2. Problem Statement

Build the linear regression model using scikit learn in boston data to predict 'Price'

based on other dependent variable.

Here is the code to load the data

import numpy as np

import pandas as pd

import scipy.stats as stats

import matplotlib.pyplot as plt

import sklearn

from sklearn.datasets import load\_boston

boston = load\_boston()

bos = pd.DataFrame(boston.data)

bos.head()

bos.info()

print(boston.data.shape)

print(boston.feature\_names)

bos.columns = boston.feature\_names

print(bos.head())

print(boston.target.shape)

bos['PRICE'] = boston.target

print(bos.head())

print(bos.describe())

X = bos.drop('PRICE', axis = 1)

Y = bos['PRICE']

from sklearn.model\_selection import train\_test\_split

X\_train, X\_test, Y\_train, Y\_test = sklearn.model\_selection.train\_test\_split(X, Y, test\_size = 0.33, random\_state = 5)

print(X\_train.shape)

print(X\_test.shape)

print(Y\_train.shape)

print(Y\_test.shape)

from sklearn.linear\_model import LinearRegression

lm = LinearRegression()

lm.fit(X\_train, Y\_train)

Y\_pred = lm.predict(X\_test)

plt.scatter(Y\_test, Y\_pred)

plt.xlabel("Prices: $Y\_i$")

plt.ylabel("Predicted prices: $\\hat{Y}\_i$")

plt.title("Prices vs Predicted prices: $Y\_i$ vs ${Y}\_i$")

mse = sklearn.metrics.mean\_squared\_error(Y\_test, Y\_pred)

print(mse)











