HW2&3_question8_Naive_Bayes

March 25, 2021

[1]:
```
import pandas as pd

# read the iris dataset which is csv format
col_names = ["sepal_length", "sepal_width", "petal_length", "petal_width",␣
 ↪"species"]
iris = pd.read_csv("iris.data", header=None, names=col_names)
```

[2]:
```
# map iris class name to number
iris_class = {'Iris-setosa':0, 'Iris-versicolor':1, 'Iris-virginica':2}
iris['species_tag'] = [iris_class[i] for i in iris.species]
```

[3]:
```
#split data into attributes and target/label
iris_attrs = iris.drop(['species','species_tag'], axis=1)
iris_labels = iris.species_tag
```

[4]:
```
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
# create GaussianNB
gnb = GaussianNB()
# avg of score
avg = 0

# run 10 times
for i in range(10):
    # split data into training and testing sets
    train_data, test_data, train_label, test_label =␣
 ↪train_test_split(iris_attrs, iris_labels,
                                                                    ␣
 ↪random_state=None, train_size=0.7)
    # fit the model on the training data
    gnb.fit(train_data, train_label)
    # see how the model preforms
    avg = avg + gnb.score(test_data, test_label)

# average accuracy
print('avg:',avg/10)
```

```
avg: 0.9377777777777778
```