

Machine Learning

Tarea #1

Entrega: Marzo 7 de 2018, 11:59 p.m.

1. El archivo `datap1.zip` contiene datos para un problema de regresión lineal. Se incluyen datos de entrenamiento `xTRAINcod.txt`, `yTRAINcod.txt` y datos de prueba de entrada `xTESTcod.txt` donde `cod` es el código de su grupo (ver `sicua+`). Usted debe utilizar únicamente los datos correspondientes a su grupo. Reserve 200 datos (del conjunto de entrenamiento) para evaluar todos sus modelos.
 - a) Implemente (en su lenguaje de programación favorito) el algoritmo de descenso de gradiente estocástico visto en clase.
 - b) Usando su implementación, entrene modelos diferentes usando 50,200, 500 y 1000 datos de entrenamiento. Evalúe estos modelos en los datos de validación que ha reservado. Grafique la evolución del error cuadrático a medida que su algoritmo transcurre. Analice los resultados obtenidos.
 - c) Evalúe el modelo entrenado con 1000 datos en el conjunto de prueba de entrada `xTESTcod`. Adjunte a su informe los valores de salida correspondientes en un archivo plano `xTESTcod.txt`, donde `cod` es el código de su grupo.
2. En este problema se debe resolver un problema de clasificación de música por género, usando un modelo de regresión logística. Obtenga el conjunto de datos [MSD genre dataset](#), derivado de la base de datos Million Song Dataset. Este es un problema multiclase, del cual he derivado problemas binarios aproximadamente balanceados. Cada grupo debe resolver el problema indicado en la siguiente tabla:

	Clase 1	Clase 2
1,2,3	(jazz and blues)	(soul and reggae)
4,5,6	(dance and electronica)	(jazz and blues)
7,8,9	(jazz and blues)∪(soul and reggae)∪(dance and electronica)	(folk)
10,11	(metal)∪(punk)	(dance and electronica)
12,13	(metal)∪(punk)	(jazz and blues)∪(classical)

- a) Implemente (en su lenguaje de programación favorito) el algoritmo de ascenso de gradiente estocástico para la función de error en regresión logística visto en clase.
- b) Usando su implementación, entrene el modelo de regresión logística con 80 % de los datos. Grafique la evolución del error a medida que su algoritmo transcurre. Evalúe su modelo en el restante 20 % de los datos. Analice sus resultados.