PRAC1: MAP Y DATOS GAUSSIANOS
CRITERIOS DE CLASIFICACIÓN BASADOS EN LA MAXIMIZACIÓN DE LA PROBABILIDAD A POSTERIORI

Asignatura: CLP Clasificación de Patrones.
Optativa de Grados
ETSETB-UPC
TSC-UPC

1 Objetivos de la práctica 1

Los objetivos de la presente práctica consisten en:

- Familiarizarse con el software de la toolbox de Matlab "Statistics and Machine Learning" en su aplicación de clasificación.
- Generar bases de datos Gaussianas y evaluar parámetros de las mismas.
- Aplicar clasificadores MAP (Lineales y Cuadráticos) sobre las clases generadas.

2 Estudio previo

1. En la página de mathworks.com: 'Discriminant Analysis': http://es.mathworks.com/help/stats/discriminant-analysis.html, se recomienda documentarse sobre el uso de la función fitcdiscr.m aplicada a clasificación.

2. Para una matriz de tipo $\mathbf{C} = \frac{1}{d}\sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$; $d = 2$ obtenga los dos autovalores teóricos en función de los parámetros ρ i σ^2 .

3 Laboratorio: creación de bases de datos Gaussianas.

Para la presente práctica no es necesario que ejecute el fichero de actualización de path.

3.1 PARTE 1: Análisis de curvas ROC

El fichero proporcionado para la realización de esta parte es *prac1_gauss3.m*

Se trabaja con vectores de dimensión d=3 (variable n_feat) y C=2 clases (variable n_classes). Las 3 coordenadas son estadísticamente independientes tal como describen las matrices de covarianza (Caso 1 explicado en clase).

La siguiente fórmula proporciona el modelo que siguen las muestras correspondientes a cada una de las dos clases.

$$\mathbf{x}|c : N(\mathbf{m}_c, \mathbf{C}); \quad c = 1..2; \quad \mathbf{C} = \tfrac{1}{d}\sigma^2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}; \quad d = 3$$

donde se define $SNR = 10\log_{10}\left(\dfrac{Energia\ promedio}{\sigma^2}\right) = 10\log_{10}\left(\dfrac{E\left[\mathbf{m}_c^T\mathbf{m}_c\right]}{\sigma^2}\right)$.

Edite el fichero *prac1_gauss3.m* y analice el código, identificando las partes principales del programa. Mediante la opción "help" de Matlab puede obtener información sobre las funciones invocadas en el programa.

Ejecute *prac1_gauss3.m* para cada uno de los cuatro casos siguientes:

SNR=3,0,-3,-10 dB.

Observe las probabilidades de error, las matrices de confusión y las curvas ROC obtenidas mediante los clasificadores:

- **Lineal** (LC): Fronteras de decisión lineales. Por defecto presupone probabilidades a priori proporcionales al número de elementos por clase e idéntica matriz de covarianza entre las clases.
- **Cuadrático (QC):** Fronteras de decisión cuadráticas. Por defecto presupone proporcionales al número de elementos por clase y estima diferentes matrices de covarianza para cada una de las clases.

<u>A incluir en el documento entregable con comentarios pertinentes:</u>
- Genere una tabla que incluya la probabilidad de error obtenida mediante clasificador lineal (LC) y probabilidad de error obtenida mediante clasificador cuadrático (QC) para cada uno de los valores de SNR.
- Incluya las curvas ROC obtenidas con SNR=-10dB y con SNR = -3 dB.
- Comente porqué la distancia cuadrática de Mahalanobis entre las dos clases da distintos resultados según se invoque la función 'mahal'.

3.2 PARTE 2: Autovalores de matriz de covarianza y forma de los clusters.

El fichero proporcionado para la realización de esta parte es <u>*prac1_QPSK.m*</u>

Se propone utilizar la modulación QPSK como base de trabajo. Por tanto los vectores de características son de dimensión d=2 y se tienen C=4 clases (o símbolos).

<u>**QPSK Caso 2**</u>

Inicialmente se utilizan matrices de covarianza iguales para todas las clases (Caso 2 explicado en clase) pero no diagonales.

$$\mathbf{x}\big|c : N(\mathbf{m}_c,\mathbf{C}); \quad c=1..4; \quad \mathbf{C}=\tfrac{1}{d}\sigma^2\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}; \quad d=2$$

<u>A realizar en el laboratorio:</u>
Edite el fichero *prac1_QPSK.m* y analice el código, identificando las partes principales del programa.

Ejecute *prac1_QPSK.m* con relación señal a ruido: SNR=10dB para cada uno de los dos casos siguientes:
- Parámetro $\text{ro} = \rho = 0$

- Parámetro ro = $\rho = +0.5$

A incluir en el documento entregable con comentarios pertinentes:

Para cada uno de los 2 casos a testear, según el valor dado a la variable ρ

- Diagrama de scatter y fronteras de decisión.
- Matrices de confusión.
- Probabilidades de error.

Observe que apenas hay diferencias entre ambos clasificadores y comente los resultados.

QPSK Caso 3 explicado en clase.

A realizar en el laboratorio:
Edite el fichero *prac1_QPSK*.m para generar la modulación QPSK en la que cada clase (o símbolo) se genera con una matriz de covarianza distinta a la del resto de clases:

SNR= + 10 dB, + 5 dB

- Símbolo 1: ro = +0.5
- Símbolo 2: ro = 0
- Símbolo 3: ro = −0.5
- Símbolo 4: ro = +0.8

A incluir en el documento entregable con comentarios pertinentes:

Para cada uno de los dos valores de SNR propuestos:

- Probabilidades de error al aplicar LC y QC.
- Autovalores teóricos para cada una de las cuatro matrices de covarianza aplicando las expresiones obtenidas en el estudio previo.
- Autovalores calculados a partir de las matrices de covarianza mediante la instrucción:

  Autoval=eig(squeeze(M_covar(:,:,c)))

  Donde *M_covar* es la variable de matrices de covarianza.
- Diagrama de scatter al aplicar LC y QC. Compare la forma de los clúster con los autovalores de las distintas matrices de covarianza.

A realizar en el laboratorio:
Para el caso de SNR=10 dB, multiplique la matriz de covarianza únicamente de la primera clase por un valor alto (por ejemplo sigma(1)=30) y compare en esta simulación el error de clasificación entre LC y QC, verificando el mejor comportamiento del discriminarte cuadrático en este caso. Observe el scatter y la forma de las fronteras que separan la clase 1 del resto de clases.

NOTA IMPORTANTE
Entregue por Atenea un único documento en pdf que incluya los resultados solicitados en las 2 partes de esta práctica.