

## Lab Regression 1 (8/4/2022)

ລະຫັດນັກສຶກສາ: 205Q001019

ຊື່ ແລະ ນາມສະກຸນ: ທ ນຸຊິວ ເຮີ

ຈົ່ງຕອບຄໍາຖາມຕໍ່ໄປນີ້ໃຫ້ສໍາເລັດ:

### 1. Regression

1.1. ຈາກຊຸດຂໍ້ມູນ ( Data Set) **Salary\_Data.csv**, ຈົ່ງບອກຈຳນວນຖັນ, ແຖວ (shape) ແລະ ເພີ່ມຂໍ້ມູນໃນຕາຕະລາງລຸ່ມນີ້ໃຫ້ສໍາເລັດ:

ຈຳນວນຖັນ, ແຖວ (shape)

ຖັນ	2
ແຖວ	30

	YearsExperience	Salary
0	1.1	39343.0
1	1.3	46205.0
2	1.5	37731.0
3	2.0	43525.0
4	2.2	39891.0

1.2. ຈາກຊຸດຂໍ້ມູນ (Data Set) **Salary\_Data.csv**. ຈົ່ງຂຽນຄໍາສັ່ງເພື່ອແຍກຊຸດຂໍ້ມູນອອກເປັນສອງພາກສ່ວນ ຄື: ຊຸດທົດສອບ 75 ແລະ ຊຸດທົດສອບ 25 ?

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.25, random_state = 0)
```

✓ 0.2s

1.3. ຈົ່ງຮຽນຊຸດຂໍ້ມູນຂໍ້ 1.2 ດ້ວຍໂມເດວ LinearRegression ເພື່ອຄາດເດົາເງິນເດືອນ (Salary) ຂອງພະນັກງານທີ່ມີປະສົບ 5 ປີ ແລະ 9 ປີ?

➤ ມີປະສົບ 5 ປີ

```
y_pred_5_years = regressor.coef_*5 + regressor.intercept_ #salary 5 year
y_pred_5_years
```

✓ 0.3s

array([73885.24377647])

➤ ມີປະສົບ 9 ປີ

```
y_pred_9_years = regressor.coef_*9 + regressor.intercept_ #salary 9 year
y_pred_9_years
```

✓ 0.3s

array([111404.08574425])

1.4 ຈົ່ງສະແດງຂໍ້ມູນຊຸດຮຽນ (X\_train, y\_train) ດ້ວຍ scatter ຫຼື seaborn Graph ບົນພື້ນຖານຊຸດຄໍາສັ່ງ matplotlib

```
plt.scatter(X_train, y_train, color = 'red')
plt.plot(X_train, regressor.predict(X_train), color = 'blue')
plt.title('Salary vs Experience (Training set)')
plt.xlabel('Years of Experience')
plt.ylabel('Salary')
plt.show()
```

✓ 0.1s



1.5 ຈົ່ງສະແດງຂໍ້ມູນຊຸດທົດສອບ ( $X_{\text{test}}$ ,  $y_{\text{test}}$ ) ດ້ວຍ scatter ຫຼື seaborn Graph ບົນພື້ນຖານຊຸດຄຳສັ່ງ matplotlib

```
plt.scatter(X_test, y_test, color = 'red')
plt.plot(X_train, regressor.predict(X_train), color = 'blue')
plt.title('Salary vs Experience (Test set)')
plt.xlabel('Years of Experience')
plt.ylabel('Salary')
plt.show()
```

✓ 0.1s



1.6 ຖາມວ່າພະນັກງານສ່ວນຫຼາຍມີເງິນເດືອນຢູ່ຊ່ວງໃດ ແລະ ມີປະສິບການເຮັດວຽກຊ່ວງໃດ

- \* ພະນັກງານສ່ວນຫຼາຍມີເງິນເດືອນຢູ່ຊ່ວງ 40000-60000
- \* ມີປະສິບການເຮັດວຽກຊ່ວງ 2-4 ປີ

1.7 ຈົ່ງຄິດໄລ່ **R squared** (ຄວາມຖືກຕ້ອງໃນການຄາດເດົາ)

- `regressor.score(X_train, y_train)`  
0.9395413526983522

1.8 ເມື່ອ  $y = ax + b$  (Simple Linear Regression), ຈົ່ງຄິດໄລ່ **b intercept** ແລະ **a-coef**

- b -intercept

```
regressor.intercept_ |  
✓ 0.6s  
26986.691316737248
```

- a-coef

```
regressor.coef_  
✓ 0.3s  
array([9379.71049195])
```