

Práctica 5. Clasificación Bayesiana

Objetivo

El objetivo es resolver mediante clasificación Bayesiana un problema real de clasificación multi-clase: el reconocimiento de dígitos manuscritos. Igual que en la práctica anterior, utilizaremos una versión reducida del conjunto de datos MNIST, y usaremos para la clasificación los niveles de intensidad de los 400 píxeles de cada dígito. Utilizaremos modelos Gaussianos para los atributos de cada clase.

Estudio previo (es necesario prepararlo **por escrito, antes** de acudir a la práctica)

1. Repasa las transparencias de clase y estudia las funciones auxiliares proporcionadas para esta práctica. Escribe el algoritmo de entrenamiento y clasificación multi-clase utilizando clasificación Bayesiana con atributos Gaussianos.

Desarrollo de la práctica

Copia a tu directorio de trabajo los ficheros proporcionados, y comprueba que funcionan correctamente en Matlab. A continuación escribe los programas necesarios para resolver la clasificación, siguiendo los siguientes pasos:

2. **Entrenamiento y clasificación con modelos Gaussianos regularizados.** Programa la función `entrenarGaussianas` para que aprenda el modelo Gaussiano de cada clase, tanto en el caso general como en el caso de Bayes ingenuo. Programa la función `clasificacionBayesiana`, para que haga la clasificación de un conjunto de muestras, utilizando los modelos Gaussianos entrenados.
3. **Bayes ingenuo.** Basándote en el código de la práctica anterior, programa el entrenamiento y clasificación multi-clase, usando Bayes ingenuo. Para ello separa un 20% de los datos para validación, y encuentre el mejor valor para el parámetro de regularización. Re-entrena con todos los datos para el mejor valor de λ , y utiliza los datos de test para calcular la matriz de confusión y los valores de precisión y recall para cada dígito, y visualiza las confusiones con la función `verConfusiones`. ¿Qué dígitos son los más problemáticos?
4. **Covarianzas completas.** Repite el apartado anterior en el caso general con matrices de covarianzas completas, y compara con los resultados obtenidos con Bayes ingenuo. ¿Cual de los dos modelos funciona mejor? Compara también con los resultados de la práctica anterior.

A entregar (en Moodle, dentro de un fichero .zip)

- Programa `P5.m`, junto con las funciones auxiliares que hayas programado, que vaya mostrando por pantalla los resultados de todos los apartados.
- Si no presentas la práctica durante la sesión, además deberás entregar la memoria de la práctica en un fichero `P5.pdf` ó `P5.doc` con los resultados de todos los apartados, su interpretación y las conclusiones que hayas obtenido.