KNN_iris

July 5, 2022

[ ]:
```
import pandas as pd
import seaborn as sns
import numpy as np
from sklearn.datasets import load_iris
# load iris flower data from scikit-learn

iris = load_iris()
```

[ ]:
```
df = pd.DataFrame(data = iris.data, columns = iris.feature_names)
df['target'] = iris.target
df
```

[ ]:
```
     sepal length (cm)  sepal width (cm)  petal length (cm)  petal width (cm)  \
0                  5.1               3.5                1.4               0.2
1                  4.9               3.0                1.4               0.2
2                  4.7               3.2                1.3               0.2
3                  4.6               3.1                1.5               0.2
4                  5.0               3.6                1.4               0.2
..                 …                 …                  …                 …
145                6.7               3.0                5.2               2.3
146                6.3               2.5                5.0               1.9
147                6.5               3.0                5.2               2.0
148                6.2               3.4                5.4               2.3
149                5.9               3.0                5.1               1.8

     target
0         0
1         0
2         0
3         0
4         0
..      …
145       2
146       2
147       2
148       2
149       2
```

```
[150 rows x 5 columns]
```

[ ]:
```
df.columns=['sl','sw','pl','pw','label']
df
```

[ ]:
```
      sl   sw   pl   pw  label
0    5.1  3.5  1.4  0.2      0
1    4.9  3.0  1.4  0.2      0
2    4.7  3.2  1.3  0.2      0
3    4.6  3.1  1.5  0.2      0
4    5.0  3.6  1.4  0.2      0
..   …  …  …  …    …
145  6.7  3.0  5.2  2.3      2
146  6.3  2.5  5.0  1.9      2
147  6.5  3.0  5.2  2.0      2
148  6.2  3.4  5.4  2.3      2
149  5.9  3.0  5.1  1.8      2

[150 rows x 5 columns]
```

[ ]:
```
from sklearn.model_selection import train_test_split
train, test = train_test_split(df, test_size = 0.2)
```

[ ]:
```
from sklearn.neighbors import KNeighborsClassifier
```

[ ]:
```
knn = KNeighborsClassifier(n_neighbors = 5)
```

[ ]:
```
x_train = train[['pl','pw']]
y_train = train[['label']]
```

[ ]:
```
knn.fit(x_train, y_train.values.ravel())
```

[ ]:
```
KNeighborsClassifier()
```

[ ]:
```
x_test = test[['pl','pw']]
y_test = test[['label']]
```

[ ]:
```
predictions = knn.predict(x_test)
predictions
```

[ ]:
```
array([2, 2, 1, 0, 1, 0, 0, 1, 0, 2, 1, 2, 0, 0, 2, 2, 1, 0, 1, 1, 1, 1,
       2, 1, 1, 1, 1, 1, 0, 0])
```

[ ]:
```
comparison = pd.DataFrame(
    {'pred' : predictions, 'truth' : y_test.values.ravel()})
comparison
```

[ ]:

	pred	truth
0	2	2
1	2	2
2	1	1
3	0	0
4	1	1
5	0	0
6	0	0
7	1	1
8	0	0
9	2	2
10	1	1
11	2	2
12	0	0
13	0	0
14	2	2
15	2	2
16	1	1
17	0	0
18	1	1
19	1	1
20	1	1
21	1	1
22	2	1
23	1	1
24	1	1
25	1	1
26	1	1
27	1	1
28	0	0
29	0	0

[ ]: