

Examen Mat II -24 de Abril, 2020.

Fecha límite de entrega: Lunes 27 de Abril a las 9:00AM

Descripción de base de datos. La base de datos se compone de datos recolectados de mediciones a tres tipos de flores (categorías) ω_i =(setosa, versicolor, virginica), las mediciones que se realizaron a cada categoría (clase) son el x_1 =ancho y x_2 =largo de la hoja, esto es $x = [x_1, x_2]$.

Problema. Desarrollar un clasificador Bayesiano, tomando en cuenta lo siguiente.

1. Describir las distribuciones asumiendo comportamiento Gaussiano.
2. Determinar y graficar las distribuciones *a posteriori* $P(\omega_i|x)$
3. Asumiendo lo anterior desarrollar un clasificador asumiendo a) $\Sigma_i = \sigma$; b) $\Sigma_i = \Sigma$; c) $\Sigma_i = \text{arbitraria}$.
4. Determinar el error de clasificación para el caso a y b usando la aproximación de error promedio $P(\text{error})$ dada por a) Chernoff and b) Bhattacharyya *bounds*.
5. Probar el clasificador usando el 50% de los datos tomados de forma aleatoria. Registrar y comparar el error de clasificación con el teórico.

Requerimientos: Seguir los siguientes puntos en la elaboración de tu reporte.

1. Describir adecuadamente, ecuaciones, formulas o supuestos usados.
2. Presentar gráficos para: a) distribución de datos; b) aproximación usando distribuciones normales, c) grafico de distribución *a posteriori*, d) graficar el clasificador.
3. Programar en Matlab o Python.
4. El reporte puede enviarse en *pdf* o Word.
5. Enviar archivo de programa por separado.