HW2&3_question7_KNN_Regression

March 25, 2021

[8]:
```
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)
```

[9]:
```
#split data into attributes and target/label
iris_attrs = iris.drop(['petal_width','species'], axis=1)
iris_labels = iris.petal_width
```

[11]:
```
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsRegressor
from sklearn.metrics import mean_squared_error

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)
    # set 5 neighbors of knn
    knn = KNeighborsRegressor(n_neighbors = 5, weights='distance')
    # fit the model on the training data
    knn.fit(train_data, train_label)
    # make prediction
    knn_pred = knn.predict(test_data)
    # calculate by Mean square error
    error = mean_squared_error(test_label,knn_pred)
    avg = avg + error

# average accuracy and best-fit of k-neighbor
print('avg of MSE',avg/10)
```

```
avg of MSE 0.03751805248738778
```