Homework 4

a. Compute the linear regression solution (i.e., best fit line) for this dataset. Use the entire dataset to train and find the best fit line. Give the expression for the best fit line and compute the error performance on the training dataset. Recall that the error performance is measured by the sum of squared errors.

In [1]:
```
import numpy as np
import scipy.stats as st
import sklearn.linear_model as lm
import matplotlib.pyplot as plt
%matplotlib inline
```

In [2]:
```
#data set
d=np.array([[0,394.33],[4,329.5],[8,291],[12,255.17],[16,229.33],[20,204.83],[24,179],[28,163.83],[32,150.33]])
n=len(d)

#simple linear regression model
Sxy=np.zeros(n)
Sxx=np.zeros(n)
x_mean=np.mean(d[:,0])
y_mean=np.mean(d[:,1])
for i in range(n):
    Sxy[i]=(d[i,0]-x_mean)*(d[i,1]-y_mean)
    Sxx[i]=(d[i,0]-x_mean)**2
S_xy=np.sum(Sxy)
S_xx=np.sum(Sxx)
k=S_xy/S_xx
b=y_mean-k*x_mean

x_tr = np.linspace(0, 35, 200)
y_tr = k*x_tr+b
```