UNIVERSIDAD
DE GRANADA

Aprendizaje Automático. Complejidad de H y Modelos Lineales

Ruido y complejidad, Algoritmo de aprendizaje del Perceptrón, Regresión Logística

Ricardo Ruiz Fernández de Alba

Escuela Técnica Ingeniería Informática y Matemáticas

DECSAI

Universidad de Granada

12 de abril de 2022

Índice general

1

Sobre la complejidad de H y el ruido

En este ejercicio debemos aprender la dificultad que introduce la aparición de ruido en las etiquetas a la hora de elegir la clase de funciones más adecuada. Haremos uso de tres funciones incluidas en el fichero template trabajo2.py:

- `simula_unif(N, dim, rango)`: calcula una lista de N vectores de dimensión dim. Cada vector contiene dim números aleatorios uniformes en el intervalo rango.
- `simula_gauss(N, dim, sigma)`: calcula una lista de longitud N de vectores de dimensión dim, donde cada posición del vector contiene un número aleatorio extraído de una distribución Gaussiana de media 0 y varianza dada, para cada dimension, por la posición del vector sigma.
- `simula_recta(intervalo)`: simula de forma aleatoria los parámetros, $v = (a, b)$ de una recta, $y = ax + b$, que corta al cuadrado $[-50, 50] \times [-50, 50]$.

1.1 | Ejercicio 1

Dibujar gráficas con las nubes de puntos simuladas con las siguientes condiciones

1.1.1 | simula_unif

Considere $N = 50$, dim $= 2$, $rango = [-50, 50]$ con `simula_unif (N, dim, rango)`

1.1.2 | simula_gauss

Considere $N = 50$, $dim = 2$ **y** $sigma = [5, 7]$ **con** `simula_gauss(N, dim, sigma)`

1.2 | Ejercicio 2

Vamos a valorar la influencia del ruido en la selección de la complejidad de la clase de funciones. Con ayuda de la función `simula_unif(100, 2, [-50, 50])` generamos una muestra de puntos 2D a los que vamos añadir una etiqueta usando el signo de la función $f(x,y) = y - ax - b$, es decir el signo de la distancia de cada punto a la recta simulada con `simula_recta()`.

1.2.1 | Dibujo de puntos con etiqueta y recta usada

Dibujar un gráfico 2D donde los puntos muestren (use colores) el resultado de su etiqueta. Dibuje también la recta usada para etiquetar. Observe que todos los puntos están bien clasificados respecto de la recta.

1.2.2 | Añadir ruido aleatorio

Modifique de forma aleatoria un 10 % de las etiquetas positivas y otro 10 % de las negativas y guarde los puntos con sus nuevas etiquetas. Dibuje de nuevo la gráfica anterio. Ahora habrá puntos mal clasificados respecto de la recta.

1.2.3 | Otras fronteras de clasificación

Supongamos ahora que las siguientes funciones definen la frontera de clasificación de los puntos de la muestra en lugar de una recta

- $f(x,y) = (x-10)^2 + (y-20)^2 - 400$
- $f(x,y) = 0.5(x+10)^2 + (y-20)^2 - 400$
- $f(x,y) = 0.5(x-10)^2 - (y+20)^2 - 400$
- $f(x,y) = y - 20x^2 - 5x + 3$

Visualizar el etiquetado generado en el apartado 2b junto con la gráfica de cada una de las funciones. Comparar las regiones positivas y negativas de estas nuevas funciones con las obtenidas en el caso de la recta. Argumente si estas funciones más complejas son mejores clasificadores que la función lineal. Observe las gráficas y diga qué consecuencias extrae sobre la influencia de la modificación de etiquetas en el proceso de aprendizaje. Explique el razonamiento.

2

Modelos Lineales