

BÀI TUẦN 05: K NEAREST NEIGHBORS

1.Thông tin sinh viên

DƯƠNG MINH LƯỢNG-18521071

2.Source

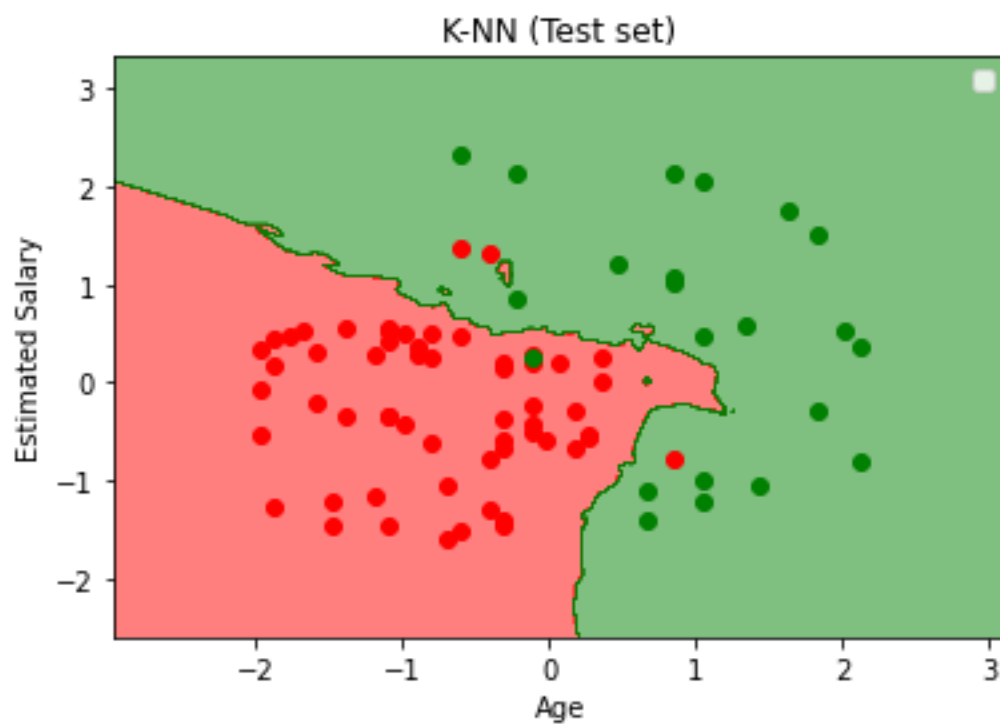
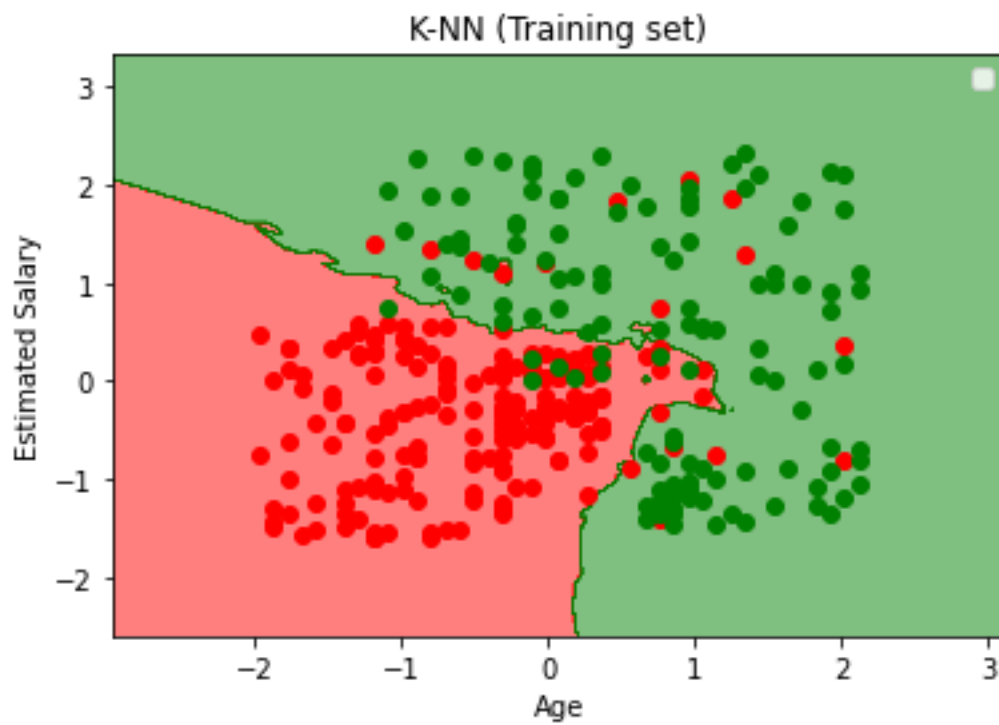
```
1. import numpy as np
2. import matplotlib.pyplot as plt
3. import pandas as pd
4. from matplotlib.colors import ListedColormap
5.
6. # Importing the dataset
7. dataset = pd.read_csv('Social_Network_Ads.csv')
8. X = dataset.iloc[:, [2, 3]].values
9. y = dataset.iloc[:, -1].values
10. from sklearn.model_selection 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. from sklearn.neighbors import KNeighborsClassifier
17. classifier = KNeighborsClassifier(n_neighbors = 5)
18. classifier.fit(X_train, y_train)
19. y_pred = classifier.predict(X_test)
20. from sklearn.metrics import confusion_matrix
21. cm = confusion_matrix(y_test, y_pred)
22. print("Test\n",cm)
23. cm1 = confusion_matrix(y_train,classifier.predict(X_train))
24. print("Train\n",cm1)
25. def VisualizingDataset(X_, Y_):
26.     X1 = X_[:, 0]
27.     X2 = X_[:, 1]
28.     for i, label in enumerate(np.unique(Y_)):
29.         plt.scatter(X1[Y_ == label], X2[Y_ == label],color =
            ListedColormap(("red", "green"))(i))
30. def VisualizingResult(model, X_):
```

```

31. X1 = X[:, 0]
32. X2 = X[:, 1]
33. X1_range = np.arange(start= X1.min()-1, stop= X1.max()+1,step =
    0.01)
34. X2_range = np.arange(start= X2.min()-1, stop= X2.max()+1,step =
    0.01)
35. X1_matrix, X2_matrix = np.meshgrid(X1_range, X2_range)
36. X_grid= np.array([X1_matrix.ravel(), X2_matrix.ravel()]).T
37. Y_grid= model.predict(X_grid).reshape(X1_matrix.shape)
38. plt.contourf(X1_matrix, X2_matrix, Y_grid, alpha = 0.5,cmap =
    ListedColormap(("red", "green")))
39.
40. VisualizingResult(classifier, X_train)
41. VisualizingDataset(X_train, y_train)
42. plt.title('K-NN (Training set)')
43. plt.xlabel('Age')
44. plt.ylabel('Estimated Salary')
45. plt.legend()
46. plt.show()
47.
48. VisualizingResult(classifier, X_test)
49. VisualizingDataset(X_test, y_test)
50. plt.title('K-NN (Test set)')
51. plt.xlabel('Age')
52. plt.ylabel('Estimated Salary')
53. plt.legend()
54. plt.show()
55.

```

3. Kết quả



	0	1
0	55	3
1	1	21

Nhận xét:

- Có 55 người không mua hàng dự đoán không mua hàng
- Có 3 người không mua hàng nhưng dự đoán là mua hàng
- Có 1 người mua hàng nhưng dự đoán không mua hàng
- Có 21 người mua hàng dự đoán mua hàng
- Tổng số dự đoán đúng là $55+21=76$
- Tổng số dự đoán sai là $3+1=4$
- Xác suất sai là $4/80=0.05=5\%$