18F1

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.

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

3.2 2. Sea $t = \boldsymbol{w}^T \boldsymbol{\phi}(\boldsymbol{x}) + \epsilon$;
t, un variable aleatoria que toma valores en $\mathbb{R}$;
$\boldsymbol{x}$, un vector de $\mathbb{R}^{in}$;
$\boldsymbol{\phi} = (1, \phi_1(\boldsymbol{x}), ..., \phi_{M-1}(\boldsymbol{x}))^T$ con $\phi_i(\cdot)$ $(i = 1, ..., M-1)$ un conjunto de $M-1$ transformaciones (no necesariamente lineales) escalares sobre $\mathbb{R}^{in}$;
$\boldsymbol{w} = (w_0, w_1, ..., w_{M-1})^T$, un vector de $\mathbb{R}^M$; y, finalmente, ϵ una variable aleatoria distribuida según una gaussiana $\mathcal{N}(0, \beta^{-1})$. Sea $(\boldsymbol{x}_1, ..., \boldsymbol{x}_N)$ y $(t_1, ..., t_N)$ un conjunto de valores de $\boldsymbol{x}$ y t obtenidos a partir de las distribuciones descritas por todo lo anterior.

Defínase $\boldsymbol{\Phi}$ como la matriz de dimensiones $N \times M$ en que la fila j es igual a $(1, \phi_1(\boldsymbol{x}_j), ..., \phi_{M-1}(\boldsymbol{x}_j))$.

Demuestre que la matriz $\Phi(\Phi^T\Phi)^{-1}\Phi^T$ proyecta cualquier vector v en el espacio generado por los vectores columna de ϕ. Utilice este resultado para probar que la solución de mínimos cuadrados dada por $w = (\Phi^T\Phi)^{-1}\Phi^T t$ corresponde a una proyección ortogonal del vector t sobre la variedad (manifold en inglés) lineal S definida por los vectores columna de $\boldsymbol{\Phi}$, $\psi_k = (\phi_k(\boldsymbol{x}_1), ..., \phi_k(\boldsymbol{x}_N))$ como se muestra en la Fig. 1.

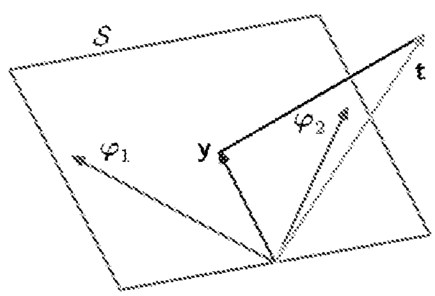

Figura 1:

4.1 3. Dado un conjunto de puntos $\{\boldsymbol{x}_n\}$, podemos definir su envolvente convexa (*convex hull*) como el conjunto de todos los puntos $\boldsymbol{x}$ tales que

$$\boldsymbol{x} = \sum_n \alpha_n \boldsymbol{x}_n$$

donde $\alpha_n \geq 0$ y $\sum_n \alpha_n = 1$. Considere un segundo conjunto de puntos $\boldsymbol{y}_n$ y su envolvente convexa. Por definición, los dos conjuntos de puntos $\{\boldsymbol{x}_n\}$ y $\{\boldsymbol{y}_n\}$ serán linealmente separables si existe un vector $\hat{\boldsymbol{w}}$ y un escalar w_0 tales que para todo $\boldsymbol{x}_n$, $\hat{\boldsymbol{w}}^T\boldsymbol{x}_n + w_0 > 0$, y para todo $\boldsymbol{y}_n$, $\hat{\boldsymbol{w}}^T\boldsymbol{y}_n + w_0 < 0$. Demuestre que si sus envolventes convexas tienen intersección no nula, entonces los dos conjuntos de puntos no pueden ser linealmente separables y *vice versa*: si son linealmente separables entonces sus envolventes convexas tienen interesección nula.

18F2

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.

2.33 1. Considere la distribución de probabilidad conjunta $p(\boldsymbol{x}, \boldsymbol{y})$ definida por las distribuciones marginal y condicional siguientes:

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

donde $\boldsymbol{\mu}$ y $\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ son las medias de la distribución gaussiana, y $\boldsymbol{\Lambda}$ y $\boldsymbol{L}$ son las matrices de precisión. Utilice la técnica de completar el cuadrado para obtener una expresión analítica de la media y covarianza de $p(\boldsymbol{x}|\boldsymbol{y})$.

3.11 2. Utilice la identidad matricial

$$(\boldsymbol{M} + \boldsymbol{v}\boldsymbol{v}^T)^{-1} = \boldsymbol{M}^{-1} - \frac{(\boldsymbol{M}^{-1}\boldsymbol{v})(\boldsymbol{v}^T\boldsymbol{M}^{-1})}{1 + \boldsymbol{v}^T\boldsymbol{M}^{-1}\boldsymbol{v}}$$

y suponga que

$$\boldsymbol{S}_{N+1}^{-1} = \boldsymbol{S}_N^{-1} + \beta\boldsymbol{\phi}_{N+1}\boldsymbol{\phi}_{N+1}^T \tag{1}$$

para demostrar que la incertidumbre de un modelo predictivo bayesiano construido a partir de N observaciones, y dada por

$$\sigma_N^2(\boldsymbol{x}) = \frac{1}{\beta} + \boldsymbol{\phi}(\boldsymbol{x})^T\boldsymbol{S}_N\boldsymbol{\phi}(\boldsymbol{x}) \tag{2}$$

(con la interpretación habitual de los símbolos utilizados en el texto base) satisface que $\sigma_{N+1}^2(\boldsymbol{x}) \leq \sigma_N^2(\boldsymbol{x})$.

4.8 3. Sabiendo que en un problema de clasificación con dos clases

$$p(\mathcal{C}_1|\boldsymbol{x}) = \frac{1}{1 + \exp(-\ln\frac{p(\boldsymbol{x}|\mathcal{C}_1)p(\mathcal{C}_1)}{p(\boldsymbol{x}|\mathcal{C}_2)p(\mathcal{C}_2)})} = \frac{1}{1 + \exp(-a)} = \sigma(a) \tag{3}$$

y suponiendo un modelo generativo en el que las verosimilitudes (likelihoods) de las dos clases vienen dadas por dos gaussianas de medias $\boldsymbol{\mu}_1$ y $\boldsymbol{\mu}_2$, pero la misma varianza $\boldsymbol{\Sigma}$, demuestre que

$$p(\mathcal{C}_1|\boldsymbol{x}) = \sigma(\boldsymbol{w}^T\boldsymbol{x} + w_0) \tag{4}$$

con $\boldsymbol{w} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ y

$$w_0 = -\frac{1}{2}\boldsymbol{\mu}_1^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1 + \frac{1}{2}\boldsymbol{\mu}_2^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_2 + \ln\frac{p(\mathcal{C}_1)}{p(\mathcal{C}_2)}. \tag{5}$$

18SO

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.

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

   - $\mathbb{E}[x] = \mathbb{E}_y[\mathbb{E}_x[x|y]]$
   - $var[x] = \mathbb{E}_y[var_x[x|y]] + var_y[\mathbb{E}_x[x|y]]$

   donde $\mathbb{E}_x[x|y]$ 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.

3.16 2. Deduzca la ecuación para (el logaritmo de) la evidencia de un modelo de regresión lineal con hiperparámetros α y β

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

   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.

4.1 3. Dado un conjunto de puntos $\{\boldsymbol{x}_n\}$, podemos definir su envolvente convexa (*convex hull*) como el conjunto de todos los puntos $\boldsymbol{x}$ tales que

$$\boldsymbol{x} = \sum_n \alpha_n \boldsymbol{x}_n$$

   donde $\alpha_n \geq 0$ y $\sum_n \alpha_n = 1$. Considere un segundo conjunto de puntos $\boldsymbol{y}_n$ y su envolvente convexa. Por definición, los dos conjuntos de puntos $\{\boldsymbol{x}_n\}$ y $\{\boldsymbol{y}_n\}$ serán linealmente separables si existe un vector $\hat{\boldsymbol{w}}$ y un escalar w_0 tales que para todo $\boldsymbol{x}_n$, $\hat{\boldsymbol{w}}^T\boldsymbol{x}_n + w_0 > 0$, y para todo $\boldsymbol{y}_n$, $\hat{\boldsymbol{w}}^T\boldsymbol{y}_n + w_0 < 0$. Demuestre que si sus envolventes convexas tienen intersección no nula, entonces los dos conjuntos de puntos no pueden ser linealmente separables y *vice versa*: si son linealmente separables entonces sus envolventes convexas tienen interesección nula.

18SR

EXAMEN DE RESERVA NO DISPONIBLE

El contenido de este examen de reserva no está disponible, conforme al acuerdo del Consejo de Gobierno de la UNED de 11 de noviembre de 2015, en el que se acordó:

- No publicar los exámenes de reserva no utilizados en la valija virtual de centros nacionales.

17F1

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.

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

3.16 2. Deduzca la ecuación para (el logaritmo de) la evidencia de un modelo de regresión lineal con hiperparámetros α y β

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

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.

4.10 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

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

y el de la matriz de covarianza, viene dado por

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

con

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

17F2

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.

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

- $\mathbb{E}[x] = \mathbb{E}_y[\mathbb{E}_x[x|y]]$
- $var[x] = \mathbb{E}_y[var_x[x|y]] + var_y[\mathbb{E}_x[x|y]]$

donde $\mathbb{E}_x[x|y]$ 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.

3.16 2. Deduzca la ecuación para (el logaritmo de) la evidencia de un modelo de regresión lineal con hiperparámetros α y β

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

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.

4.1 3. Dado un conjunto de puntos $\{\boldsymbol{x}_n\}$, podemos definir su envolvente convexa (*convex hull*) como el conjunto de todos los puntos $\boldsymbol{x}$ tales que

$$\boldsymbol{x} = \sum_n \alpha_n \boldsymbol{x}_n$$

donde $\alpha_n \geq 0$ y $\sum_n \alpha_n = 1$. Considere un segundo conjunto de puntos $\boldsymbol{y}_n$ y su envolvente convexa. Por definición, los dos conjuntos de puntos $\{\boldsymbol{x}_n\}$ y $\{\boldsymbol{y}_n\}$ serán linealmente separables si existe un vector $\hat{\boldsymbol{w}}$ y un escalar w_0 tales que para todo $\boldsymbol{x}_n$, $\hat{\boldsymbol{w}}^T \boldsymbol{x}_n + w_0 > 0$, y para todo $\boldsymbol{y}_n$, $\hat{\boldsymbol{w}}^T \boldsymbol{y}_n + w_0 < 0$. Demuestre que si sus envolventes convexas tienen intersección no nula, entonces los dos conjuntos de puntos no pueden ser linealmente separables y *vice versa*: si son linealmente separables entonces sus envolventes convexas tienen interesección nula.

17SO

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.

2.33 1. Considere la distribución de probabilidad conjunta $p(\boldsymbol{x}, \boldsymbol{y})$ definida por las distribuciones marginal y condicional siguientes:

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

donde $\boldsymbol{\mu}$ y $\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ son las medias de la distribución gaussiana, y $\boldsymbol{\Lambda}$ y $\boldsymbol{L}$ son las matrices de precisión. Utilice la técnica de completar el cuadrado para obtener una expresión analítica de la media y covarianza de $p(\boldsymbol{x}|\boldsymbol{y})$.

4.10 2. 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

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

17SR

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.

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

3.2 2. Sea $t = \boldsymbol{w}^T \boldsymbol{\phi}(\boldsymbol{x}) + \epsilon$;
t, un variable aleatoria que toma valores en $\mathbb{R}$;
$\boldsymbol{x}$, un vector de $\mathbb{R}^{in}$;
$\boldsymbol{\phi} = (1, \phi_1(\boldsymbol{x}), ..., \phi_{M-1}(\boldsymbol{x}))^T$ con $\phi_i(\cdot)$ $(i = 1, ..., M-1)$ un conjunto de $M - 1$ transformaciones (no necesariamente lineales) escalares sobre $\mathbb{R}^{in}$;
$\boldsymbol{w} = (w_0, w_1, ..., w_{M-1})^T$, un vector de $\mathbb{R}^M$; y, finalmente, ϵ una variable aleatoria distribuida según una gaussiana $\mathcal{N}(0, \beta^{-1})$. Sea $(\boldsymbol{x}_1, ..., \boldsymbol{x}_N)$ y $(t_1, ..., t_N)$ un conjunto de valores de $\boldsymbol{x}$ y t obtenidos a partir de las distribuciones descritas por todo lo anterior.

Defínase $\boldsymbol{\Phi}$ como la matriz de dimensiones $N \times M$ en que la fila j es igual a $(1, \phi_1(\boldsymbol{x}_j), ..., \phi_{M-1}(\boldsymbol{x}_j))$.

Demuestre que la matriz $\Phi(\Phi^T\Phi)^{-1}\Phi^T$ proyecta cualquier vector v en el espacio generado por los vectores columna de ϕ. Utilice este resultado para probar que la solución de mínimos cuadrados dada por $w = (\Phi^T\Phi)^{-1}\Phi^T t$ corresponde a una proyección ortogonal del vector t sobre la variedad (manifold en inglés) lineal S definida por los vectores columna de $\boldsymbol{\Phi}$, $\psi_k = (\phi_k(\boldsymbol{x}_1), ..., \phi_k(\boldsymbol{x}_N))$ como se muestra en la Fig. 1.

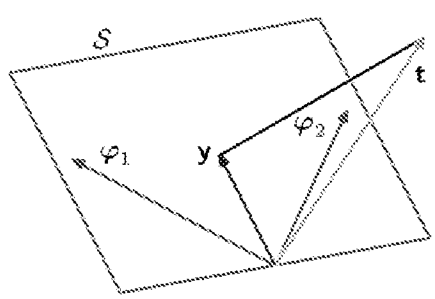

Figura 1:

4.1 3. Dado un conjunto de puntos $\{\boldsymbol{x}_n\}$, podemos definir su envolvente convexa (*convex hull*) como el conjunto de todos los puntos $\boldsymbol{x}$ tales que

$$\boldsymbol{x} = \sum_n \alpha_n \boldsymbol{x}_n$$

donde $\alpha_n \geq 0$ y $\sum\limits_n \alpha_n = 1$. Considere un segundo conjunto de puntos $\boldsymbol{y}_n$ y su envolvente convexa. Por definición, los dos conjuntos de puntos $\{\boldsymbol{x}_n\}$ y $\{\boldsymbol{y}_n\}$ serán linealmente separables si existe un vector $\hat{\boldsymbol{w}}$ y un escalar w_0 tales que para todo $\boldsymbol{x}_n$, $\hat{\boldsymbol{w}}^T \boldsymbol{x}_n + w_0 > 0$, y para todo $\boldsymbol{y}_n$, $\hat{\boldsymbol{w}}^T \boldsymbol{y}_n + w_0 < 0$. Demuestre que si sus envolventes convexas tienen intersección no nula, entonces los dos conjuntos de puntos no pueden ser linealmente separables y *vice versa*: si son linealmente separables entonces sus envolventes convexas tienen interesección nula.

16F1

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.

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

- $\mathbb{E}[x] = \mathbb{E}_y[\mathbb{E}_x[x|y]]$
- $var[x] = \mathbb{E}_y[var_x[x|y]] + var_y[\mathbb{E}_x[x|y]]$

donde $\mathbb{E}_x[x|y]$ 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.

3.16 2. Deduzca la ecuación para (el logaritmo de) la evidencia de un modelo de regresión lineal con hiperparámetros α y β

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

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.

4.1 3. Dado un conjunto de puntos $\{\boldsymbol{x}_n\}$, podemos definir su envolvente convexa (*convex hull*) como el conjunto de todos los puntos $\boldsymbol{x}$ tales que

$$\boldsymbol{x} = \sum_n \alpha_n \boldsymbol{x}_n$$

donde $\alpha_n \geq 0$ y $\sum_n \alpha_n = 1$. Considere un segundo conjunto de puntos $\boldsymbol{y}_n$ y su envolvente convexa. Por definición, los dos conjuntos de puntos $\{\boldsymbol{x}_n\}$ y $\{\boldsymbol{y}_n\}$ serán linealmente separables si existe un vector $\hat{\boldsymbol{w}}$ y un escalar w_0 tales que para todo $\boldsymbol{x}_n$, $\hat{\boldsymbol{w}}^T\boldsymbol{x}_n + w_0 > 0$, y para todo $\boldsymbol{y}_n$, $\hat{\boldsymbol{w}}^T\boldsymbol{y}_n + w_0 < 0$. Demuestre que si sus envolventes convexas tienen intersección no nula, entonces los dos conjuntos de puntos no pueden ser linealmente separables y *vice versa*: si son linealmente separables entonces sus envolventes convexas tienen interesección nula.

16F2

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.

1.32 1. Considere un vector $\boldsymbol{x}$ de variables aleatorias contínuas, que sigue una distribución de probabilidad dada $p(\boldsymbol{x})$ y entropía asociada $H[\boldsymbol{x}]$. Suponga que sobre el vector $\boldsymbol{x}$ se aplica una transformación lineal no singular que transforma $\boldsymbol{x}$ en $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$. Demuestre que la entropía correspondiente al vector $\boldsymbol{y}$ viene dada por $H[\boldsymbol{y}] = H[\boldsymbol{x}] + ln|\boldsymbol{A}|$, donde $|\boldsymbol{A}|$ denota el determinante de $\boldsymbol{A}$.

3.2 2. Sea $t = \boldsymbol{w}^T\boldsymbol{\phi}(\boldsymbol{x}) + \epsilon$;
t, un variable aleatoria que toma valores en $\mathbb{R}$;
$\boldsymbol{x}$, un vector de $\mathbb{R}^{in}$;
$\boldsymbol{\phi} = (1, \phi_1(\boldsymbol{x}), ..., \phi_{M-1}(\boldsymbol{x}))^T$ con $\phi_i(\cdot)$ $(i = 1, ..., M-1)$ un conjunto de $M-1$ transformaciones (no necesariamente lineales) escalares sobre $\mathbb{R}^{in}$;
$\boldsymbol{w} = (w_0, w_1, ..., w_{M-1})^T$, un vector de $\mathbb{R}^M$; y, finalmente, ϵ una variable aleatoria distribuida según una gaussiana $\mathcal{N}(0, \beta^{-1})$. Sea $(\boldsymbol{x}_1, ..., \boldsymbol{x}_N)$ y $(t_1, ..., t_N)$ un conjunto de valores de $\boldsymbol{x}$ y t obtenidos a partir de las distribuciones descritas por todo lo anterior.

Defínase $\boldsymbol{\Phi}$ como la matriz de dimensiones $N \times M$ en que la fila j es igual a $(1, \phi_1(\boldsymbol{x}_j), ..., \phi_{M-1}(\boldsymbol{x}_j))$.

Demuestre que la matriz $\Phi(\Phi^T\Phi)^{-1}\Phi^T$ proyecta cualquier vector v en el espacio generado por los vectores columna de ϕ. Utilice este resultado para probar que la solución de mínimos cuadrados dada por $w = (\Phi^T\Phi)^{-1}\Phi^T t$ corresponde a una proyección ortogonal del vector t sobre la variedad (manifold en inglés) lineal S definida por los vectores columna de $\boldsymbol{\Phi}$, $\psi_k = (\phi_k(\boldsymbol{x}_1), ..., \phi_k(\boldsymbol{x}_N))$ como se muestra en la Fig. 1.

4.1 3. Dado un conjunto de puntos $\{\boldsymbol{x}_n\}$, podemos definir su envolvente convexa (*convex hull*) como el conjunto de todos los puntos $\boldsymbol{x}$ tales que

$$\boldsymbol{x} = \sum_n \alpha_n \boldsymbol{x}_n$$

donde $\alpha_n \geq 0$ y $\sum\limits_n \alpha_n = 1$. Considere un segundo conjunto de puntos $\boldsymbol{y}_n$ y su envolvente convexa. Por definición, los dos conjuntos de puntos $\{\boldsymbol{x}_n\}$ y $\{\boldsymbol{y}_n\}$ serán linealmente separables si existe un vector $\hat{\boldsymbol{w}}$ y un escalar w_0 tales que para todo $\boldsymbol{x}_n$, $\hat{\boldsymbol{w}}^T\boldsymbol{x}_n + w_0 > 0$, y para todo $\boldsymbol{y}_n$, $\hat{\boldsymbol{w}}^T\boldsymbol{y}_n + w_0 < 0$.

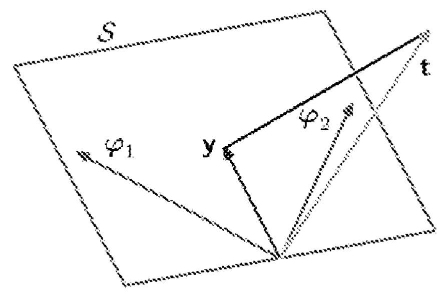

Figura 1:

Demuestre que si sus envolventes convexas tienen intersección no nula, entonces los dos conjuntos de puntos no pueden ser linealmente separables y *vice versa*: si son linealmente separables entonces sus envolventes convexas tienen interesección nula.

16SO

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.

2.33 1. Considere la distribución de probabilidad conjunta $p(\boldsymbol{x}, \boldsymbol{y})$ definida por las distribuciones marginal y condicional siguientes:

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

donde $\boldsymbol{\mu}$ y $\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ son las medias de la distribución gaussiana, y $\boldsymbol{\Lambda}$ y $\boldsymbol{L}$ son las matrices de precisión. Utilice la técnica de completar el cuadrado para obtener una expresión analítica de la media y covarianza de $p(\boldsymbol{x}|\boldsymbol{y})$.

3.11 2. Utilice la identidad matricial

$$(\boldsymbol{M} + \boldsymbol{v}\boldsymbol{v}^T)^{-1} = \boldsymbol{M}^{-1} - \frac{(\boldsymbol{M}^{-1}\boldsymbol{v})(\boldsymbol{v}^T\boldsymbol{M}^{-1})}{1 + \boldsymbol{v}^T\boldsymbol{M}^{-1}\boldsymbol{v}}$$

y suponga que

$$\boldsymbol{S}_{N+1}^{-1} = \boldsymbol{S}_N^{-1} + \beta\boldsymbol{\phi}_{N+1}\boldsymbol{\phi}_{N+1}^T \tag{1}$$

para demostrar que la incertidumbre de un modelo predictivo bayesiano construido a partir de N observaciones, y dada por

$$\sigma_N^2(\boldsymbol{x}) = \frac{1}{\beta} + \boldsymbol{\phi}(\boldsymbol{x})^T\boldsymbol{S}_N\boldsymbol{\phi}(\boldsymbol{x}) \tag{2}$$

(con la interpretación habitual de los símbolos utilizados en el texto base) satisface que $\sigma_{N+1}^2(\boldsymbol{x}) \leq \sigma_N^2(\boldsymbol{x})$.

4.8 3. Sabiendo que en un problema de clasificación con dos clases

$$p(\mathcal{C}_1|\boldsymbol{x}) = \frac{1}{1 + \exp(-\ln\frac{p(\boldsymbol{x}|\mathcal{C}_1)p(\mathcal{C}_1)}{p(\boldsymbol{x}|\mathcal{C}_2)p(\mathcal{C}_2)})} = \frac{1}{1 + \exp(-a)} = \sigma(a) \tag{3}$$

y suponiendo un modelo generativo en el que las verosimilitudes (likelihoods) de las dos clases vienen dadas por dos gaussianas de medias $\boldsymbol{\mu}_1$ y $\boldsymbol{\mu}_2$, pero la misma varianza $\boldsymbol{\Sigma}$, demuestre que

$$p(\mathcal{C}_1|\boldsymbol{x}) = \sigma(\boldsymbol{w}^T\boldsymbol{x} + w_0) \tag{4}$$

con $\boldsymbol{w} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ y

$$w_0 = -\frac{1}{2}\boldsymbol{\mu}_1^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1 + \frac{1}{2}\boldsymbol{\mu}_2^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_2 + \ln\frac{p(\mathcal{C}_1)}{p(\mathcal{C}_2)}. \tag{5}$$

16SR

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.

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

3.16 2. Deduzca la ecuación para (el logaritmo de) la evidencia de un modelo de regresión lineal con hiperparámetros α y β

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

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.

4.10 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

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

15F1

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.

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

- $\mathbb{E}[x] = \mathbb{E}_y[\mathbb{E}_x[x|y]]$
- $var[x] = \mathbb{E}_y[var_x[x|y]] + var_y[\mathbb{E}_x[x|y]]$

donde $\mathbb{E}_x[x|y]$ 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.

3.16 2. Deduzca la ecuación para (el logaritmo de) la evidencia de un modelo de regresión lineal con hiperparámetros α y β

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

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

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

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

4.1 3. Dado un conjunto de puntos $\{\mathbf{x}_n\}$, podemos definir su envolvente convexa (*convex hull*) como el conjunto de todos los puntos $\mathbf{x}$ tales que

$$\mathbf{x} = \sum_n \alpha_n \mathbf{x}_n$$

donde $\alpha_n \geq 0$ y $\sum_n \alpha_n = 1$. Considere un segundo conjunto de puntos $\mathbf{y}_n$ y su envolvente convexa. Por definición, los dos conjuntos de puntos $\{\mathbf{x}_n\}$ y $\{\mathbf{y}_n\}$ serán linealmente separables si existe un vector $\hat{\mathbf{w}}$ y un escalar w_0 tales que para todo $\mathbf{x}_n$, $\hat{\mathbf{w}}^T\mathbf{x}_n + w_0 > 0$, y para todo $\mathbf{y}_n$, $\hat{\mathbf{w}}^T\mathbf{y}_n + w_0 < 0$. Demuestre que si sus envolventes convexas tienen intersección no nula, entonces los dos conjuntos de puntos no pueden ser linealmente separables y *vice versa*: si son linealmente separables entonces sus envolventes convexas tienen interesección nula.

15F2

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.

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

3.2 2. Sea $t = \boldsymbol{w}^T \boldsymbol{\phi}(\boldsymbol{x}) + \epsilon$;
t, un variable aleatoria que toma valores en $\mathbb{R}$;
$\boldsymbol{x}$, un vector de $\mathbb{R}^{in}$;
$\boldsymbol{\phi} = (1, \phi_1(\boldsymbol{x}), ..., \phi_{M-1}(\boldsymbol{x}))^T$ con $\phi_i(\cdot)$ $(i = 1, ..., M-1)$ un conjunto de $M-1$ transformaciones (no necesariamente lineales) escalares sobre $\mathbb{R}^{in}$; $\boldsymbol{w} = (w_0, w_1, ..., w_{M-1})^T$, un vector de $\mathbb{R}^M$; y, finalmente, ϵ una variable aleatoria distribuida según una gaussiana $\mathcal{N}(0, \beta^{-1})$. Sea $(\boldsymbol{x}_1, ..., \boldsymbol{x}_N)$ y $(t_1, ..., t_N)$ un conjunto de valores de $\boldsymbol{x}$ y t obtenidos a partir de las distribuciones descritas por todo lo anterior.

Defínase $\boldsymbol{\Phi}$ como la matriz de dimensiones $N \times M$ en que la fila j es igual a $(1, \phi_1(\boldsymbol{x}_j), ..., \phi_{M-1}(\boldsymbol{x}_j))$.

Demuestre que la matriz $\Phi(\Phi^T\Phi)^{-1}\Phi^T$ proyecta cualquier vector v en el espacio generado por los vectores columna de ϕ. Utilice este resultado para probar que la solución de mínimos cuadrados dada por $w = (\Phi^T\Phi)^{-1}\Phi^T t$ corresponde a una proyección ortogonal del vector t sobre la variedad (manifold en inglés) lineal S definida por los vectores columna de $\boldsymbol{\Phi}$, $\psi_k = (\phi_k(\boldsymbol{x}_1), ..., \phi_k(\boldsymbol{x}_N))$ como se muestra en la Fig. 1.

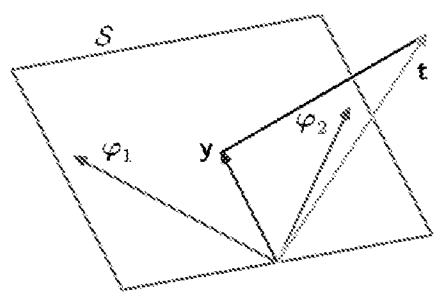

Figura 1:

4.1 3. Dado un conjunto de puntos $\{\boldsymbol{x}_n\}$, podemos definir su envolvente convexa (*convex hull*) como el conjunto de todos los puntos $\boldsymbol{x}$ tales que

$$\boldsymbol{x} = \sum_n \alpha_n \boldsymbol{x}_n$$

donde $\alpha_n \geq 0$ y $\sum\limits_n \alpha_n = 1$. Considere un segundo conjunto de puntos $\boldsymbol{y}_n$ y su envolvente convexa. Por definición, los dos conjuntos de puntos $\{\boldsymbol{x}_n\}$ y $\{\boldsymbol{y}_n\}$ serán linealmente separables si existe un vector $\hat{\boldsymbol{w}}$ y un escalar w_0 tales que para todo $\boldsymbol{x}_n$, $\hat{\boldsymbol{w}}^T \boldsymbol{x}_n + w_0 > 0$, y para todo $\boldsymbol{y}_n$, $\hat{\boldsymbol{w}}^T \boldsymbol{y}_n + w_0 < 0$. Demuestre que si sus envolventes convexas tienen intersección no nula, entonces los dos conjuntos de puntos no pueden ser linealmente separables y *vice versa*: si son linealmente separables entonces sus envolventes convexas tienen interesección nula.

15SO

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.

2.33 1. Considere la distribución de probabilidad conjunta $p(\boldsymbol{x}, \boldsymbol{y})$ definida por las distribuciones marginal y condicional siguientes:

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

donde $\boldsymbol{\mu}$ y $\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ son las medias de la distribución gaussiana, y $\boldsymbol{\Lambda}$ y $\boldsymbol{L}$ son las matrices de precisión. Utilice la técnica de completar el cuadrado para obtener una expresión analítica de la media y covarianza de $p(\boldsymbol{x}|\boldsymbol{y})$.

3.11 2. Utilice la identidad matricial

$$(\boldsymbol{M} + \boldsymbol{v}\boldsymbol{v}^T)^{-1} = \boldsymbol{M}^{-1} - \frac{(\boldsymbol{M}^{-1}\boldsymbol{v})(\boldsymbol{v}^T\boldsymbol{M}^{-1})}{1 + \boldsymbol{v}^T\boldsymbol{M}^{-1}\boldsymbol{v}}$$

para demostrar que la incertidumbre de un modelo predictivo bayesiano construido a partir de N observaciones, y dada por

$$\sigma_N^2(\boldsymbol{x}) = \frac{1}{\beta} + \boldsymbol{\phi}(\boldsymbol{x})^T\boldsymbol{S}_N\boldsymbol{\phi}(\boldsymbol{x}) \tag{1}$$

(con la interpretación habitual de los símbolos utilizados en el texto base) satisface que $\sigma_{N+1}^2(\boldsymbol{x}) \leq \sigma_N^2(\boldsymbol{x})$.

4.8 3. Sabiendo que en un problema de clasificación con dos clases

$$p(\mathcal{C}_1|\boldsymbol{x}) = \frac{1}{1 + \exp(-\ln\frac{p(\boldsymbol{x}|\mathcal{C}_1)p(\mathcal{C}_1)}{p(\boldsymbol{x}|\mathcal{C}_2)p(\mathcal{C}_2)})} = \frac{1}{1 + \exp(-a)} = \sigma(a) \tag{2}$$

y suponiendo un modelo generativo en el que las verosimilitudes (likelihoods) de las dos clases vienen dadas por dos gaussianas de medias $\boldsymbol{\mu}_1$ y $\boldsymbol{\mu}_2$, pero la misma varianza $\boldsymbol{\Sigma}$, demuestre que

$$p(\mathcal{C}_1|\boldsymbol{x}) = \sigma(\boldsymbol{w}^T\boldsymbol{x} + w_0) \tag{3}$$

con $\boldsymbol{w} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ y

$$w_0 = -\frac{1}{2}\boldsymbol{\mu}_1^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1 + \frac{1}{2}\boldsymbol{\mu}_2^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_2 + \ln\frac{p(\mathcal{C}_1)}{p(\mathcal{C}_2)}. \tag{4}$$

15SR

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.

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

3.16 2. Deduzca la ecuación para (el logaritmo de) la evidencia de un modelo de regresión lineal con hiperparámetros α y β

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

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.

4.10 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

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

14F1

Para la realización del examen se permite un único libro (sea cual sea) 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.

1.32 1. Considere un vector $\boldsymbol{x}$ de variables aleatorias contínuas, que sigue una distribución de probabilidad dada $p(\boldsymbol{x})$ y entropía asociada $H[\boldsymbol{x}]$. Suponga que sobre el vector $\boldsymbol{x}$ se aplica una transformación lineal no singular que transforma $\boldsymbol{x}$ en $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$. Demuestre que la entropía correspondiente al vector $\boldsymbol{y}$ viene dada por $H[\boldsymbol{y}] = H[\boldsymbol{x}] + ln|\boldsymbol{A}|$, donde $|\boldsymbol{A}|$ denota el determinante de $\boldsymbol{A}$.

3.2 2. Sea $t = \boldsymbol{w}^T \boldsymbol{\phi}(\boldsymbol{x}) + \epsilon$;
t, un variable aleatoria que toma valores en $\mathbb{R}$;
$\boldsymbol{x}$, un vector de $\mathbb{R}^{in}$;
$\boldsymbol{\phi} = (1, \phi_1(\boldsymbol{x}), ..., \phi_{M-1}(\boldsymbol{x}))^T$ con $\phi_i(\cdot)$ $(i = 1, ..., M-1)$ un conjunto de $M-1$ transformaciones (no necesariamente lineales) escalares sobre $\mathbb{R}^{in}$;
$\boldsymbol{w} = (w_0, w_1, ..., w_{M-1})^T$, un vector de $\mathbb{R}^M$; y, finalmente, ϵ una variable aleatoria distribuida según una gaussiana $\mathcal{N}(0, \beta^{-1})$. Sea $(\boldsymbol{x}_1, ..., \boldsymbol{x}_N)$ y $(t_1, ..., t_N)$ un conjunto de valores de $\boldsymbol{x}$ y t obtenidos a partir de las distribuciones descritas por todo lo anterior.

Defínase $\boldsymbol{\Phi}$ como la matriz de dimensiones $N \times M$ en que la fila j es igual a $(1, \phi_1(\boldsymbol{x}_j), ..., \phi_{M-1}(\boldsymbol{x}_j))$.

Demuestre que la matriz $\Phi(\Phi^T\Phi)^{-1}\Phi^T$ proyecta cualquier vector v en el espacio generado por los vectores columna de ϕ. Utilice este resultado para probar que la solución de mínimos cuadrados dada por $w = (\Phi^T\Phi)^{-1}\Phi^T t$ corresponde a una proyección ortogonal del vector t sobre la variedad (manifold en inglés) lineal S definida por los vectores columna de $\boldsymbol{\Phi}$, $\psi_k = (\phi_k(\boldsymbol{x}_1), ..., \phi_k(\boldsymbol{x}_N))$ como se muestra en la Fig. 1.

4.1 3. Dado un conjunto de puntos $\{\boldsymbol{x}_n\}$, podemos definir su envolvente convexa (*convex hull*) como el conjunto de todos los puntos $\boldsymbol{x}$ tales que

$$\boldsymbol{x} = \sum_n \alpha_n \boldsymbol{x}_n$$

donde $\alpha_n \geq 0$ y $\sum_n \alpha_n = 1$. Considere un segundo conjunto de puntos $\boldsymbol{y}_n$ y su envolvente convexa. Por definición, los dos conjuntos de puntos $\{\boldsymbol{x}_n\}$ y $\{\boldsymbol{y}_n\}$ serán linealmente separables si existe un vector $\hat{\boldsymbol{w}}$ y un escalar w_0 tales que para todo $\boldsymbol{x}_n$, $\hat{\boldsymbol{w}}^T\boldsymbol{x}_n + w_0 > 0$, y para todo $\boldsymbol{y}_n$, $\hat{\boldsymbol{w}}^T\boldsymbol{y}_n + w_0 < 0$. Demuestre que si sus envolventes convexas tienen intersección no nula, entonces los dos conjuntos de puntos no pueden ser linealmente separables y *vice versa*: si son linealmente separables entonces sus envolventes convexas tienen interesección nula.

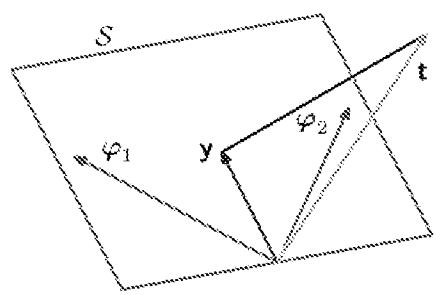

Figura 1:

14F2

Para la realización del examen se permite un único libro (sea cual sea) 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.

2.33 1. Considere la distribución de probabilidad conjunta $p(\boldsymbol{x}, \boldsymbol{y})$ definida por las distribuciones marginal y condicional siguientes:

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

donde $\boldsymbol{\mu}$ y $\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ son las medias de la distribución gaussiana, y $\boldsymbol{\Lambda}$ y $\boldsymbol{L}$ son las matrices de precisión. Utilice la técnica de completar el cuadrado para obtener una expresión analítica de la media y covarianza de $p(\boldsymbol{x}|\boldsymbol{y})$.

3.11 2. Utilice la identidad matricial

$$(\boldsymbol{M} + \boldsymbol{v}\boldsymbol{v}^T)^{-1} = \boldsymbol{M}^{-1} - \frac{(\boldsymbol{M}^{-1}\boldsymbol{v})(\boldsymbol{v}^T\boldsymbol{M}^{-1})}{1 + \boldsymbol{v}^T\boldsymbol{M}^{-1}\boldsymbol{v}}$$

para demostrar que la incertidumbre de un modelo predictivo bayesiano construido a partir de N observaciones, y dada por

$$\sigma_N^2(\boldsymbol{x}) = \frac{1}{\beta} + \boldsymbol{\phi}(\boldsymbol{x})^T\boldsymbol{S}_N\boldsymbol{\phi}(\boldsymbol{x}) \tag{1}$$

(con la interpretación habitual de los símbolos utilizados en el texto base) satisface que $\sigma_{N+1}^2(\boldsymbol{x}) \leq \sigma_N^2(\boldsymbol{x})$.

4.8 3. Sabiendo que en un problema de clasificación con dos clases

$$p(\mathcal{C}_1|\boldsymbol{x}) = \frac{1}{1 + \exp(-\ln\frac{p(\boldsymbol{x}|\mathcal{C}_1)p(\mathcal{C}_1)}{p(\boldsymbol{x}|\mathcal{C}_2)p(\mathcal{C}_2)})} = \frac{1}{1 + \exp(-a)} = \sigma(a) \tag{2}$$

y suponiendo un modelo generativo en el que las verosimilitudes (likelihoods) de las dos clases vienen dadas por dos gaussianas de medias $\boldsymbol{\mu}_1$ y $\boldsymbol{\mu}_2$, pero la misma varianza $\boldsymbol{\Sigma}$, demuestre que

$$p(\mathcal{C}_1|\boldsymbol{x}) = \sigma(\boldsymbol{w}^T\boldsymbol{x} + w_0) \tag{3}$$

con $\boldsymbol{w} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ y

$$w_0 = -\frac{1}{2}\boldsymbol{\mu}_1^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1 + \frac{1}{2}\boldsymbol{\mu}_2^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_2 + \ln\frac{p(\mathcal{C}_1)}{p(\mathcal{C}_2)}. \tag{4}$$

Para la realización del examen se permite un único libro (sea cual sea) 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.

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

- $\mathbb{E}[x] = \mathbb{E}_y[\mathbb{E}_x[x|y]]$
- $var[x] = \mathbb{E}_y[var_x[x|y]] + var_y[\mathbb{E}_x[x|y]]$

donde $\mathbb{E}_x[x|y]$ 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.

3.16 2. Deduzca la ecuación para (el logaritmo de) la evidencia de un modelo de regresión lineal con hiperparámetros α y β

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

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.

4.1 3. Dado un conjunto de puntos $\{\boldsymbol{x}_n\}$, podemos definir su envolvente convexa (*convex hull*) como el conjunto de todos los puntos $\boldsymbol{x}$ tales que

$$\boldsymbol{x} = \sum_n \alpha_n \boldsymbol{x}_n$$

donde $\alpha_n \geq 0$ y $\sum_n \alpha_n = 1$. Considere un segundo conjunto de puntos $\boldsymbol{y}_n$ y su envolvente convexa. Por definición, los dos conjuntos de puntos $\{\boldsymbol{x}_n\}$ y $\{\boldsymbol{y}_n\}$ serán linealmente separables si existe un vector $\hat{\boldsymbol{w}}$ y un escalar w_0 tales que para todo $\boldsymbol{x}_n$, $\hat{\boldsymbol{w}}^T\boldsymbol{x}_n + w_0 > 0$, y para todo $\boldsymbol{y}_n$, $\hat{\boldsymbol{w}}^T\boldsymbol{y}_n + w_0 < 0$. Demuestre que si sus envolventes convexas tienen intersección no nula, entonces los dos conjuntos de puntos no pueden ser linealmente separables y *vice versa*: si son linealmente separables entonces sus envolventes convexas tienen interesección nula.

14SR

Para la realización del examen se permite un único libro (sea cual sea) 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.

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

3.2 2. Sea $t = \boldsymbol{w}^T \boldsymbol{\phi}(\boldsymbol{x}) + \epsilon$;
t, un variable aleatoria que toma valores en $\mathbb{R}$;
$\boldsymbol{x}$, un vector de $\mathbb{R}^{in}$;
$\boldsymbol{\phi} = (1, \phi_1(\boldsymbol{x}), ..., \phi_{M-1}(\boldsymbol{x}))^T$ con $\phi_i(\cdot)$ $(i = 1, ..., M-1)$ un conjunto de $M-1$ transformaciones (no necesariamente lineales) escalares sobre $\mathbb{R}^{in}$;
$\boldsymbol{w} = (w_0, w_1, ..., w_{M-1})^T$, un vector de $\mathbb{R}^M$; y, finalmente, ϵ una variable aleatoria distribuida según una gaussiana $\mathcal{N}(0, \beta^{-1})$. Sea $(\boldsymbol{x}_1, ..., \boldsymbol{x}_N)$ y $(t_1, ..., t_N)$ un conjunto de valores de $\boldsymbol{x}$ y t obtenidos a partir de las distribuciones descritas por todo lo anterior.

Defínase $\boldsymbol{\Phi}$ como la matriz de dimensiones $N \times M$ en que la fila j es igual a $(1, \phi_1(\boldsymbol{x}_j), ..., \phi_{M-1}(\boldsymbol{x}_j))$.

Demuestre que la matriz $\Phi(\Phi^T\Phi)^{-1}\Phi^T$ proyecta cualquier vector v en el espacio generado por los vectores columna de ϕ. Utilice este resultado para probar que la solución de mínimos cuadrados dada por $w = (\Phi^T\Phi)^{-1}\Phi^T t$ corresponde a una proyección ortogonal del vector t sobre la variedad (manifold en inglés) lineal S definida por los vectores columna de $\boldsymbol{\Phi}$, $\psi_k = (\phi_k(\boldsymbol{x}_1), ..., \phi_k(\boldsymbol{x}_N))$ como se muestra en la Fig. 1.

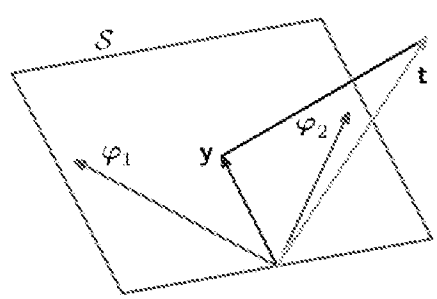

Figura 1:

4.1 3. Dado un conjunto de puntos $\{\boldsymbol{x}_n\}$, podemos definir su envolvente convexa (*convex hull*) como el conjunto de todos los puntos $\boldsymbol{x}$ tales que

$$\boldsymbol{x} = \sum_n \alpha_n \boldsymbol{x}_n$$

donde $\alpha_n \geq 0$ y $\sum_n \alpha_n = 1$. Considere un segundo conjunto de puntos $\boldsymbol{y}_n$ y su envolvente convexa. Por definición, los dos conjuntos de puntos $\{\boldsymbol{x}_n\}$ y $\{\boldsymbol{y}_n\}$ serán linealmente separables si existe un vector $\hat{\boldsymbol{w}}$ y un escalar w_0 tales que para todo $\boldsymbol{x}_n$, $\hat{\boldsymbol{w}}^T \boldsymbol{x}_n + w_0 > 0$, y para todo $\boldsymbol{y}_n$, $\hat{\boldsymbol{w}}^T \boldsymbol{y}_n + w_0 < 0$. Demuestre que si sus envolventes convexas tienen intersección no nula, entonces los dos conjuntos de puntos no pueden ser linealmente separables y *vice versa*: si son linealmente separables entonces sus envolventes convexas tienen interesección nula.