

In this Lab, we will practice in Numpy, Pandas and simple linear regression. We will use the dataset mtcars.

```
In [19]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

Use the dataset cars.csv

```
In [12]: cars_df = pd.read_csv(filepath_or_buffer='cars.csv')
cars_df.head()
```

```
Out[12]:
```

	make	model	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
0	Mazda	Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
1	Mazda	Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
2	Datsun	Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
3	Hornet	Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
4	Hornet	Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2

1.Convert the column mpg to numpy array and print the array. (5)

```
In [1]: #Enter your code here
```

2. calculate how many cars' mpg is under 15. (5)

```
In [2]: #Enter your code here
```

3.Show the models that their mpg is under 15. (10)

```
In [3]: #Enter your code here
```

4.Show how many unique make does the dataset have and print them out. (10)

```
In [4]: #Enter your code here
```

5.Show how many cars each make has. (10)

In [5]: *#Enter your code here*

6. Create the scatter plot with X axis is hp and Y axis is mpg.(10)

In [6]: *#Enter your code here*

7. Train the linear model using hp as variable and MPG as target and print  $\theta_0, \theta_1$  (10)

(hint: use the formula:  $\theta = (X^T X)^{-1} X^T y$ )

In [7]: *#Enter your code here*

8. Plot the scatter plot again with the line showing the linear model.(10)

In [8]: *#Enter your code here*

9. Randomly choose 3 datapoint from the dataset, use the linear model to predict the MPG with HP, show the difference between predict value and real value. (10)

In [9]: *#Enter your code here*

10. Explain the why some differences are large, some are not(10)

In [13]: *#Enter your answer here*

11. Find  $R^2$  and use  $R^2$  to explain whether the model is accurate.(10)

In [14]: *#Enter your code here*