

BÀI TUẦN 05: LOGISTIC REGRESSION

1.Thông tin sinh viên

DUƠNG MINH LƯỢNG-18521071

2.Source

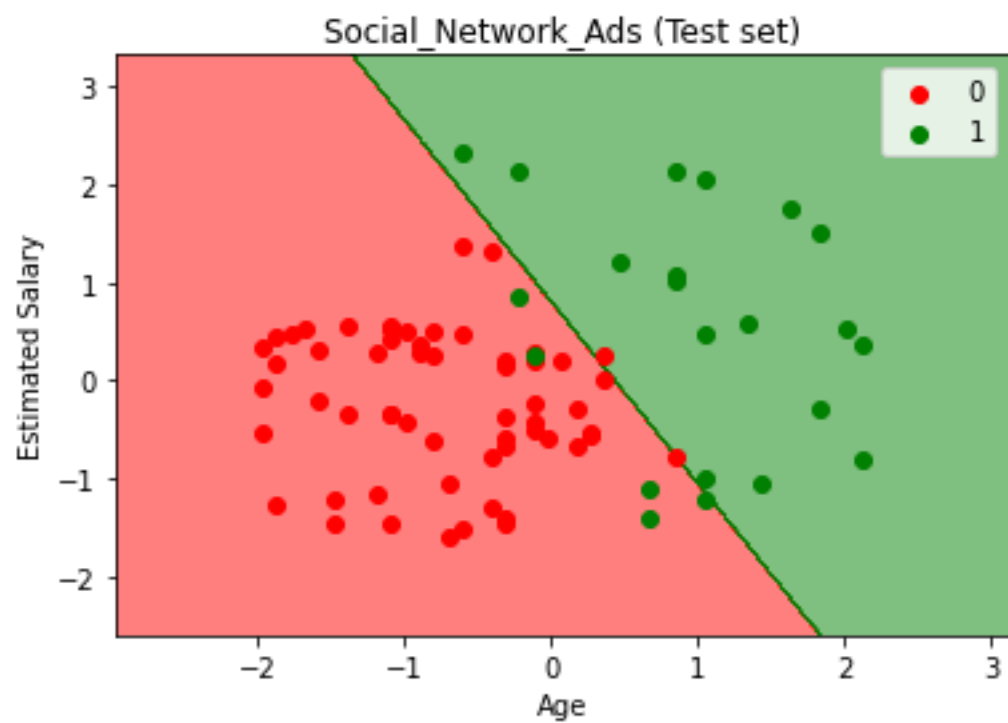
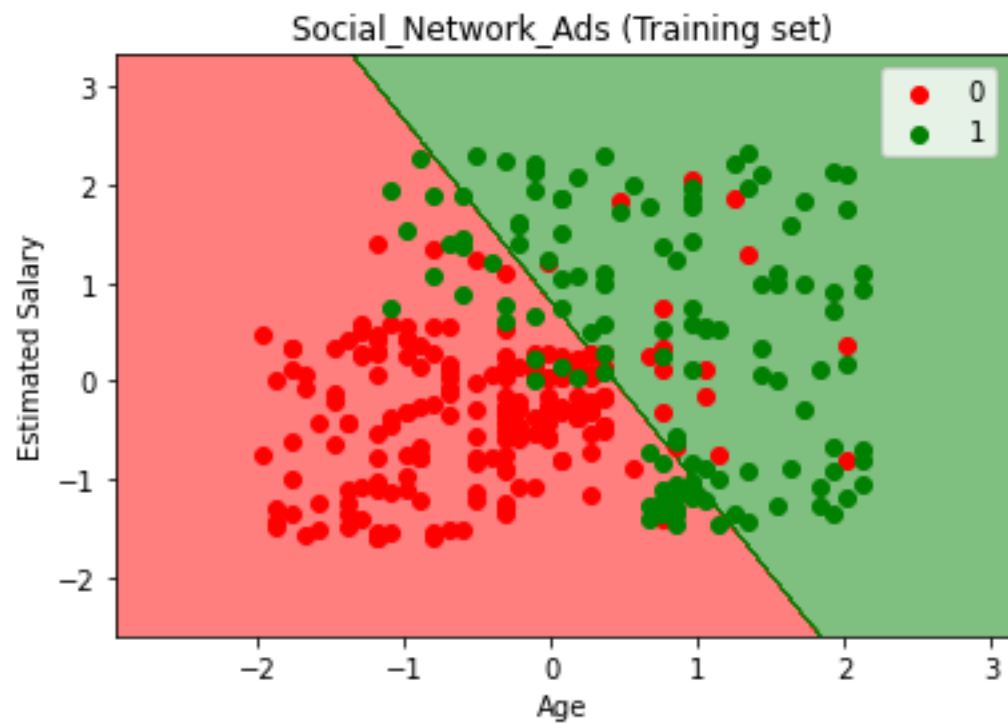
```
1. import numpy as np
2. import matplotlib.pyplot as plt
3. import pandas as pd
4. from matplotlib.colors import ListedColormap
5. # Importing the dataset
6. dataset = pd.read_csv('Social_Network_Ads.csv')
7. X = dataset.iloc[:, [2, 3]].values
8. y = dataset.iloc[:, -1].values
9. # Splitting the dataset into the Training set and Test set
10. from sklearn.model_selection import train_test_split
11. X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2,
    random_state = 0)
12. from sklearn.preprocessing import StandardScaler
13. sc = StandardScaler()
14. X_train = sc.fit_transform(X_train)
15. X_test = sc.transform(X_test)
16.
17. from sklearn.linear_model import LogisticRegression
18. classifier = LogisticRegression(random_state = 0)
19. classifier.fit(X_train, y_train)
20.
21. y_pred = classifier.predict(X_test)
22. # Making the Confusion Matrix
23. from sklearn.metrics import confusion_matrix
24. cm = confusion_matrix(y_test, y_pred)
25. print(cm)
26. def VisualizingDataset(X_, Y_):
27.     X1 = X_[:, 0]
28.     X2 = X_[:, 1]
29.     for i, label in enumerate(np.unique( Y_)):
30.         plt.scatter(X1[Y_ == label], X2[Y_ == label],color =
            ListedColormap(("red", "green"))(i),label = label)
```

```

31. def VisualizingResult(model, X_):
32.     X1 = X_[:, 0]
33.     X2 = X_[:, 1]
34.     X1_range = np.arange(start= X1.min()-1, stop= X1.max()+1,step =
        0.01)
35.     X2_range = np.arange(start= X2.min()-1, stop= X2.max()+1,step =
        0.01)
36.     X1_matrix, X2_matrix = np.meshgrid(X1_range, X2_range)
37.     X_grid= np.array([X1_matrix.ravel(), X2_matrix.ravel()]).T
38.     Y_grid= model.predict(X_grid).reshape(X1_matrix.shape)
39.     plt.contourf(X1_matrix, X2_matrix, Y_grid, alpha = 0.5,cmap =
        ListedColormap(("red", "green")))
40. VisualizingResult(classifier, X_train)
41. VisualizingDataset(X_train, y_train)
42. plt.title('Social_Network_Ads (Training set)')
43. plt.xlabel('Age')
44. plt.ylabel('Estimated Salary')
45. plt.legend()
46. plt.show()
47. VisualizingResult(classifier, X_test)
48. VisualizingDataset(X_test, y_test)
49. plt.title('Social_Network_Ads (Test set)')
50. plt.xlabel('Age')
51. plt.ylabel('Estimated Salary')
52. plt.legend()
53. plt.show()

```

3. Kết quả



— Confusion Matrix được in ra từ test là:

	0	1
0	57	1
1	5	17

Nhận xét:

- Theo ma trận trên, số lượng dữ liệu được phân loại đúng là $57 + 17 = 74$ điểm dữ liệu.
- Số lượng dữ liệu phân loại sai là $1 + 5 = 6$ điểm dữ liệu.
- Tỷ lệ điểm dữ liệu phân loại sai là $6/80 \approx 0.075 \approx 7.5\%$