# UT3 TA2

## Ej1:

Cuales atributos son necesarios:  
1 Crimen y tiene outlier

2 Zona

5 NOX concentración de oxido nitrico

7 AGE

8 DIS

10 TAX

11 PTRATIO cantidad de maestros en proximidad a la ciudad

12 B – es un atributo racista que mide la cantidad de población de color en la ciudad

13 Lstat Porcentaje de pobreza en la población.

14 MEDV medianas de valores de ocupación propietaria para casas en $1000.

Tenemos data que es de tipo, numérico y binomial.

Encontramos outliers en todos los atributos

El objetivo del dataset es hallar los valores de las medias del precio de viviendas en la zona suburbana de Boston.

## Ej2:

### Feature selection:

Permite elegir el método de elegir las features a utilizarse, implementa unaserie de métodos para decidirlo, como greedy, M5 prime, T-Test, iterative T-test. Estos utilizan el peso de los atributos para decidir.

### Eliminate colinear features:

Estos atributos son aquellos que dependen de otros atributos independientes, y terminan afectando el rendimiento gral del modelo. Nos facilita el trabajo de selección y detección

### Use bias:

Permite definir si utilizar el coeficiente independiente o no en la función final.

## Resultados:

### Sin feature selection:

 squared\_error: 28.014 +/- 76.102

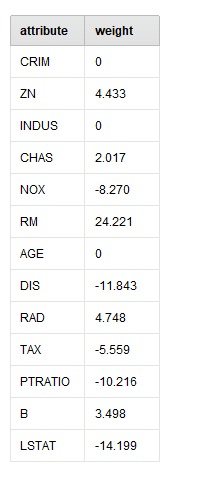
### Sin bias



### Con datos normalizados

 squared\_error: 28.389 +/- 77.467

Vector de pesos de atributos:



### Como mejorarlo

Remover variables sesgadas, filtrar los outliers (que hay muchos), transformar atributos con distribuciones desiguales.