mathbf{w}_j = \lambda_j\mathbf{S}_C\mathbf{w}_j$, y que pueden calcularse a partir de la relacion:

$$\left|\mathbf{S}_B - \lambda_j \mathbf{S}_C\right| = 0$$

Si $\mathbf{S}_C$ es definida positiva, los autovectores pueden normalizarse de forma que $\mathbf{w}_i^T \mathbf{S}_C \mathbf{w}_j = \delta_{ij}$ y $\mathbf{w}_i^T \mathbf{S}_B \mathbf{w}_j = \lambda_{ij}\delta_{ij}$. Supongamos que $n(< d)$ vectores $\mathbf{w}_j$ se agrupan en las columas de la matriz **W**, de forma que se construyen las matrices $\tilde{\mathbf{S}}_C = \mathbf{W}^T \mathbf{S}_C \mathbf{W}$ y $\tilde{\mathbf{S}}_B = \mathbf{W}^T \mathbf{S}_B \mathbf{W}$.

a) Demuestre que $\tilde{\mathbf{S}}_C$ es la identidad (de dimensiones *n* x *n*) y que $\tilde{\mathbf{S}}_B$ es una matriz diagonal cuyos elementos son los autovalores generalizados $\lambda_j$.
b) Calcule el valor de $J = \left|\tilde{\mathbf{S}}_B\right| / \left|\tilde{\mathbf{S}}_C\right|$.
c) Suponga que las matrices $\mathbf{S}_B$ y $\mathbf{S}_C$ se calculan a partir de los vectores de observaciones **x** de distintas clases, como la matriz de dispersion interclases e intraclases, respectivamente. Demuestre que el valor de $J = \left|\tilde{\mathbf{S}}_B\right| / \left|\tilde{\mathbf{S}}_C\right|$ queda inalterado si los vectores se modifican de la siguiente forma $\mathbf{y} = \mathbf{Q}\mathbf{D}\mathbf{W}^{\mathbf{T}}\mathbf{x}$, donde **Q** es una matriz de rotacion (y por tanto ortogonal), y **D** es una matriz diagonal no singular.

Notas: Tenga en cuenta que $\mathbf{W}^T\mathbf{W} = \mathbf{I}$ y que $\mathbf{W}\mathbf{W}^T$ es una matrix singular. Úsese la propiedad $|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$

**Ejercicio 5**

Sea la transformación $\mathbf{y}_n = \mathbf{W}^T(\mathbf{x}_n - \mathbf{m})$, demuestre que si $\mathbf{W} = \mathbf{w}$ el valor máximo de $\sum_{n=1}^{N} |y_n|^2$ y con la restricción de que el vector **w** es de norma igual a la unidad, se produce cuando **w** es el autovector asociado al máximo autovalor de la matriz $\mathbf{S} = \sum_{n=1}^{N} (\mathbf{x}_n - \mathbf{m})(\mathbf{x}_n - \mathbf{m})^T$. A la vista de los resultados decida cual es la matriz **W** de dimensiones dxk (k<d) que maximiza la norma de los vectores transformados $\mathbf{y}_n = \mathbf{W}^T(\mathbf{x}_n - \mathbf{m})$.

**Ejercicio 6**

Demuestre que el máximo rango de la matriz de covarianza inter clases $\mathbf{S}_B = \sum_{i=1}^{c} N_i(\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T$ es *c-1*.

**Ejercicio 7**

Sea la transformación $\mathbf{y}_n = \mathbf{W}^T(\mathbf{x}_n - \mathbf{m})$, demuestre que si $\mathbf{W} = \mathbf{w}$ el valor máximo del cociente $\mathbf{W} = \arg\max_{\mathbf{W}} \frac{\left|\mathbf{W}^T\mathbf{S}_B\mathbf{W}\right|}{\left|\mathbf{W}^T\mathbf{S}_C\mathbf{W}\right|}$, se produce cuando **w** es el autovector generalizado asociado al máximo autovalor de $\mathbf{S}_B\mathbf{w}_j = \lambda_j\mathbf{S}_C\mathbf{w}_j$. A la vista de los resultados decida cual es la matriz **W** de dimensiones dxk (k<d) que maximiza la norma de los vectores transformados $\mathbf{y}_n = \mathbf{W}^T(\mathbf{x}_n - \mathbf{m})$.

ANEXO

## Gradientes vectoriales

Sea $a(\mathbf{x})$ una función escalar dependiente del vector **x** de dimensión Nx1 $x \in \mathbb{R}^N$ . El gradiente de *a* respecto a **x** es el vector:

$$\frac{\partial a(\mathbf{x})}{\partial \mathbf{x}} = \nabla_{\mathbf{x}}(a(\mathbf{x})) = \begin{pmatrix} \dfrac{\partial a(\mathbf{x})}{\partial x_1} \\ .. \\ \dfrac{\partial a(\mathbf{x})}{\partial x_N} \end{pmatrix}$$

Algunos casos particulares:

1. Matriz Identidad $\quad \nabla_{\mathbf{x}}(\mathbf{x}) = \mathbf{I}_{NxN}$

2. Operador Lineal $\nabla_{\mathbf{x}}\left(\mathbf{b}^T\mathbf{x}\right) = \nabla_{\mathbf{x}}\left(\mathbf{x}^T\mathbf{b}\right) = \mathbf{b}$ donde **b** no depende de **x**

3. Producto escalar de vectores dependientes de **x**:
$$\nabla_{\mathbf{x}}\left(\mathbf{a}^T(\mathbf{x})\mathbf{b}(\mathbf{x})\right) = \nabla_{\mathbf{x}}^T\left(\mathbf{a}(\mathbf{x})\right)\mathbf{b}(\mathbf{x}) + \mathbf{a}^T(\mathbf{x})\nabla_{\mathbf{x}}\left(\mathbf{b}(\mathbf{x})\right)$$

4. Forma Cuadrática:
   - Matriz Simétrica $\mathbf{Q} = \mathbf{Q}^T \in \mathbb{R}^{NxN} \Rightarrow \quad \nabla_{\mathbf{x}}\left(\mathbf{x}^T\mathbf{Q}\mathbf{x}\right) = 2\mathbf{Q}\mathbf{x}$
   - Matriz NO Simétrica $\mathbf{Q} \in \mathbb{R}^{NxN} \Rightarrow \quad \nabla_{\mathbf{x}}\left(\mathbf{x}^T\mathbf{Q}\mathbf{x}\right) = \mathbf{Q}\mathbf{x} + \mathbf{Q}^T\mathbf{x}$

5. Forma Cuadrática Parametrizada ( $\mathbf{x} \in \mathbb{R}^{NxN}$; $\quad \mathbf{a}(\mathbf{x}) \in \mathbb{R}^{MxM}$ ):
   - Matriz Simétrica $\mathbf{Q} = \mathbf{Q}^T \in \mathbb{R}^{MxM} \quad \Rightarrow$
     $\nabla_{\mathbf{x}}\left(\mathbf{a}^T(\mathbf{x})\mathbf{Q}\mathbf{a}(\mathbf{x})\right) = 2\nabla_{\mathbf{x}}\left(\mathbf{a}(\mathbf{x})\right)\mathbf{Q}\mathbf{a}(\mathbf{x})$

**Gradientes matriciales**

Sea $a(\mathbf{R})$ una función escalar que depende de una matriz $\mathbf{R} \in \mathbb{R}^{MxN}$. La función gradiente de *a* respecto a la matriz se define como:

$$\nabla_{\mathbf{R}}(a) = \begin{pmatrix} \frac{\partial a(\mathbf{R})}{r_{11}} & \frac{\partial a(\mathbf{R})}{r_{12}} & \cdots & \frac{\partial a(\mathbf{R})}{r_{1N}} \\ \frac{\partial a(\mathbf{R})}{r_{21}} & \frac{\partial a(\mathbf{R})}{r_{22}} & \cdots & \frac{\partial a(\mathbf{R})}{r_{2N}} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial a(\mathbf{R})}{r_{M1}} & \frac{\partial a(\mathbf{R})}{r_{M2}} & \cdots & \frac{\partial a(\mathbf{R})}{r_{MN}} \end{pmatrix}$$

Cada vector columna corresponde al gradiente vectorial del escalar *a* respecto a una columna de **R**.

Algunos casos particulares:

Sea $\mathbf{R} \in \mathbb{R}^{NxN}$. El determinante de la matriz puede calcularse como:

$$\det(\mathbf{R}) = |\mathbf{R}| = \det(\mathbf{R}) = \sum_{j=1}^{N} r_{ij} (\mathbf{A}_{\mathbf{R}})_{ij}$$

donde $\mathbf{A}_{\mathbf{R}}$ es la matriz de adjuntos o cofactores:

$$(\mathbf{A}_{\mathbf{R}})_{ij} = (-1)^{i+j} \det(\mathbf{R}_{ij})$$

$\mathbf{R}_{ij}$ es la matriz **R** en la que se ha eliminado la fila *i* y la columna *j*.

1. Gradiente del determinante
   - Si la matriz NO es simétrica: $\nabla_{\mathbf{R}}(\det(\mathbf{R})) = \det(\mathbf{R})\mathbf{R}^{-T}$
   - Si la matriz es simétrica: $\mathbf{R} = \mathbf{R}^T \quad \nabla_{\mathbf{R}}(\det(\mathbf{R})) = \det(\mathbf{R})\left(2\mathbf{R}^{-1} - diag(\mathbf{R}^{-1})\right)$
2. $\nabla_{\mathbf{R}}(\ln(\det(\mathbf{R}))) = \frac{1}{\det(\mathbf{R})}\nabla_{\mathbf{R}}(\det(\mathbf{R}))$
3. $\nabla_{\mathbf{R}}\left(\text{traza}\left(\mathbf{ARB}\right)\right) = \mathbf{A}^T\mathbf{B}^T$
4. $\nabla_{\mathbf{R}}\left(\text{traza}\left(\mathbf{AR}^{-1}\mathbf{B}\right)\right) = -\left(\mathbf{R}^{-1}\mathbf{BAR}^{-1}\right)^T$