Homework Lab 5.1 Logistic Regression (29/4/2022)

ລະຫັດນັກສຶກສາ: 205Q0010.19
ຊື່ ແລະ ນາມສະກຸນ: ທ້າວ ນູຊົ່ວ ເຮີ
ຈົ່ງຕອບຄຳຖາມຕໍ່ໄປນີ້ໃຫ້ສຳເລັດດ້ວຍການນຳໃຊ້ຄຳສັ່ງຂອງ Python:
1.ຈາກຊຸດຂໍ້ມູນ (Data Set) Social_Network_Ads.csv, ຈົ່ງບອກຈຳນວນຖັນ, ແຖວ (shape) ແລະ ເພີ່ມຂໍ້ມູນໃນຕາຕະລາງລຸ່ມນີ້ໃຫ້ສຳເລັດ:

ຖັນ	5
ແຖວ	400

	User ID	Gender	Age	EstimatedSalary	Purchased
	15624510	Male	19	19000	0
	15810944	Male	35	20000	0
	15668575	Female	26	43000	0

2. ຈົ່ງກຳນົດຕົວປ່ຽນເອກະລາດ (Independent Variables X) ໃຫ້ເປັນ User ID ແລະ EstimatedSalary. ກຳນົດຕົວປ່ຽນຕາມ (Dependent Variables y) ໃຫ້ເປັນ Purchased.

```
dataset = pd.read_csv('Social_Network_Ads.csv')
X = dataset.iloc[:, [0, 3]].values
y = dataset.iloc[:, -1].values
```

3. ຫຼັງຈາກຕຽມຊຸດຂໍ້ມູນສຳເລັດ, ຈົ່ງຂຽນຄຳສັ່ງເພື່ອແຍກຊຸດຂໍ້ມູນອອກເປັນສອງພາກສ່ວນຄື: ຊຸດຮຽນ 80 ແລະ ຊຸດທົດສອບ 20 ?

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.20, random_state = 0)
```

4. ຈົ່ງທຳການຕຽມຊຸດຂໍ້ມູນ (Preprocessing) ດ້ວຍການເຮັດ StandardScaler ຂອງຕົວປ່ຽນອິດສະຫຼະ X_trainແລະ X_test.

```
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)
```

5. ຈົ່ງສ້າງໂມເດວ Logistic Regression ແລະ ທຳການປະມວນຜົນ (fit) ຊຸດຂໍ້ມູນຮຽນຈາກຂໍ້ 4

```
from sklearn.linear_model import LogisticRegression
classifier = LogisticRegression(random_state = 0)
classifier.fit(X_train, y_train)
```

6. ຈົ່ງທົດສອບໂມເດວດ້ວຍການpredict(X_test).

```
y_pred = classifier.predict(X_test)
print(y_pred)
```

```
[0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 1 0 0 1 0 1 0 1 0 0 0 0 1 0 0 0 0 0 0
 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0
 0 0 0 0 0 1]
```

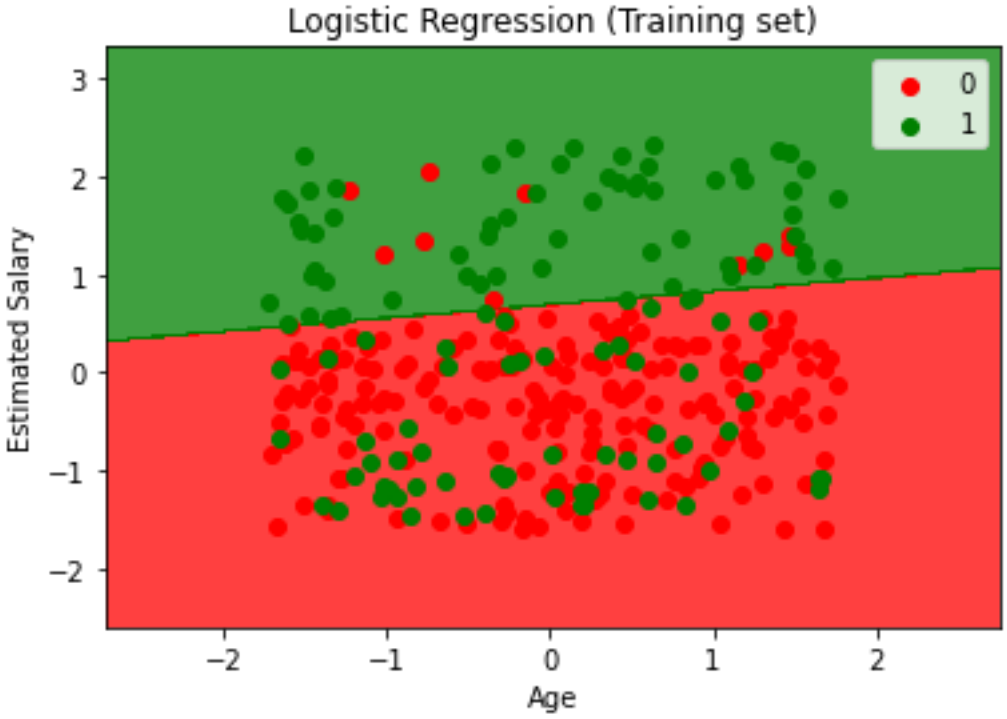
7. ຈົ່ງທຳການprocessing ດ້ວຍconfusion_matrix

```
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test, y_pred)
print(cm)
```

```
[[55 3]
 [12 10]]
```

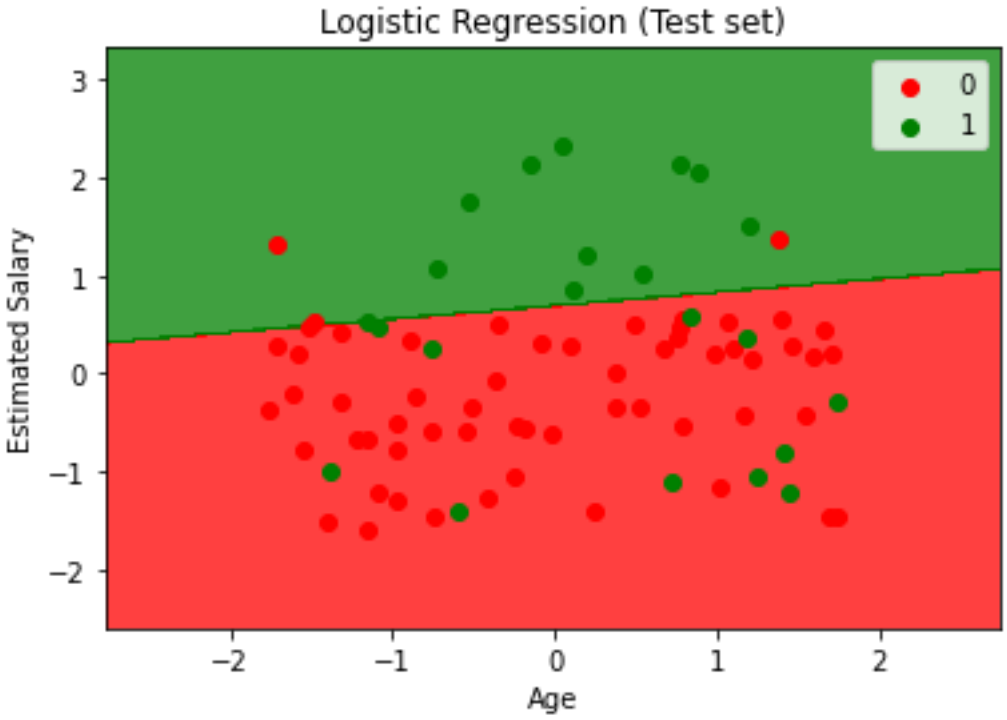
8.ຈົ່ງສະແດງຂໍ້ມູນຊຸດຮຽນ (X_train, y_train) ດ້ວຍGraph ບົນພື້ນຖານຊຸດຄຳສັ່ງ matplotlib.

```
from matplotlib.colors import ListedColormap
X_set, y_set = X_train, y_train
X1, X2 = np.meshgrid(np.arange(start = X_set[:, 0].min() - 1, stop = X_set[:, 0].max() + 1, step = 0.01),
                     np.arange(start = X_set[:, 1].min() - 1, stop = X_set[:, 1].max() + 1, step = 0.01))
plt.contourf(X1, X2, classifier.predict(np.array([X1.ravel(), X2.ravel()]).T).reshape(X1.shape),
             alpha = 0.75, cmap = ListedColormap(('red', 'green')))
plt.xlim(X1.min(), X1.max())
plt.ylim(X2.min(), X2.max())
for i, j in enumerate(np.unique(y_set)):
    plt.scatter(X_set[y_set == j, 0], X_set[y_set == j, 1],
                c = ListedColormap(('red', 'green'))(i), label = j)
plt.title('Logistic Regression (Training set)')
plt.xlabel('Age')
plt.ylabel('Estimated Salary')
plt.legend()
plt.show()
```



9.ຈົ່ງສະແດງຂໍ້ມູນຊຸດຮຽນ (X_test, y_test) ດ້ວຍGraph ບົນພື້ນຖານຊຸດຄຳສັ່ງ matplotlib.

```
from matplotlib.colors import ListedColormap
X_set, y_set = X_test, y_test
X1, X2 = np.meshgrid(np.arange(start = X_set[:, 0].min() - 1, stop = X_set[:, 0].max() + 1, step = 0.01),
                     np.arange(start = X_set[:, 1].min() - 1, stop = X_set[:, 1].max() + 1, step = 0.01))
plt.contourf(X1, X2, classifier.predict(np.array([X1.ravel(), X2.ravel()]).T).reshape(X1.shape),
             alpha = 0.75, cmap = ListedColormap(('red', 'green')))
plt.xlim(X1.min(), X1.max())
plt.ylim(X2.min(), X2.max())
for i, j in enumerate(np.unique(y_set)):
    plt.scatter(X_set[y_set == j, 0], X_set[y_set == j, 1],
                c = ListedColormap(('red', 'green'))(i), label = j)
plt.title('Logistic Regression (Test set)')
plt.xlabel('Age')
plt.ylabel('Estimated Salary')
plt.legend()
plt.show()
```



10. ຈົ່ງກຳນົດຕົວປ່ຽນເອກະລາດ (Independent Variables X) ໃຫ້ເປັນ User ID, Ageແລະ EstimatedSalary. ກຳນົດຕົວປ່ຽນຕາມ (Dependent Variables y) ໃຫ້ເປັນ Purchased.

```
dataset = pd.read_csv('Social_Network_Ads.csv')
X = dataset.iloc[:, [0,2, 3]].values
y = dataset.iloc[:, -1].values
```

11. ຫຼັງຈາກຕຽມຊຸດຂໍ້ມູນສຳເລັດ, ຈົ່ງຂຽນຄຳສັ່ງເພື່ອແຍກຊຸດຂໍ້ມູນອອກເປັນສອງພາກສ່ວນຄື: ຊຸດຮຽນ 85 ແລະ ຊຸດທົດສອບ 15 ?

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.15, random_state = 0)
```

12. ຈົ່ງທຳການຕຽມຊຸດຂໍ້ມູນ (Preprocessing) ດ້ວຍການເຮັດ StandardScaler ຂອງຕົວປ່ຽນອິດສະຫຼະ X_trainແລະ X_test.

```
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)
```

13. ຈົ່ງສ້າງໂມເດວ Logistic Regression ແລະ ທຳການປະມວນຜົນ (fit) ຊຸດຂໍ້ມູນຮຽນຈາກຂໍ້ 1.12

```
from sklearn.linear_model import LogisticRegression
classifier = LogisticRegression(random_state = 0)
classifier.fit(X_train, y_train)
```

14. จั่งทีดสอบโมเดอด้อยฑานpredict(X_test).

```
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test, y_pred)
print(cm)
```

[0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 1 0 0 1 0 1 0 1 0 0 0 0 0 0 1 0 0 0 0
 0 0 1 0 0 0 0 1 0 0 1 0 1 1 0 0 0 1 0 0 0 0 0]

15 จั่งทำฑานprocessing ด้อยconfusion_matrix

```
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test, y_pred)
print(cm)
```

[[44 1]
 [3 12]]