
Práctica 4: Regresión logística regularizada

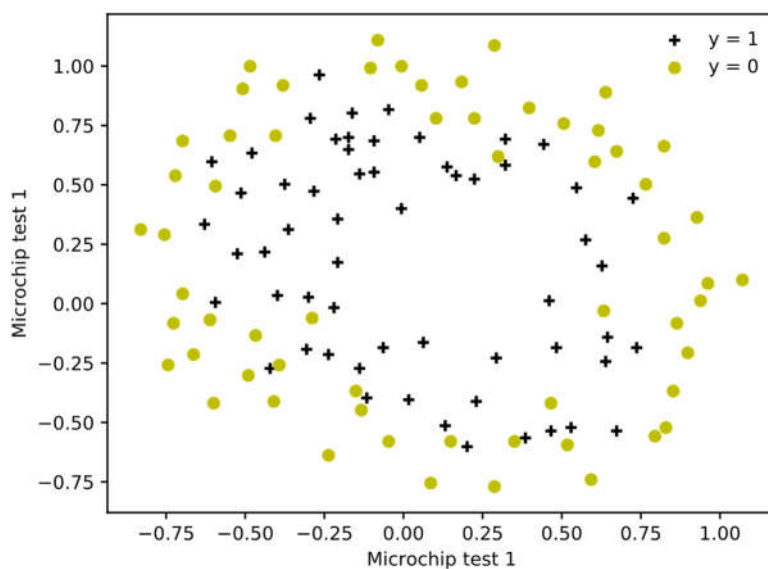
Material proporcionado:

Fichero	Explicación
data4.csv	Datos para la práctica.

1. Regresión logística regularizada

En esta práctica utilizarás la regresión logística regularizada para encontrar una función que pueda predecir si un microchip pasará o no el control de calidad, a partir del resultado de dos tests a los que se somete a los microchips (data4.csv).

Empieza visualizando los datos, para observar que no son linealmente separables:



1.1. Mapeo de los atributos

Una forma de obtener un mejor ajuste a los ejemplos de entrenamiento usando el método de regresión logística es añadir nuevos atributos a la descripción de los ejemplos, combinando los atributos originales. Utiliza la clase `sklearn.preprocessing.PolynomialFeatures` para extender cada ejemplo de entrenamiento con los términos polinómicos de x_1 y x_2 hasta la sexta potencia, completando así un total de 28 atributos para cada ejemplo:

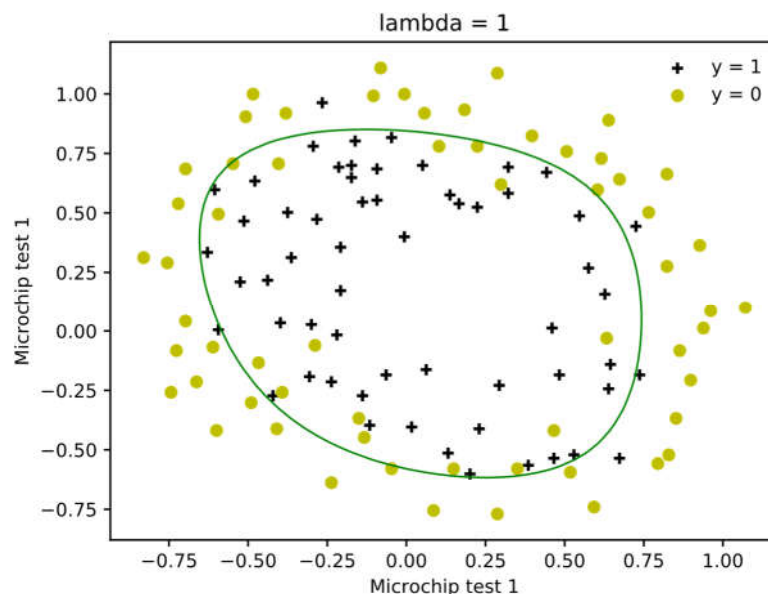
1.2. Cálculo de la función de coste y su gradiente

Implementa una función que devuelva el valor de la función de coste y otra que devuelva un vector con los valores del gradiente de la misma función para la versión regularizada de la regresión logística.

1.3. Cálculo del valor óptimo de los parámetros

Utiliza el descenso de gradiente para obtener el valor óptimo de los parámetros para la versión regularizada de la función de coste.

El resultado debería ser similar al que se muestra en el figura:



1.4. Efectos de la regularización

Experimenta con distintos valores del parámetro λ para ver cómo afecta el término de regularización al aprendizaje logístico, comparando las gráficas resultantes y evaluando el resultado del aprendizaje sobre los ejemplos de entrenamiento.

2. Entrega de la práctica

La práctica debe entregarse utilizando el mecanismo de entregas del campus virtual, no más tarde del 8 de marzo.

Se entregará un único fichero en formato notebook que contenga el código desarrollado y los comentarios y gráficas que se estimen más adecuados para explicar los resultados obtenidos.