

Práctica 3. Regresión Logística

Objetivo

El objetivo es aplicar regresión logística para clasificación binaria, utilizando técnicas de regularización y validación cruzada. Se trata de clasificar un tipo de vino, a partir de un conjunto de atributos medidos del mismo.

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. Comprueba la función `minFunc` y sus diferentes opciones. Escribe el algoritmo de k-fold cross-validation para elegir el valor del parámetro de regularización.

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 regresión logística, siguiendo los siguientes pasos:

1. **Regresión logística básica.** Queremos detectar los vinos de la clase 1, a partir de los atributos 6 y 10. El programa `P3demo.m` carga y muestra los datos correspondientes. Como verás, los datos ya están normalizados, y permutados aleatoriamente.
 - a. Entrena la regresión logística básica y dibuja la superficie de separación obtenida.
 - b. Discute qué métricas son adecuadas para este problema, y calcúlalas con los datos de entrenamiento y los datos de test. Analiza los resultados.
 - c. Para un vino que tiene atributo 6 igual a 0.6, dibuja una gráfica con la probabilidad de que sea de clase 1 en función del valor del atributo 10.
2. **Regularización.** Queremos detectar los vinos de la clase 1, a partir de los atributos 6 y 10, haciendo expansión de atributos de grado 6, utilizando la función `expandir2.m` proporcionada.
 - a. Elige el parámetro de regularización `landa` mediante k-fold. Dibuja las curvas de evolución de la métrica elegida, con los datos de entrenamiento y de validación y analiza los resultados.
 - b. Re-entrena el modelo con mejor `landa` y el modelo con `landa=0` con todos los datos de entrenamiento. Dibuja las superficies de separación y evalúa ambos modelos.
 - c. Repite al apartado 1.c para el mejor modelo.
3. **Curvas Precisión/Recall.** Utilizando los datos de test, dibuja en una misma figura las curvas de precisión/recall de los diferentes modelos entrenados, y analízalas. Si queremos garantizar que el 90% de los vinos clasificados como clase 1 realmente lo sean, ¿qué habría que hacer? ¿Cuál de los distintos modelos es mejor para conseguirlo?

A entregar en Moodle

- Memoria de la práctica en un fichero `P3.pdf` con el escaneado del estudio previo, los resultados de todos los apartados, su interpretación y las conclusiones que hayas obtenido.
- Programa `P3.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.