

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

- A. Repasa las transparencias de teoría y estudia a fondo las funciones de Scikit-Learn que se indican en la última transparencia.

Desarrollo de la práctica

Queremos detectar los vinos de la clase 1, a partir de los atributos 6 y 10. El programa `P3demo.py` carga y muestra los datos correspondientes. Como verás, los datos ya están normalizados, y permutados aleatoriamente. A continuación, escribe los programas necesarios para **entrenar, evaluar y comparar entre sí distintos modelos** de regresión logística binaria, siguiendo los siguientes pasos:

1. **Métricas.** Discute qué métricas son adecuadas para este problema, y programa una función que las calcule con los datos de entrenamiento y los datos de test.
2. **Regresión logística básica.** Entrena la regresión logística básica. Dibuja la superficie de separación y evalúa el modelo. Para distintos vinos que tuvieran el atributo 6 igual a 0.6 y distintos valores del atributo 10 entre 0 y 1, ¿qué probabilidad diría tu clasificador de que pertenecieran a la clase 1? Dibuja la gráfica de probabilidad en función del valor del atributo 10 e interprétala.
3. **Expansión de atributos.** Haz la expansión polinómica de atributos de **grado 6** y entrena un modelo **sin regularizar**. Dibuja la superficie de separación y evalúa el modelo.
4. **Regularización.** Programa la regresión logística regularizada, con expansión polinómica de **grado 6**, eligiendo el parámetro de regularización mediante **k-fold**. Dibuja las curvas de evolución de la métrica elegida cuando varía la regularización con los datos de entrenamiento y de validación y analiza los resultados. Dibuja la superficie de separación y evalúa el modelo. Para un vino que tiene atributo 6 igual a 0.6, dibuja la gráfica de la probabilidad de que sea de clase 1 en función del valor del atributo 10, e interprétala.
5. **Curvas Precisión/Recall.**
 - a. Utilizando los datos de test, dibuja en una misma figura las curvas precisión/recall de los diferentes modelos entrenados, y analízalas.
 - b. Queremos garantizar que el 90% de los vinos clasificados como clase 1 realmente lo sean. ¿Podemos conseguirlo con cada uno de los modelos entrenados? Toma todos los modelos con los que sea factible, ajústalos para conseguirlo, y compara las soluciones.
6. Escribe en la memoria un **resumen de las conclusiones** principales obtenidas.

A entregar en Moodle

- Un notebook `P3.ipynb` con los programas de cada apartado, los resultados, su interpretación y las conclusiones que hayas obtenido.
- O, si lo prefieres, un fichero `P3.py` con el código y otro `P3.pdf` con los resultados de todos los apartados, su interpretación y las conclusiones que hayas obtenido.

Recuerda:

- Trae la práctica preparada para aprovechar la sesión de prácticas al máximo.
- Si te atascas, pregunta en la sesión o en tutorías.
- Debes citar correctamente todas las fuentes utilizadas.
- Tienes 5 días desde tu sesión para depositar la práctica en Moodle.
- Deberás defenderla en tu próxima sesión de prácticas.