

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)

- A. 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. Atención, la instrucción `load('MNISTdata2.mat')` carga tanto las variables `X` e `y` que se usan para el entrenamiento, como `Xtest` e `ytest` que sólo se usarán para la evaluación final del modelo. A continuación, escribe los programas necesarios para resolver la clasificación, siguiendo los siguientes pasos:

1. **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.
2. **Bayes ingenuo.** Basándote en el código de la práctica anterior, programa el entrenamiento y clasificación multi-clase, usando Bayes ingenuo, buscando el mejor valor para el parámetro de regularización. Re-entrena el mejor modelo obtenido con todos los datos de entrenamiento, y evalúalo. Analiza qué dígitos son más problemáticos, y qué confusiones son más habituales.
3. **Covarianzas completas.** Repite el apartado anterior en el caso general con matrices de covarianzas completas.
4. **Comparación de modelos.** Compara los resultados obtenidos con regresión logística (Práctica 4), Bayes ingenuo, y Bayes completo. Analiza y discute a qué crees que se deben las diferencias.

A entregar en Moodle

- Memoria de la práctica en un fichero `P5.pdf` con el escaneado del estudio previo, los resultados de todos los apartados, su interpretación y las conclusiones que hayas obtenido.
- Programa `P5.m` que vaya ejecutando cada apartado y mostrando por pantalla los resultados obtenidos, junto con las funciones auxiliares que hayas programado, en un fichero `.zip`

Recuerda que:

- Tienes 5 días desde tu sesión de prácticas para depositarla en Moodle.
- Deberás defender la práctica en tu próxima sesión de prácticas.