

PRAC0: ANÁLISIS DE GAUSSIANIDAD

Asignatura: CLP: Clasificación de Patrones.

Optativa de grado

ETSETB

UPC

UPC-TSC-D5

1	Objetivos de la práctica 0	2
1.1	Análisis de gaussianidad	2
1.2	BRAIN data base	2
2	Laboratorio	3
2.1	Pasos a seguir en el laboratorio.....	3
2.2	Test de gaussianidad de distribuciones sintéticas	3
2.3	Observación de base de datos Brain	3
2.4	Estimación de intervalos de confianza.....	3
2.5	Test de gaussianidad de base de datos Brain	4

1 Objetivos de la práctica 0

- Familiarizarse con las técnicas básicas de test de gaussianidad utilizadas para medir una distribución formada a partir de un conjunto de medidas.
- Observación de la base de datos BRAIN

1.1 Análisis de gaussianidad

Se propone que consulte el documento de presentación de esta práctica para entender los siguientes test a realizar sobre una distribución de muestras:

- Histogramas
- Cumulative Density Function (CDF)
- Scatter Plot (Test de separabilidad)
- High order moments: Skewness and Kurtosis
- NormPlot
- Confidence interval for the mean value.
- Testing the good of fitness of a distribution: the Chi-squared test.

1.2 BRAIN data base

The database BRAIN is formed by 5 images. All of them correspond to a single human brain cut.

Images 1,2,3,4: Obtained by Magnetic Resonance (MR)
Image 5: Obtained by Positron Emission Tomography (PET)

Each image has 256×256 pixels = 65536 = N . Each pixel corresponds to one of four classes.

- White (White matter) Class 2
- Grey (Gray matter) Class 1
- LCR (Cephaloraquideal liquid) Class 3
- Background there are a lot of pixels outside the brain contour Class 4 (not used in this practice).

In this database each feature vector is formed taking $d = 5$ pixels. Each feature j in a vector is obtained from the same horizontal-vertical pixel at the j image.

The whole database has been recorded in the file

“PRBB_Brain2”:

Data base size: 65536 vectors \times 5 features.

A reduced Data base containing fewer pixels has been recorded in the file

“Brain_Prac0”:

The purpose of this sub-set is to be used in training in future sessions.

Data base size: 10682 vectors \times 5 features.

2 Laboratorio

2.1 Pasos a seguir en el laboratorio.

Nombre de usuario: cla11*

Password: A elegir (el actual es cla)

RECOMENDACIONES:

- Cree directamente un directorio de trabajo: por ejemplo, DIR_CLA en modo local en el escritorio o DIR_CLA en el servidor H:/soft.
- Copiar y/o ejecutar el archivo *act_path_al.m*
- Descargar y descomprimir el directorio *prac0_Test* (Atenea) en el directorio DIR_CLA
- En Matlab vaya a DIR_CLA/prac0 y realice la práctica, según las siguientes instrucciones.

2.2 Test de gaussianidad de distribuciones sintéticas

Ejecute *prac0_test* y verifique las opciones de medida de gaussianidad para las distribuciones normal, Rayleigh, laplaciana y uniforme

2.3 Observación de base de datos Brain

Ejecute “*prac0_brain*” (este programa utiliza la base de datos de training *Brain_Prac0*)

El programa realiza las siguientes operaciones:

- 1) Evalúa los parámetros kurtosis y skewness para cada característica de cada clase.
 - 2) Presenta el “scatter plot” de la base de datos.
 - 3) Evalúa las funciones Histograma, CDF y Normalplot para cada característica únicamente de una de las clases, seleccionada previamente.
- Observe y comente todas las gráficas obtenidas.
 - A partir de la observación del “scatter plot”, proponga el subconjunto de 2 características o coordenadas que resulte más discriminativo para este caso.
 - Cambie la clase analizada en la última parte del programa y copie los resultados obtenidos en el documento de resultados.

2.4 Estimación de intervalos de confianza

Edite el programa “*prac0_brain*” y sávelo con un nuevo nombre para no confundirlo con el original.

Para cada una de las 3 clases de la base de datos, proporcione los intervalos de confianza al estimar la media de la característica 1 con niveles de confianza de 95%, 99% y 99,9%. Disponga todos los valores solicitados en una tabla.

Sugerencia: Utilice las funciones de matlab “*tin*” y “*var*” para calcular la tabla anterior.

2.5 Test de gaussianidad de base de datos Brain

En el mismo fichero de trabajo del apartado anterior edite nuevo código Matlab para aceptar o rechazar las siguientes hipótesis de decisión utilizando el método Chi-squared.

- Hipótesis nula H_0 : La característica 1 de la clase 1 es gaussiana con un nivel de significancia (α) de 0.001
- Hipótesis nula H_0 : La característica 1 de la clase 2 es gaussiana con un nivel de significancia (α) de 0.001
- Hipótesis nula H_0 : La característica 1 de la clase 3 es gaussiana con un nivel de significancia (α) de 0.001

En cada una de las decisiones facilite los valores de P-value y los grados de libertad de la distribución resultante para el test. De nuevo puede organizar en una tabla todos los valores solicitados en este apartado.

Sugerencia: Utilice la función de Matlab: “chi2gof”

NOTA IMPORTANTE

Entregue en Atenea los resultados de los apartados 2.3, 2.4 y 2.5 en un único documento en formato pdf. Incluya en el documento el nombre de los dos integrantes del grupo, así como las líneas de código generadas en los apartados 2.4 y 2.5.

Deposite el documento en Atenea según la tarea y plazo correspondiente.