

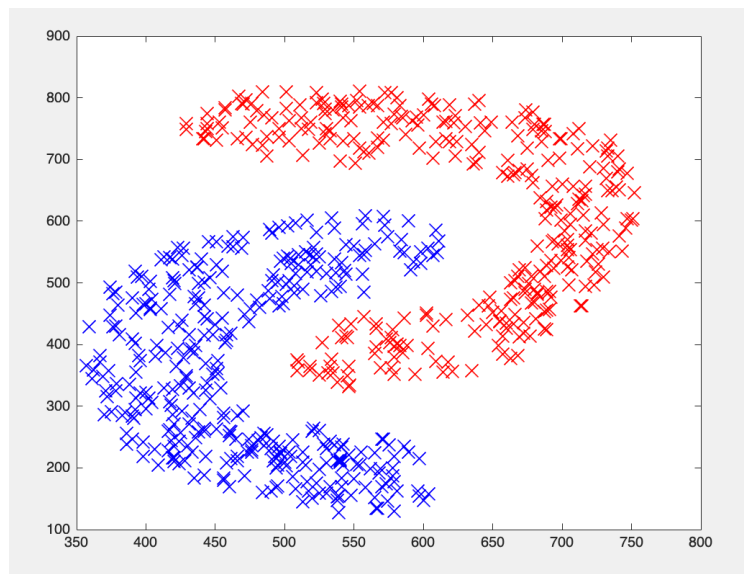


Inteligencia Artificial
Grado en Ingeniería Informática en Sistemas de Información - Curso 2020/21
HOMEWORK #2: Regresión Logística

Bienvenidos a la segunda tarea de la asignatura Inteligencia Artificial. Los modelos de aprendizaje de regresión logística presentan una gran flexibilidad para separar conjuntos de datos mediante fronteras no lineales. En esta tarea aprenderás a usar la regresión logística para separar dos clases que no se pueden separar bien a través de una línea recta.

Problema 1

En este problema implementaremos una regresión logística para el conjunto de datos que se muestra en la figura.



1. Cargue y visualice el conjunto de datos tal como se muestra en la figura.
2. Implementar una regresión logística. En este caso no se hará una evaluación del modelo mediante un conjunto de test sino que simplemente queremos analizar como la frontera de decisión aprendida a partir de todo el conjunto de datos separa las dos clases representadas por color azul y rojo. Para ello, obtenga un modelo usando todo el conjunto de datos y haga una predicción de todo el conjunto de datos, imprimiendo por pantalla la tasa de acierto y la gráfica con la frontera de decisión usando para ello la función `plotDecisionBoundary`.

Nota: No se debe usar ninguna función de optimización avanzada sino que debe implementar el método de descenso del gradiente para regresión logística.

Con el objeto de verificar el código tenga en cuenta las siguientes soluciones parciales:

Función coste:

```
Cost at initial theta (zeros): 0.693147
Gradient at initial theta (zeros):
-0.025148
22.073225
50.710059
```

Función descenso del gradiente para `num_iter=10` y `alfa=1`

```
Theta at iteration 1: 0.025148 -22.073225 -50.710059
Theta at iteration 2: 0.550296 228.207840 138.297337
Theta at iteration 3: 0.075444 -66.219675 -152.130178
Theta at iteration 4: 0.600592 184.061391 36.877219
Theta at iteration 5: 0.125740 -110.366124 -253.550296
Theta at iteration 6: 0.650888 139.914941 -64.542899
```



```
Theta at iteration 7: 0.176036 -154.512574 -354.970414
Theta at iteration 8: 0.701183 95.768491 -165.963018
Theta at iteration 9: 1.034024 247.021450 -17.233728
Theta at iteration 10: 0.559172 -47.406065 -307.661243
```

3. Como se puede comprobar en la figura del apartado 2, los datos no son separables linealmente y se necesita una frontera de decisión mucho más compleja que una simple recta. Una solución es crear más atributos mediante la función `mapFeature.m`, que mapea los atributos en términos polinomiales de x_1 y x_2 hasta el grado 2 ($x_1, x_2, x_1^2, x_2^2, x_1 \cdot x_2$) y devuelve una matriz con 6 atributos (incluyendo el atributo $x_0=1$), permitiendo de esta manera obtener una frontera de decisión definida por una curva de grado 2 que se pueda adaptar mejor a los puntos. Modificar la regresión logística del apartado 2, para obtener un modelo a partir de los datos mapeados con 6 atributos. Una vez obtenido el modelo, haga una predicción de todo el conjunto de datos, imprimiendo por pantalla la tasa de acierto y la gráfica con la frontera de decisión. En este apartado, usar la función de optimización avanzada `fmincg.m` disponible en el material para obtener los parámetros theta óptimos. La función `fmincg` es similar a la función `fminunc` y tiene los mismos argumentos de entrada y salida.

Nota:

Debe elegir el número de iteraciones de forma que al menos se obtenga una tasa de acierto mayor o igual que 90%.