ML3 Classification Analysis

October 26, 2023

[1]:
```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import confusion_matrix, accuracy_score, precision_score,␣
 ↪recall_score

data = pd.read_csv("C:/Users/hp/Downloads/Practical_Data/Social_Network_Ads.
 ↪csv")
data.head()
```

```
C:\Users\hp\anaconda3\lib\site-packages\scipy\__init__.py:146: UserWarning: A
NumPy version >=1.16.5 and <1.23.0 is required for this version of SciPy
(detected version 1.25.2
  warnings.warn(f"A NumPy version >={np_minversion} and <{np_maxversion}"
```

[1]:

	User ID	Gender	Age	EstimatedSalary	Purchased
0	15624510	Male	19	19000	0
1	15810944	Male	35	20000	0
2	15668575	Female	26	43000	0
3	15603246	Female	27	57000	0
4	15804002	Male	19	76000	0

[2]:
```
X = data.iloc[:, [2, 3]]  # Features (Age and EstimatedSalary columns)
y = data['Purchased']     # Target variable

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
 ↪random_state=42)
```

[3]:
```
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)

k = 5  # Number of neighbors
knn_classifier = KNeighborsClassifier(n_neighbors=k)
knn_classifier.fit(X_train_scaled, y_train)
```

```
y_pred = knn_classifier.predict(X_test_scaled)
confusion_matrix(y_test, y_pred)
```

[3]:
```
array([[48,  4],
       [ 3, 25]], dtype=int64)
```

[4]:
```
accuracy = accuracy_score(y_test, y_pred)
error_rate = 1 - accuracy
precision = precision_score(y_test, y_pred)
recall = recall_score(y_test, y_pred)

print("Accuracy:", accuracy*100,"%")
print("Error Rate:", error_rate)
print("Precision:", precision)
print("Recall:", recall)
```

```
Accuracy: 91.25 %
Error Rate: 0.08750000000000002
Precision: 0.8620689655172413
Recall: 0.8928571428571429
```

[ ]: