assignment6

May 2, 2024

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

[2]:
```
df = pd.read_csv("Social_Network_Ads.csv")
```

[3]:
```
df.head()
```

[3]:

	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

[4]:
```
df_x = df.iloc[:, [2, 3]]
df_x.head()
```

[4]:

	Age	EstimatedSalary
0	19	19000
1	35	20000
2	26	43000
3	27	57000
4	19	76000

[5]:
```
df_y = df.iloc[:, 4]
df_y
```

[5]:
```
0      0
1      0
2      0
3      0
4      0
```

```
      ..
395    1
396    1
397    1
398    0
399    1
Name: Purchased, Length: 400, dtype: int64
```

[6]:
```
X_train, X_test, y_train, y_test = train_test_split(df_x, df_y, test_size=0.20,␣
↪random_state=2)
```

[7]:
```
sc = preprocessing.StandardScaler()
```

[8]:
```
x_train = sc.fit_transform(X_train)
```

[9]:
```
log_reg = LogisticRegression(random_state=0)
```

[10]:
```
log_reg.fit(x_train, y_train)
```

[10]:
```
LogisticRegression(random_state=0)
```

[11]:
```
y_pred = log_reg.predict(X_test)
```

```
c:\ProgramData\anaconda3\lib\site-packages\sklearn\base.py:413: UserWarning: X
has feature names, but LogisticRegression was fitted without feature names
  warnings.warn(
```

[12]:
```
cm = confusion_matrix(y_test, y_pred)
```

[13]:
```
cm
```

[13]:
```
array([[ 0, 48],
       [ 0, 32]], dtype=int64)
```

[14]:
```
ac = accuracy_score(y_true=y_test, y_pred=y_pred)
```

[15]:
```
ac
```

[15]:
```
0.4
```

[16]:
```
from sklearn.metrics import classification_report
```

[17]:
```
report = classification_report(y_test, y_pred)
```

```
c:\ProgramData\anaconda3\lib\site-
packages\sklearn\metrics\_classification.py:1344: UndefinedMetricWarning:
Precision and F-score are ill-defined and being set to 0.0 in labels with no
```

```
predicted samples. Use `zero_division` parameter to control this behavior.
  _warn_prf(average, modifier, msg_start, len(result))
c:\ProgramData\anaconda3\lib\site-
packages\sklearn\metrics\_classification.py:1344: UndefinedMetricWarning:
Precision and F-score are ill-defined and being set to 0.0 in labels with no
predicted samples. Use `zero_division` parameter to control this behavior.
  _warn_prf(average, modifier, msg_start, len(result))
c:\ProgramData\anaconda3\lib\site-
packages\sklearn\metrics\_classification.py:1344: UndefinedMetricWarning:
Precision and F-score are ill-defined and being set to 0.0 in labels with no
predicted samples. Use `zero_division` parameter to control this behavior.
  _warn_prf(average, modifier, msg_start, len(result))
```

[18]:
```
print(report)
```

```
              precision    recall  f1-score   support

           0       0.00      0.00      0.00        48
           1       0.40      1.00      0.57        32

    accuracy                           0.40        80
   macro avg       0.20      0.50      0.29        80
weighted avg       0.16      0.40      0.23        80
```

[19]:
```
print(1-ac)
```

```
0.6
```

[20]:
```
# Find precision score
pr_sc = precision_score(y_test, y_pred)
pr_sc
```

[20]: 0.4

[21]:
```
rec_score = recall_score(y_test, y_pred)
rec_score
```

[21]: 1.0