Exercise 2: Play Goft

Cho dữ liệu play golf trong tập tin playgoft_data.xlsx.

Yêu cầu: Hãy đọc dữ liệu từ tập tin này, áp dụng Naive Bayes để thực hiện việc xác định có đi chơi golf hay không dựa trên các thông tin như: 'Outlook', 'Temperature', 'Humidity', 'Wind', 'Play Golf'

Yêu cầu:

1. Hãy chuẩn hóa dữ liệu cho phù hợp
2. Áp dụng Naive Bayes. Tìm kết quả
3. Cho dữ liệu Test: X_test = [["Overcast", "Cool", "High", "Strong"], ["Sunny", "Cool", "High", "Weak"]] => Y_test
4. Kiểm tra độ chính xác

In [1]:
```python
import pandas as pd
```

In [2]:
```python
df = pd.read_excel('playgolf_data.xlsx', index_col = 0)
df
```

Out[2]:

Day	Outlook	Temperature	Humidity	Wind	Play Golf
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

In [3]:
```python
# chuan hoa:
# Cột Outlook: 0: Sunny, 1: Overcast, 2: Rain
# Cột Temperature: 0: Hot, 1: Mild, 2: Cool
# Cột Humidity: 0: High, 1: Normal
# Cột Wind: 0: Weak, 1: Strong
df = df.replace("Sunny", 0)
df = df.replace("Overcast", 1)
df = df.replace("Rain", 2)
df = df.replace("Hot", 0)
df = df.replace("Mild", 1)
df = df.replace("Cool", 2)
df = df.replace("High", 0)
df = df.replace("Normal", 1)
df = df.replace("Weak", 0)
df = df.replace("Strong", 1)
df = df.replace("No", 0)
df = df.replace("Yes", 1)
df
```

Out[3]:

Day	Outlook	Temperature	Humidity	Wind	Play Golf
1	0	0	0	0	0
2	0	0	0	1	0
3	1	0	0	0	1
4	2	1	0	0	1
5	2	2	1	0	1
6	2	2	1	1	0
7	1	2	1	1	1
8	0	1	0	0	0
9	0	2	1	0	1
10	2	1	1	0	1
11	0	1	1	1	1
12	1	1	0	1	1
13	1	0	1	0	1
14	2	1	0	1	0

In [4]:
```python
features = df[["Outlook","Temperature","Humidity","Wind"]]
target = df[["Play Golf"]]
```

In [5]: features

Out[5]:

Day	Outlook	Temperature	Humidity	Wind
1	0	0	0	0
2	0	0	0	1
3	1	0	0	0
4	2	1	0	0
5	2	2	1	0
6	2	2	1	1
7	1	2	1	1
8	0	1	0	0
9	0	2	1	0
10	2	1	1	0
11	0	1	1	1
12	1	1	0	1
13	1	0	1	0
14	2	1	0	1

In [6]: target

Out[6]:

Day	Play Golf
1	0
2	0
3	1
4	1
5	1
6	0
7	1
8	0
9	1
10	1
11	1
12	1
13	1
14	0

In [7]:
```python
from sklearn.naive_bayes import GaussianNB
import numpy as np
from sklearn.utils.validation import column_or_1d

#Create a Gaussian Classifier
model = GaussianNB()

# Train the model using the training sets
model.fit(features, target)
```

```
c:\program files\python36\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expecte
d. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
```

Out[7]: GaussianNB(priors=None)

In [8]:
```python
import numpy as np
# Kiểm tra độ chính xác
print("The prediction accuracy is: ", model.score(features,np.array(target))*100,"
```

```
The prediction accuracy is:  78.57142857142857 %
```

In [9]:
```python
class_names = model.classes_
class_names
```

Out[9]: array([0, 1], dtype=int64)

In [10]:
```python
# X_test = [["Overcast", "Cool", "High", "Strong"], ["Sunny", "Cool", "High", "Wea
X_test = [[1, 2, 0, 1], [0, 2, 0, 0]]
y_pred = model.predict(X_test)
y_pred
```

Out[10]: array([0, 1], dtype=int64)

In [ ]: