



Universidad Simón Bolívar
Departamento de Computación y T.I.
Inteligencia Artificial II
Enero-Marzo 2017

PROYECTO 1 - REGRESIÓN LINEAL MÚLTIPLE (15 PUNTOS)

Descripción

1. Implemente el algoritmo de Descenso del Gradiente para resolver una Regresión Lineal Múltiple en el lenguaje de programación de su preferencia entre C, C++, Java o Python.
2. Pruebe su implementación del algoritmo con los siguiente conjuntos de datos simples:
 - 1.1) Peso corporal en función del peso de cerebro (1 rasgo) disponible en:
<http://people.sc.fsu.edu/~jburkardt/datasets/regression/x01.txt>
 - 1.2) Tasa de Mortalidad (3 rasgos) disponible en:
<http://people.sc.fsu.edu/~jburkardt/datasets/regression/x08.txt>

Para los datos sobre peso (1.1) :

- a) Muestre la curva de convergencia ($J(\theta)$ vs. Iteraciones) para $\alpha = 0,1$.
- b) Realice un *scatteplot* de los datos junto con la curva que minimiza la función de costo.

Para los datos sobre mortalidad (1.2) :

- a) Muestre la curva de convergencia ($J(\theta)$ vs. Iteraciones) para $\alpha = 0,1$.
 - b) Muestre en conjunto las curvas de convergencia ($J(\theta)$ vs. Iteraciones) para los siguientes valores de α : $\{0,1; 0,3; 0,5; 0,7; 0,9; 1,0\}$.
 - c) Normalice los datos y muestre la curva de convergencia ($J(\theta)$ vs. Iteraciones) para $\alpha = 0,1$.
3. Construya un modelo que incluya todas los rasgos de datos del conjunto *Ames, Iowa: Alternative to the Boston Housing Data as an End of Semester Regression Project*[1] para predecir el precio de un bien raíz disponible en: <http://www.amstat.org/publications/jse/v19n3/decock/AmesHous>.

Para esto deberá:

- a) Limpiar los datos, usando las instrucciones de artículo de DeCock [1] (sección 5)
- b) Normalizar los datos resultantes de la limpieza.
- c) Escoger el 80 % de los datos para entrenamiento, y dejar un 20 % para la prueba
- d) Evale los resultados de su modelo en sobre los datos de entrenamiento y sobre lo de prueba usando las cuatro métricas propuestas en la sección 4 de [1]

Entrega

La entrega será hasta el día lunes 20 de Febrero, a la hora de clases (1:30 am). Deberán subir al Edmodo archivo comprimido con su código y el informe (.zip o .gz) Cada grupo deberá entregar una copia impresa de su informe. El informe debe ser breve y conciso, debe incluir:

- Resumen.
- Detalles de implementación/experimentación.
- Presentación y discusión de los resultados.
- Conclusiones.

Referencias

- [1] De Cock, Dean *Journal of Statistics Education*, Volume 19, Number 3, (2011).