Curso: Machine Learning para Físicos (Temas selectos de física computacional 1)
Prof.: Dr. Huziel E. Sauceda
Ayudantes: I. Jessica Martínez Marcelo, Noé O. Rodríguez Rodríguez.

Tarea 1
Entrega: Jueves 17 de Marzo

Textos de referencia:
- Bishop, *Pattern Recognition and Machine Learning*, Springer 2006.
- Duda, Hart, y Stork, *Pattern Recognition*, Second Ed.

1. Ejercicio 1.4 del libro de Bishop.

2. Ejercicio 1.7 del libro de Bishop.

3. Ejercicio 47 del Capítulo 2 del libro de Duda.

4. Considere el conjunto de vectores S y la matriz Σ. Grafique las los vectores $\mathbf{x}$ y los vectores $\Sigma\mathbf{x}$ y $\Sigma^{-1}\mathbf{x}$.

$$S = \left\{ \mathbf{x} = \begin{bmatrix} \cos(\frac{2\pi}{8}n) \\ \sin(\frac{2\pi}{8}n) \end{bmatrix}; n = 0,\ldots,7 \right\}, \Sigma = \begin{bmatrix} 2.5 & -2 \\ -2 & 3 \end{bmatrix}.$$

5. **Estimación del error de Bayes.** La regla de decisión de Bayes para el problema de clasificación de dos clases da como resultado el error de Bayes:

$$P(\text{Err}) = \int P(\text{Err}\,|\,\mathbf{x})P(\mathbf{x})d\mathbf{x}$$

donde $P(\text{Err}\,|\,\mathbf{x}) = \min\{P(\omega_2\,|\,\mathbf{x}), P(\omega_1\,|\,\mathbf{x})\}$ es la probabilidad de error para un vector de características $\mathbf{x}$. Curiosamente, mientras que las clases a posteriori $P(\omega_1\,|\,\mathbf{x})$ y $P(\omega_2\,|\,\mathbf{x})$ a menudo se pueden expresar analíticamente y son integrables, la función de error tiene discontinuidades que impiden su integración analítica y, por lo tanto, también impiden el cálculo directo del error de Bayes.

(a). Muestre que el error completo puede ser acotado superiormente de la siguiente manera:

$$P(\text{Err}) \leq \int \frac{2}{\frac{1}{P(\omega_1|\mathbf{x})} + \frac{1}{P(\omega_2|\mathbf{x})}} P(\mathbf{x}) d\mathbf{x}$$

Tenga en cuenta que el integrando ahora es continuo y corresponde al promedio armónico de las clases a posteriori pesada por $P(\mathbf{x})$.

(b). Muestre usando el resultado **(a)** que para las distribuciones de probabilidad de una variable,

$$P(x|\omega_1) = \frac{\pi^{-1}}{1 + (x-\mu)^2} \quad \text{y} \quad P(x|\omega_2) = \frac{\pi^{-1}}{1 + (x+\mu)^2}$$

el error de Bayes puede ser acotado superiormente por:

$$P(\text{Err}) \leq \frac{2P(\omega_1)P(\omega_2)}{\sqrt{1 + 4\mu^2 P(\omega_1)P(\omega_2)}}.$$

(Sugerencia: se puede usar la identidad $\int \frac{1}{ax^2 + bx + c} dx = \frac{2\pi}{\sqrt{4ac - b^2}}$, para $b^2 < 4ac$)

6. Fronteras de decisión de Bayes. Se podría especular que, en algunos casos, los datos generados $P(\mathbf{x}|\omega_1)$ y $P(\mathbf{x}|\omega_2)$ no sirven para mejorar la precisión de un clasificador, en cuyo caso solo se debe confiar en las probabilidades a priori de cada clase $P(\omega_1)$ y $P(\omega_2)$ asumiendo aquí que son estrictamente positivas.

   Para la primera parte de este ejercicio, asumimos que los datos para cada clase son generados por las distribuciones de probabilidad laplacianas de una variable:

$$P(x|\omega_1) = \frac{1}{2\sigma} \exp(-|x-\mu|/\sigma) \quad \text{y} \quad P(x|\omega_2) = \frac{1}{2\sigma} \exp(-|x+\mu|/\sigma),$$

Donde $\mu, \sigma > 0$.

(a) Determine para qué valores de $P(\omega_1)$, $P(\omega_2)$ y de μ, σ la decisión óptima es predecir siempre la primera clase, ω_1, es decir, bajo qué condiciones $P(\text{Err} \mid x) = P(\omega_2 \mid x)$ $\forall\ x \in \mathbb{R}$.

(b) Repita el ejercicio para el caso donde los datos para cada clase son generados por una distribuciones de probabilidad Gaussiana en una dimension:

$$P(x \mid \omega_1) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) \quad \text{y} \quad P(x \mid \omega_2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x+\mu)^2}{2\sigma^2}\right),$$

Donde $\mu, \sigma > 0$.