---

Para la realización del examen se permite un único libro (sea cual sea, salvo el texto **PATTERN RECOGNITION AND MACHINE LEARNING - SOLUTIONS TO EXERCISES - TUTORS' EDITION** de M. Svensén y C.M. Bishop) sin anotaciones de ningún tipo. Si considera que el enunciado contiene algún erratum, hágalo constar en su respuesta y, **de confirmarse**, el equipo docente compensará el tiempo invertido por el estudiante. En los enunciados, asuma el empleo de la notación habitual del texto base.

---

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

2. Deduzca la ecuación para (el logaritmo de) la evidencia de un modelo de regresión lineal con hiperparámetros $\alpha$ y $\beta$

$$\ln p(\boldsymbol{t}|\alpha, \beta) = \frac{M}{2} \ln \alpha + \frac{N}{2} \ln \beta - E(\boldsymbol{m}_N) - \frac{1}{2} \ln |\boldsymbol{A}| - \frac{N}{2} \ln(2\pi)$$

sabiendo que si $p(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Lambda}^{-1})$ y $p(\boldsymbol{y}|\boldsymbol{x}) = \mathcal{N}(\boldsymbol{y}|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}, \boldsymbol{L}^{-1})$, la distribución marginal viene dada por

$$p(\boldsymbol{y}) = \mathcal{N}(\boldsymbol{y}|\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{b}, \boldsymbol{L}^{-1} + \boldsymbol{A}\boldsymbol{\Lambda}^{-1}\boldsymbol{A}^T)$$

con la interpretación habitual de los símbolos implicados utilizada en el texto base.

3. Considere un modelo generativo de clasificación de $K$ clases definido por $K$ probabilidades a priori $p(\mathcal{C}_k) = \pi_k$ y densidades de probabilidad del vector de características de entrada $\boldsymbol{\phi}$ condicionadas a la clase $p(\boldsymbol{\phi}|\mathcal{C}_k)$ dadas por distribuciones normales multi-variantes con la misma covarianza:

$$p(\boldsymbol{\phi}|\mathcal{C}_k) = \mathcal{N}(\boldsymbol{\phi}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma})$$

Suponga que se nos proporciona un conjunto de entrenamiento $\boldsymbol{\phi}_n, \boldsymbol{t}_n$ donde el subíndice $n$ toma valores $n = 1, ..., N$ y $\boldsymbol{t}_n$ es un vector binario de longitud $K$ que utiliza la codificación uno-de-$K$ (es decir, que sus componentes son $t_{nj} = I_{jk}$ si el patrón $\boldsymbol{t}_n$ pertenece a la clase $\mathcal{C}_k$). 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 a la clase $\mathcal{C}_k$.

Demuestre que el estimador máximo-verosimil de la media de la distribución de la clase $\mathcal{C}_k$ viene dado por

$$\boldsymbol{\mu}_k = \frac{1}{N_k}\sum_{n=1}^{N} t_{nk}\boldsymbol{\phi}_n$$

y el de la matriz de covarianza, viene dado por

$$\boldsymbol{\Sigma} = \sum_{k=1}^{K} \frac{N_k}{N}\boldsymbol{S}_k$$

con

$$\boldsymbol{S}_k = \frac{1}{N_k}\sum_{n=1}^{N} t_{nk}(\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)(\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)^T$$