Problem set 2

Clasificador Bayesiano, Clasificador bayesiano de distancia mínima
Aprendizaje Automático

1. P1

Obtenga la expresión analítica de la frontera de desición bayesiana para dos funciones de distribución gaussianas equiprobables con las siguientes características: $\boldsymbol{\mu_1} = [3\ 6]^t$, $\Sigma_1 = \begin{bmatrix} 1/2 & 0 \\ 0 & 2 \end{bmatrix}$, $\boldsymbol{\mu_2} = [3\ -2]^t$, y $\Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$.

2. P2

Escriba una función en Matlab que tome como entradas:

(a) Los vectores de medias.

(b) Las matrices de covarianzas de las distribuciones.

(c) Las probabilidades *a priori* de cada clase.

(d) Una matriz X con n vectores a clasificar.

La salida será un vector $\hat{Y}$ de n elementos que indica la clase enla que cada vector x de entrada fue clasificado, de acuerdo con la regla general de clasificación bayesiana.

3. P3

Considere dos clases equiprobables en un problema de clasificación bidimensional donde los vectores de características en cada una de las clases ω_1 y ω_2 están distribuidos de acuerdo con:

$$p(x|\omega_1) = \frac{1}{2\pi\sigma_1^2}\exp(-\frac{1}{2\sigma_1^2}(x-\mu_1)^t(x-\mu_1))$$
$$p(x|\omega_2) = \frac{1}{2\pi\sigma_2^2}\exp(-\frac{1}{2\sigma_2^2}(x-\mu_2)^t(x-\mu_2))$$

Con $\boldsymbol{\mu_1} = [1\ 1]^t$, $\boldsymbol{\mu_2} = [1{,}5\ 1{,}5]^t$, $\sigma_1^2 = \sigma_2^2 = 0{,}2$. Diseñe un clasificador bayesiano que:

(a) Minimice la probabilidad de error.

(b) Minimice el riesgo promedio con la matriz de penalización $l = \begin{bmatrix} 0 & 1 \\ 0{,}5 & 0 \end{bmatrix}$.

Use un generador de números aleatorios y produzca 100 vectores de cada clase, de acuerdo con las pdf's descritas.

Utilice un clasificador bayesiano de distancia mínima para clasificar los vectores generados. ¿Cuál es el porcentaje de error en cada caso?. Repita los experimentos cambiando la media de la 2a distribución a $\boldsymbol{\mu_2} = [3\ 3]^t$

4. P4

Considere una tarea de clasificación en un espacio tridimensional. Las dos clases son modeladas con distribuciones gaussianas con medias $\mu_{\mathbf{1}} = [0\ 0\ 0]^t$ y $\mu_{\mathbf{2}} = [0{,}5\ 0{,}5\ 0{,}5]^t$, respectivamente.

Asuma que las clases son equiprobables. La matriz de covarianza para ambas distribuciones es:

$$S = \begin{bmatrix} 0{,}8 & 0{,}01 & 0{,}01 \\ 0{,}01 & 0{,}2 & 0{,}01 \\ 0{,}01 & 0{,}01 & 0{,}2 \end{bmatrix}$$

Dado el punto $x = [0{,}1\ 0{,}5\ 0{,}1]^t$, clasifiquelo con:

(a) El clasificador de distancia euclidiana mínima.

(b) El clasificador de distancia de mahalanobis mínima.

Describa los resultados y explique cual es el valor de clasificación correcto y porqué.