EE2211 Introduction to Machine Learning

T14 & T18, Chua Dingjuan elechuad@nus.edu.sg

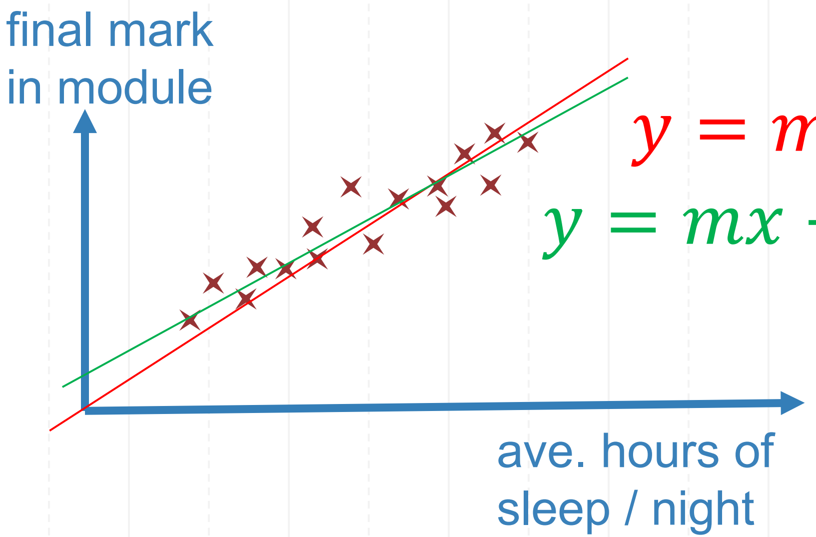
Materials @ tiny.cc/ee2211tut

Tutorial 5

Linear Regression

What is Linear Regression?

	input (s) (x)*1* … *d*	*output* *y*	
	ave. hrs of sleep / night	final mark in module	
peter x_1	5	23	y_1
jane x_2	4	45	y_2
x_3	3	89	y_3
x_4	8	46	y_4
x_5	9	90	y_5
…	8.5	80	…
x_m	…	…	y_m



$$y = mx$$

$$y = mx + c$$

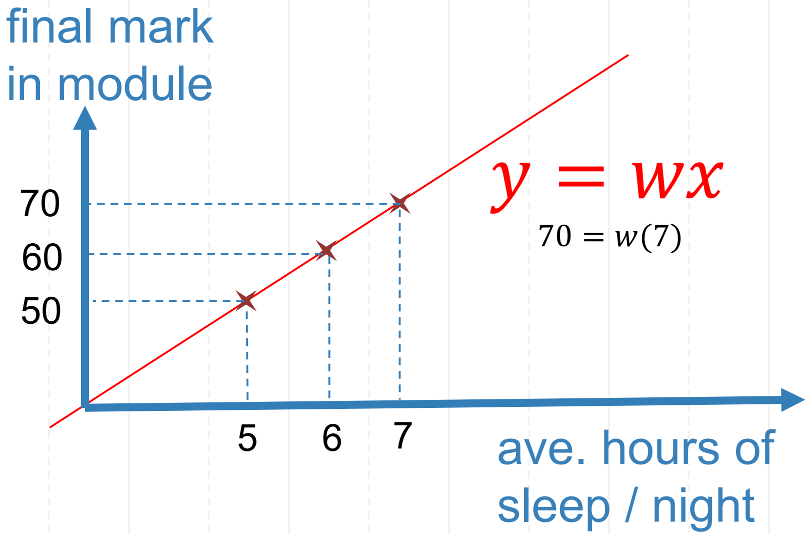
What is Linear Regression?

	1	*y*	
	ave. hrs of sleep / night	final mark in module	
x_1	5	50	y_1
x_2	6	60	y_2
x_3	7	70	y_3

$$\boldsymbol{f_w(x) = y = Xw}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} [w]$$

$$\begin{bmatrix} 50 \\ 60 \\ 70 \end{bmatrix} = \begin{bmatrix} 5 \\ 6 \\ 7 \end{bmatrix} [10]$$



What is Linear Regression?

	1 … d	y	
	ave. hrs of sleep / night	final mark in module	
x_1	5	55	y_1
x_2	6	65	y_2
x_3	7	75	y_3

$$\boldsymbol{y} = \boldsymbol{f_w}(\boldsymbol{x}) = \boldsymbol{Xw}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \end{bmatrix} \begin{bmatrix} b \\ w \end{bmatrix}$$

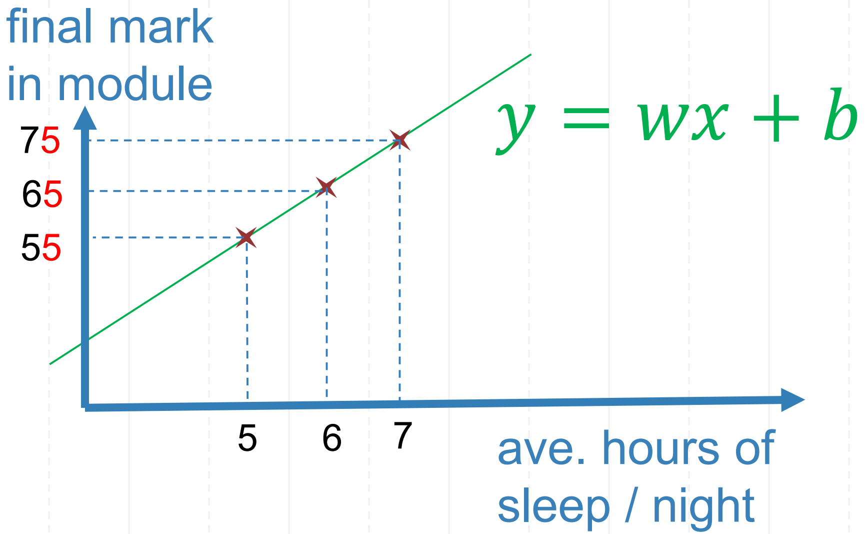
$$\begin{bmatrix} 55 \\ 65 \\ 75 \end{bmatrix} = \begin{bmatrix} 1 & 5 \\ 1 & 6 \\ 1 & 7 \end{bmatrix} \begin{bmatrix} b \\ w \end{bmatrix}$$

→ 55 = b + 5w

→ 65 = b + 6w

Solving, $\begin{bmatrix} b \\ w \end{bmatrix} = \begin{bmatrix} 5 \\ 10 \end{bmatrix}$

```
>>> y = [[55],[65],[75]]
>>> x = [[1, 5], [1, 6], [1, 7]]
>>> a(x,y)
[[ 5.]
 [10.]]
```

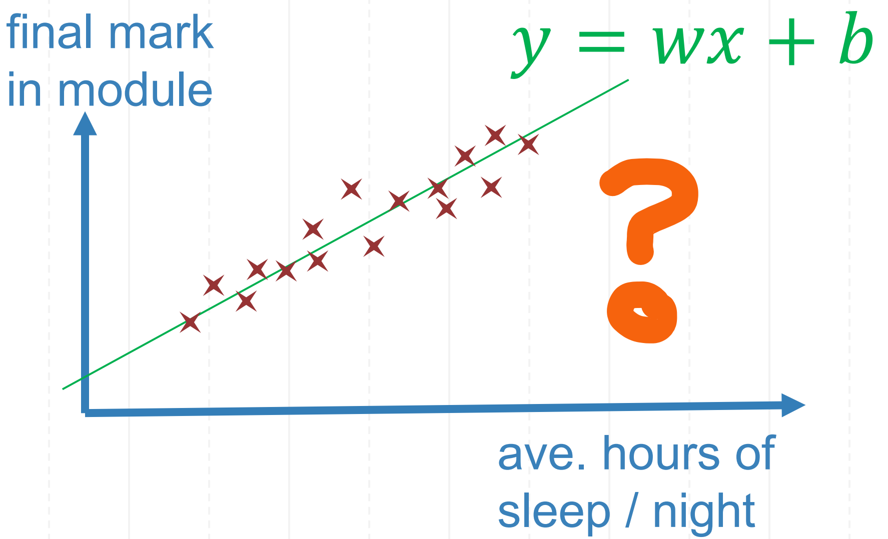


What is Linear Regression?

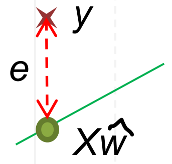
	1 … d	y	
	ave. hrs of sleep / night	final mark in module	
x_1	5	23	y_1
x_2	4	45	y_2
x_3	3	89	y_3
x_4	8	46	y_4
x_5	9	90	y_5
…	8.5	80	…
x_m	…	…	y_m

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{w}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & \dots \end{bmatrix} \begin{bmatrix} b \\ w \end{bmatrix} \Rightarrow \begin{bmatrix} 23 \\ 45 \\ \dots \end{bmatrix} = \begin{bmatrix} 1 & 5 \\ 1 & 4 \\ 1 & \dots \end{bmatrix} \begin{bmatrix} b \\ w \end{bmatrix}$$



$$\sum (\boldsymbol{e})^2 \to 0$$



Discussion of Solutions 5

Q1,2,3,4,5,6,7-9(self)

Question1

Given the following data pairs for training:

$\{x=-10\} \rightarrow \{y=5\}$

$\{x=-8\} \rightarrow \{y=5\}$

$\{x=-3\} \rightarrow \{y=4\}$

$\{x=-1\} \rightarrow \{y=3\}$

$\{x=2\} \rightarrow \{y=2\}$

$\{x=8\} \rightarrow \{y=2\}$

(a) Perform a linear regression with addition of a bias/offset term to the input feature vector and sketch the result of line fitting.

(b) Perform a linear regression without inclusion of any bias/offset term and sketch the result of line fitting.

(c) What is the effect of adding a bias/offset term to the input feature vector?

(a) SOLUTION - Overdetermined

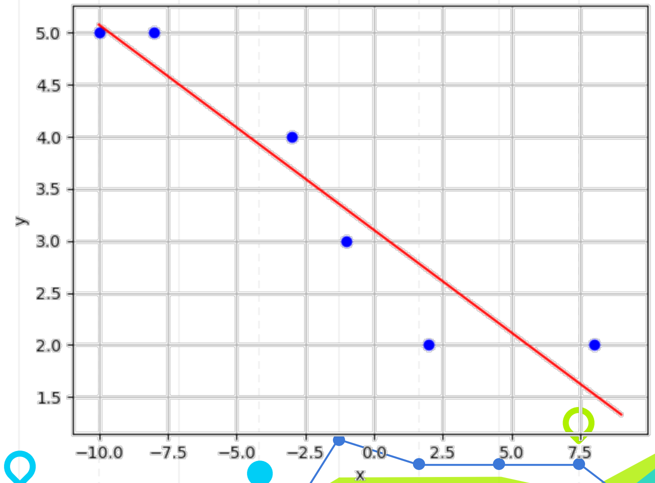
$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{w} \qquad \hat{\boldsymbol{w}} = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & \dots \end{bmatrix} \begin{bmatrix} b \\ w \end{bmatrix}$$

$$\begin{bmatrix} 5 \\ 5 \\ 4 \\ 3 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 & -10 \\ 1 & -8 \\ 1 & -3 \\ 1 & -1 \\ 1 & 2 \\ 1 & 8 \end{bmatrix} \begin{bmatrix} b \\ w \end{bmatrix}$$

$$\hat{\boldsymbol{w}} = \begin{bmatrix} 6 & -12 \\ -12 & 242 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ -10 & -8 & -3 & -1 & 2 & 8 \end{bmatrix} \begin{bmatrix} 5 \\ 5 \\ 4 \\ 3 \\ 2 \\ 2 \end{bmatrix}$$

$$\hat{\boldsymbol{w}} = \begin{bmatrix} b \\ w \end{bmatrix} \approx \begin{bmatrix} 3.1055 \\ -0.1972 \end{bmatrix}$$



Question1

(a) Perform a linear regression with addition of a bias/offset term to the input feature vector and sketch the result of line fitting.

(b) Perform a linear regression without inclusion of any bias/offset term and sketch the result of line fitting.

(c) What is the effect of adding a bias/offset term to the input feature vector?

Given the following data pairs for training:

$\{x = -10\} \rightarrow \{y = 5\}$

$\{x = -8\} \rightarrow \{y = 5\}$

$\{x = -3\} \rightarrow \{y = 4\}$

$\{x = -1\} \rightarrow \{y = 3\}$

$\{x = 2\} \rightarrow \{y = 2\}$

$\{x = 8\} \rightarrow \{y = 2\}$

(b) SOLUTION - Overdetermined

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{w}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \dots \end{bmatrix} [w]$$

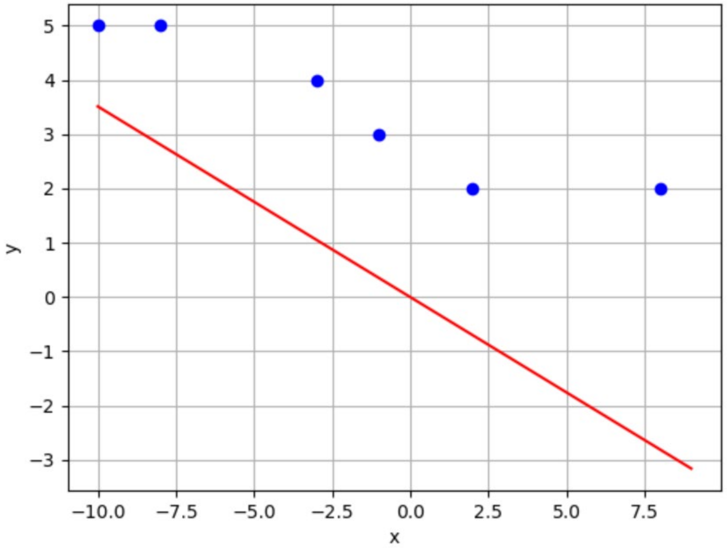
$$\begin{bmatrix} 5 \\ 5 \\ 4 \\ 3 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} -10 \\ -8 \\ -3 \\ -1 \\ 2 \\ 8 \end{bmatrix} [w]$$

$$\hat{\boldsymbol{w}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

$$\hat{\boldsymbol{w}} = [242]^{-1}[-10 \quad -8 \quad -3 \quad -1 \quad 2 \quad 8]\begin{bmatrix} 5 \\ 5 \\ 4 \\ 3 \\ 2 \\ 2 \end{bmatrix}$$

$$\hat{\boldsymbol{w}} = -0.3512$$

(c) The bias/offset term allows the line to move away from the origin (moved vertically in this case).



Question2

Given the following data pairs for training:

$\{x_1 = 1,\ x_2 = 0,\ \ x_3 = 1\} \rightarrow \{y = 1\}$

$\{x_1 = 2,\ x_2 = -1, x_3 = 1\} \rightarrow \{y = 2\}$

$\{x_1 = 1,\ x_2 = 1,\ \ x_3 = 5\} \rightarrow \{y = 3\}$

(a) Predict the following test data without inclusion of an input bias/offset term.

(b) Predict the following test data with inclusion of an input bias/offset term.

Test Data : $\{x_1 = -1,\ x_2 = 2,\ \ x_3 = 8\} \rightarrow \{y = ?\}$
$\{x_1 = 1,\ x_2 = 5,\ \ x_3 = -1\} \rightarrow \{y = ?\}$

(a) No bias term - EVEN - DETERMINED

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{w}$$

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 2 & -1 & 1 \\ 1 & 1 & 5 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix}$$

$$\widehat{\boldsymbol{w}} = \boldsymbol{X}^{-1}\boldsymbol{y}$$

$$\hat{\mathbf{w}} = \mathbf{X}^{-1}\mathbf{y} = \begin{bmatrix} 1 & 0 & 1 \\ 2 & -1 & 1 \\ 1 & 1 & 5 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 0.3333 \\ -0.6667 \\ 0.6667 \end{bmatrix}$$

$$\hat{\mathbf{y}}_t = \mathbf{X}_t\hat{\mathbf{w}} = \begin{bmatrix} -1 & 2 & 8 \\ 1 & 5 & -1 \end{bmatrix} \begin{bmatrix} 0.3333 \\ -0.6667 \\ 0.6667 \end{bmatrix} = \begin{bmatrix} 3.6667 \\ -3.6667 \end{bmatrix}$$

Question2

Given the following data pairs for training:

$\{x_1 = 1, x_2 = 0, \quad x_3 = 1\} \rightarrow \{y = 1\}$

$\{x_1 = 2, \ x_2 = -1, x_3 = 1\} \rightarrow \{y = 2\}$

$\{x_1 = 1, \ x_2 = 1, \quad x_3 = 5\} \rightarrow \{y = 3\}$

(a) Predict the following test data without inclusion of an input bias/offset term.

(b) Predict the following test data with inclusion of an input bias/offset term.

Test Data : $\{x_1 = -1, \ x_2 = 2, \quad x_3 = 8\} \rightarrow \{y = ?\}$
$\{x_1 = 1, \ x_2 = 5, \quad x_3 = -1\} \rightarrow \{y = ?\}$

(b) With bias term - UNDER - DETERMINED

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{w}$$

$$\begin{bmatrix}1\\2\\3\end{bmatrix} = \begin{bmatrix}1 & 1 & 0 & 1\\1 & 2 & -1 & 1\\1 & 1 & 1 & 5\end{bmatrix}\begin{bmatrix}b\\w_0\\w_1\\w_2\end{bmatrix}$$

$$\widehat{\boldsymbol{w}} = \boldsymbol{X}^T\left(\boldsymbol{X}\boldsymbol{X}^T\right)^{-1}\boldsymbol{y}$$

$$\hat{\mathbf{w}} = \mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{y} = \begin{bmatrix}1 & 1 & 1\\1 & 2 & 1\\0 & -1 & 1\\1 & 1 & 5\end{bmatrix}\begin{bmatrix}3 & 4 & 7\\4 & 7 & 7\\7 & 7 & 28\end{bmatrix}^{-1}\begin{bmatrix}1\\2\\3\end{bmatrix} = \begin{bmatrix}-0.1429\\0.5238\\-0.4762\\0.6190\end{bmatrix}$$

$$\hat{\mathbf{y}}_t = \mathbf{X}_t\hat{\mathbf{w}} = \begin{bmatrix}1 & -1 & 2 & 8\\1 & 1 & 5 & -1\end{bmatrix}\begin{bmatrix}-0.1429\\0.5238\\-0.4762\\0.6190\end{bmatrix} = \begin{bmatrix}3.3333\\-2.6190\end{bmatrix}$$

Question3

A college bookstore must order books two months before each semester starts. They believe that the number of books that will ultimately be sold for any particular course is related to the number of students registered for the course when the books are ordered. They would like to develop a linear regression equation to help plan how many books to order.

From past records, the bookstore obtains the number of students registered, X, and the number of books actually sold for a course, Y, for 12 different semesters. These data are shown below.

(a) Obtain a scatter plot of the number of books sold versus the number of registered students.
(b) Write down the regression equation and calculate the coefficients for this fitting.
(c) Predict the number of books that would be sold in a semester when 30 students have registered.
(d) Predict the number of books that would be sold in a semester when 5 students have registered.

Semester	Students	Books
1	36	31
2	28	29
3	35	34
4	39	35
5	30	29
6	30	30
7	31	30
8	38	38
9	36	34
10	38	33
11	29	29
12	26	26

Question3

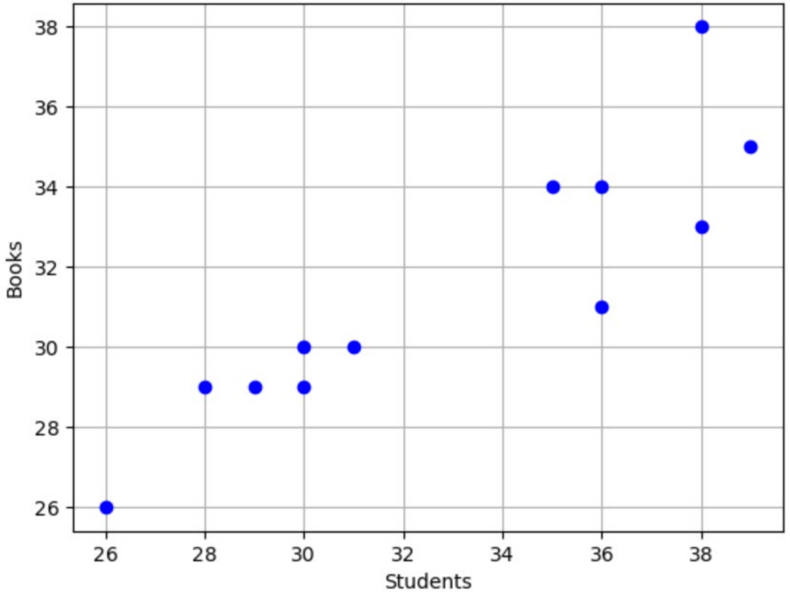
(a) Obtain a scatter plot of the number of books sold versus the number of registered students.

Semester	Students	Books
1	36	31
2	28	29
3	35	34
4	39	35
5	30	29
6	30	30
7	31	30
8	38	38
9	36	34
10	38	33
11	29	29
12	26	26

```
import matplotlib.pyplot as plt

xdata = [[36],[28],[35],[39],[30],[30],[31],[38],[36],[38],[29],[26]]

ydata = [[31],[29],[34],[35],[29],[30],[30],[38],[34],[33],[29],[26]]

#Plotting the original data points as blue points
plt.plot(xdata,ydata,'bo')
```

circles

blue



Question3

(b) Write down the regression equation and calculate the coefficients for this fitting.

Semester	Students	Books
1	36	31
2	28	29
3	35	34
4	39	35
5	30	29
6	30	30
7	31	30
8	38	38
9	36	34
10	38	33
11	29	29
12	26	26

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{w}$$

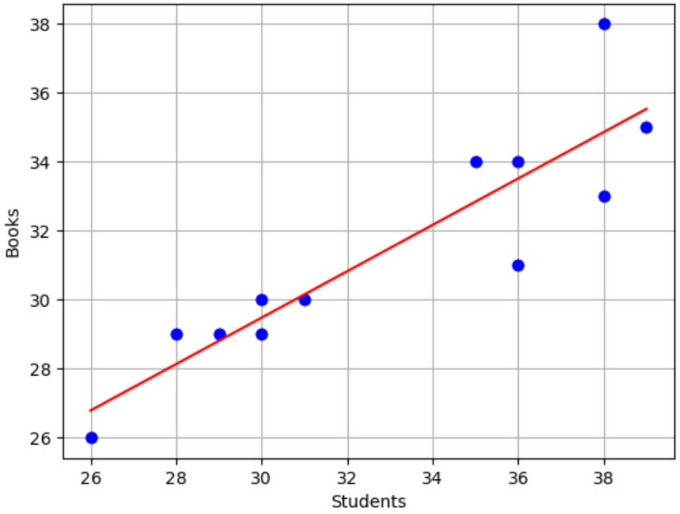
$$\begin{bmatrix} 31 \\ 29 \\ .. \\ 29 \\ 26 \end{bmatrix} = \begin{bmatrix} 1 & 36 \\ 1 & 28 \\ 1 & \dots \\ 1 & 29 \\ 1 & 26 \end{bmatrix} \begin{bmatrix} b \\ w \end{bmatrix}$$

$$\widehat{\boldsymbol{w}} = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

$$= \begin{bmatrix} 12 & 396 \\ 396 & 13288 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 36 & 28 & 35 & 39 & 30 & 30 & 31 & 38 & 36 & 38 & 29 & 26 \end{bmatrix} \begin{bmatrix} 31 \\ 29 \\ 34 \\ 35 \\ 29 \\ 30 \\ 30 \\ 38 \\ 34 \\ 33 \\ 29 \\ 26 \end{bmatrix}$$

$$= \begin{bmatrix} 9.30 \\ 0.6727 \end{bmatrix}$$

$y = 0.6727x + 9.3$



Question3

(c) Predict the number of books that would be sold in a semester when 30 students have registered.

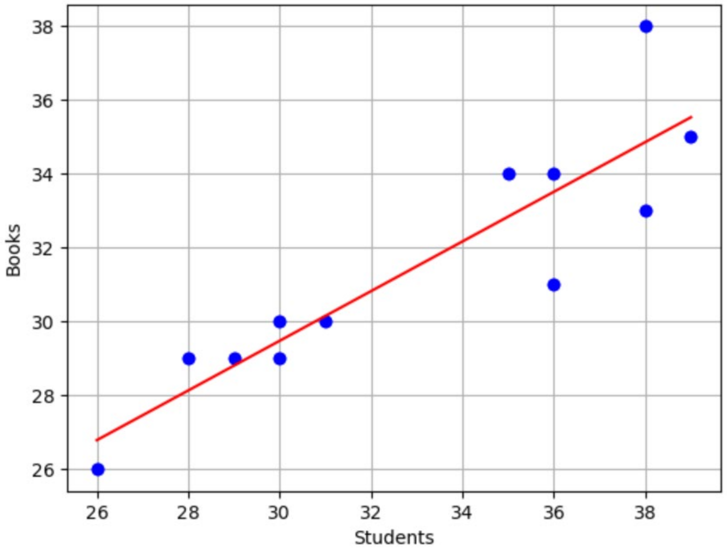
$$\hat{y}_t = \mathbf{X}_t\hat{\mathbf{w}} = [1 \ \ 30]\begin{bmatrix} 9.30 \\ 0.6727 \end{bmatrix} = 29.4818$$

(d) Predict the number of books that would be sold in a semester when 5 students have registered.

$$(\hat{y}_t = 12.6636)$$

- Prediction may be over optimistic.
  - Since 5 students is not within the range of the sampled number of students, it might not be appropriate to use the regression equation to make this prediction.
- Unknown if straight-line model would fit data at this point.
- Might not want to extrapolate far beyond the observed range.

Semester	Students	Books
1	36	31
2	28	29
3	35	34
4	39	35
5	30	29
6	30	30
7	31	30
8	38	38
9	36	34
10	38	33
11	29	29
12	26	26



Question4

Repeat Q3 using the new training data.

(a) Calculate the regression coefficients for this fitting.
(b) Predict the number of books that would be sold in a semester when 30 students have registered.
(c) Purge those duplicating data and re-fit the line and observe the impact on predicting the number of books that would be sold in a semester when 30 students have registered.
(d) Sketch and compare the two fitting lines.

Q3

Semester	Students	Books
1	36	31
2	28	29
3	35	34
4	39	35
5	30	29
6	30	30
7	31	30
8	38	38
9	36	34
10	38	33
11	29	29
12	26	26

Q4 new training data

Semester	Students	Books
1	36	31
2	26	20
3	35	34
4	39	35
5	26	20
6	30	30
7	31	30
8	38	38
9	36	34
10	38	33
11	26	20
12	26	20

Question4

(a) Calculate the regression coefficients for this fitting.

$$\hat{\mathbf{w}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = \begin{bmatrix} 12 & 387 \\ 387 & 12791 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 36 & 26 & 35 & 39 & 26 & 30 & 31 & 38 & 36 & 38 & 26 & 26 \end{bmatrix} \begin{bmatrix} 31 \\ 20 \\ 34 \\ 35 \\ 20 \\ 30 \\ 30 \\ 38 \\ 34 \\ 33 \\ 20 \\ 20 \end{bmatrix} = \begin{bmatrix} -10.4126 \\ 1.2143 \end{bmatrix}$$

(b) Predict the number of books that would be sold in a semester when 30 students have registered.

$$\hat{y}_t = \mathbf{X}_t\hat{\mathbf{w}} = [1 \;\; 30] \begin{bmatrix} -10.4126 \\ 1.2143 \end{bmatrix} = 26.0177$$

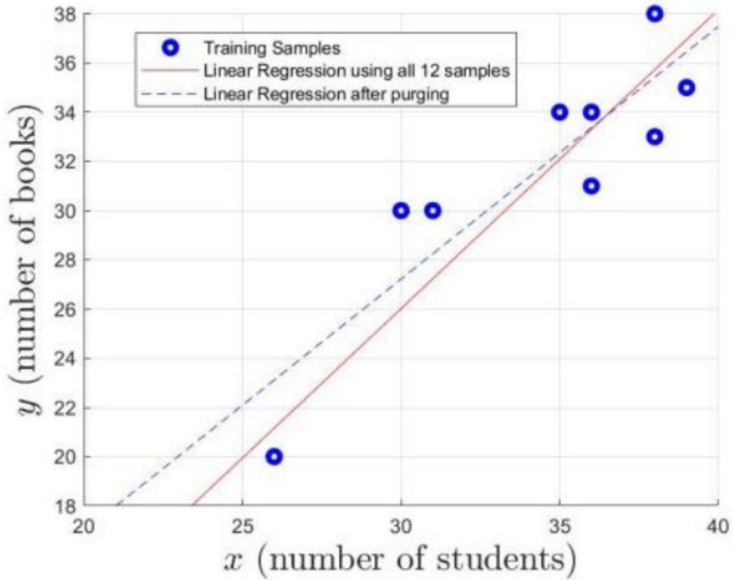
Question4

(c) Purge those duplicating data and re-fit the line and observe the impact on predicting the number of books that would be sold in a semester when 30 students have registered.

After purging duplicated data :

$$\hat{\mathbf{w}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = \begin{bmatrix} 9 & 309 \\ 309 & 10763 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 36 & 35 & 39 & 30 & 31 & 38 & 36 & 38 & 26 \end{bmatrix} \begin{bmatrix} 31 \\ 34 \\ 35 \\ 30 \\ 30 \\ 38 \\ 34 \\ 33 \\ 20 \end{bmatrix} = \begin{bmatrix} -3.5584 \\ 1.0260 \end{bmatrix}$$

$$\hat{y}_t = \mathbf{X}_t\hat{\mathbf{w}} = [1 \;\; 30] \begin{bmatrix} -3.5584 \\ 1.0260 \end{bmatrix} = 27.2208$$



- Duplicating samples can influence the learning and decision too.
- In this case, purging seems to give a more optimistic prediction for a relatively small number of students (< 37) vs more conservative prediction for a relatively large number of students (>37).

Question5

The data file “government-expenditure-on-education.csv” from https://data.gov.sg/dataset/government-expenditureon-education (Tutorial 2) depicts the government’s educational expenditure over the years.
Predict the educational expenditure of year 2021 based on linear regression. Solve the problem using Python with a plot.

	A	B
1	year	total_expenditure_on_education
2	1981	942517
3	1982	1358430
4	1983	1611647
5	1984	1769728
6	1985	1812376
7	1986	1641893
8	1987	1654115
9	1988	1604473
10	1989	1765250
11	1990	2056374
12	1991	2816371
13	1992	2597894
14	1993	2902886
15	1994	3318956
16	1995	3443857
17	1996	3771955
18	1997	4449754
19	1998	4853120
20	1999	4857488
21	2000	5867507

STEPS :

- Import the .csv file, extract year and expenditure data
- Over-determined system
    - Create the numpy arrays
    - Calculate coefficients of linear regression
- Scatter plot of expenditure vs year
- Plot linear regression line

Question5

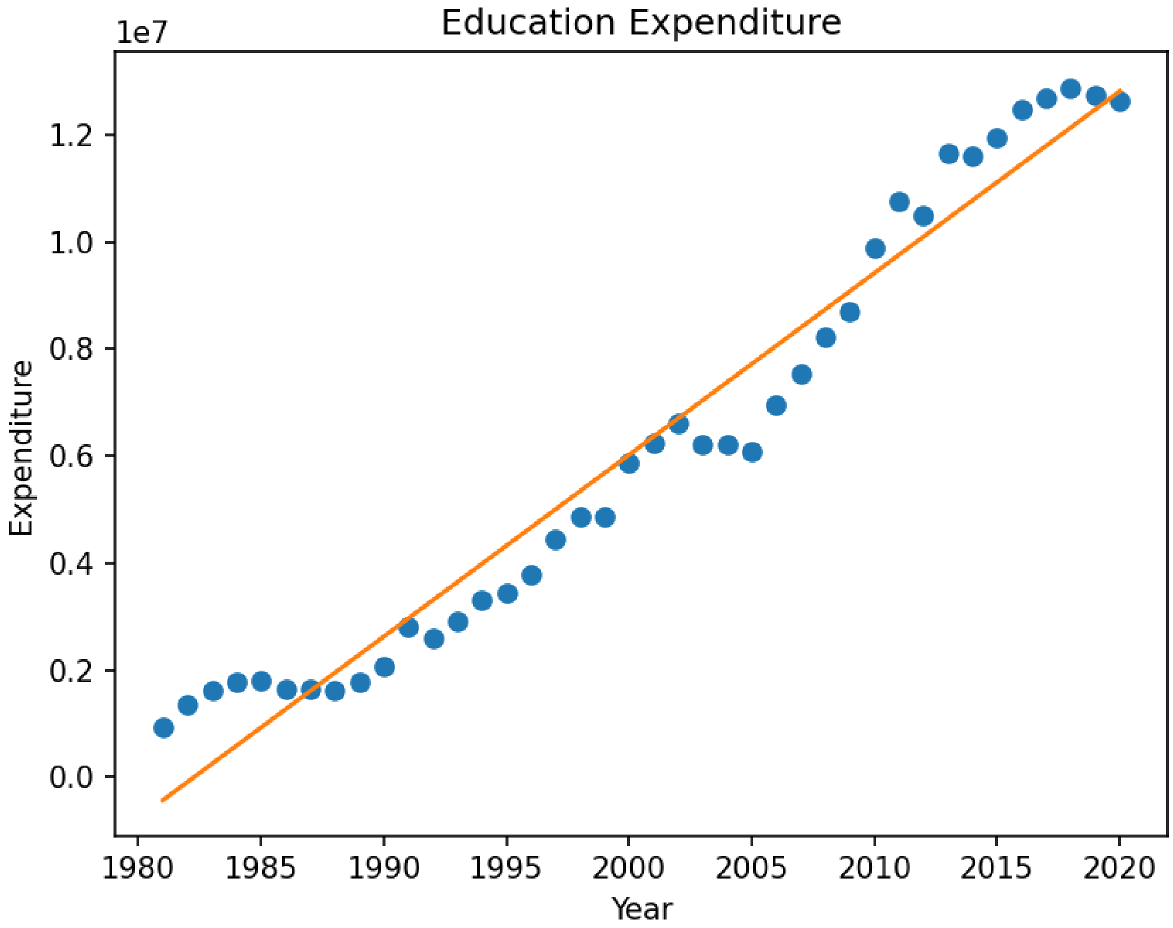
- Import the .csv file, extract year and expenditure data
- Over-determined system
  - Create the numpy arrays
  - Calculate coefficients of linear regression
- Plots

Output
[-6.4843247e+08 3.2683591e+05]
12102904.270643068

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

#Extract year and expenditure data from csv file
df = pd.read_csv("government-expenditure-on-education.csv")
expenditureList = df ['total_expenditure_on_education'].tolist()
yearList = df ['year'].tolist()

#Creating the numpy arrays

#Adding a column of ones to the year data (x data).
#You can also use np.ones() and hstack to do the following.
m_list = [[1]*len(yearList), yearList]

#Transpose required, else a row vector is created instead of a column vector
X = np.array(m_list).T

y = np.array(expenditureList)

w = np.linalg.inv(X.T @ X) @ X.T @ y
print(w)

#Creating the y-data points for the regression line
y_line = X.dot(w)

#Plotting Scatter Plot for data and Linear Regression Line
plt.plot(yearList, expenditureList, 'o', label = 'Expenditure over the years')
plt.plot(yearList, y_line)
plt.xlabel('Year')
plt.ylabel('Expenditure')
plt.title('Education Expenditure')
plt.show()

#Calculating the prediction for 2021
y_predict = np.array([1, 2021]).dot(w)
print(y_predict)
```

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \end{bmatrix}$$

Question5



Education Expenditure
1e7
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Expenditure
1980
1985
1990
1995
2000
2005
2010
2015
2020
Year

Question6

Download the CSV file for red-wine using " wine = pd.read_csv("https://archive.ics.uci.edu/ml/machine-learningdatabases/wine-quality/winequality-red.csv",sep=';') ". Use Python to perform the following tasks.

(a) Assume the given list of data is already randomly indexed (i.e., not in particular order), split the database into two sets: [0:1500] samples for regression training, and [1500:1599] samples for testing.

(b) Perform linear regression on the training set and print out the learned parameters.

(c) Perform prediction using the test set and provide the prediction accuracy in terms of the mean of squared errors (MSE).

Question6

Output

[2.22330327e+01
2.68702621e-02 -
1.12838019e+00 -
2.06141685e-01
1.22000584e-02 -
1.77718503e+00
4.29357454e-03 -
3.18953315e-03 -
1.81795124e+01 -
3.98142390e-01
8.92474793e-01
2.77147239e-01]

0.34352638122440293
0.343526381224403

```
import pandas as pd
import numpy as np
from numpy.linalg import inv
from sklearn.metrics import mean_squared_error

## get data from web
wine = pd.read_csv("winequality-red.csv",sep=';')
wine.info()
y = wine.quality
x = wine.drop('quality',axis = 1)

## Include the offset/bias term
x0 = np.ones((len(y),1))
X = np.hstack((x0,x))

## split data into training and test sets
## (Note: this exercise introduces the basic protocol of using the training-test
## partitioning of samples for evaluation assuming the list of data is already randomly indexe

## In case you really want a general random split to have a better training/test distributions:
## from sklearn.model_selection import train_test_split
## train_X,test_X,train_y,test_y = train_test_split(X,y,test_size=99/1599, random_state= 0)

train_X = X[0:1500]
train_y = y[0:1500]
test_X = X[1500:1599]
test_y = y[1500:1599]

## linear regression
w = inv(train_X.T @ train_X) @ train_X.T @ train_y
print(w)

yt_est = test_X.dot(w);

MSE = np.square(np.subtract(test_y,yt_est)).mean()
print(MSE)

MSE = mean_squared_error(test_y,yt_est)
print(MSE)
```

Question7

This question is related to understanding of modelling assumptions.
The function given by $f(\mathbf{x}) = 1 + x_1 + x_2 - x_3 - x_4$ is affine.

(a) True

(b) False

Affine function

$f(\mathbf{x}) = \boldsymbol{a}^T\mathbf{x} + b$ scalar b is called the offset (or bias)

Question8

There could be more than one answer.
Suppose $f(\mathbf{x})$ is a scalar function of d variables where $\mathbf{x}$ is a d ×1 vector. Then, without taking data points into consideration, the outcome of differentiation of $f(\mathbf{x})$ w.r.t. $\mathbf{x}$ is

a) a scalar

b) a d×1 vector

c) a d×d matrix

d) a d×d×d tensor

e) None of the above

Differentiation of a scalar function w.r.t. a vector

If $f(\mathbf{x})$ is a scalar function of d variables, $\mathbf{x}$ is a d x1 vector. Then differentiation of $f(\mathbf{x})$ w.r.t. $\mathbf{x}$ results in a d x1 vector

$$\frac{df(\mathbf{x})}{d\mathbf{x}} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_d} \end{bmatrix} \rightarrow d \times 1$$

(This video could be useful ! : https://www.youtube.com/watch?v=iWxY7VdcSH8)

Question9

The values of feature vector x and their corresponding values of target vector y are shown in the table below:

x	[3, -1, 0]	[5, 1, 2]	[9, -1, 3]	[-6, 7, 2]	[3, -2, 0]
y	[1, -1]	[-1, 0]	[1, 2]	[0, 3]	[1, -2]

Find the least square solution of w using linear regression of multiple outputs and then estimate the value of y when x = [8, 0, 2].

```
#python
import numpy as np
from numpy.linalg import inv
X = np.array([[3, 0, 0], [5, 1, 2], [9, -1, 3], [-6, 7, 2], [3, -2, 0]])
y = np.array([[1, -1], [-1, 0], [1, 2], [0, 3], [1, -2]])
w = inv(X.T @ X) @ X.T @ y
print(w)
newX=np.array([8, 0, 2])
newY=newX@w
print(newY)
```

Outputs
W =[[-0.1884058 -0.70434783]
[-0.42834138 -0.8115942]
[0.66344605 2.36231884]]
Y =[-0.18035427 -0.91014493]