Webinar_HandsOn_Part2_KNNClassifier

July 21, 2019

```
In [1]: import numpy as np
        import pandas as pd
        import matplotlib.pyplot as plt
        from sklearn.model_selection import train_test_split

In [2]: cd Downloads

/home/subarna1/Downloads


In [3]: fruits=pd.read_table('fruit_data_with_colors.txt')

In [4]: fruits.head()

Out[4]:    fruit_label fruit_name fruit_subtype  mass  width  height  color_score
        0            1      apple  granny_smith   192    8.4     7.3         0.55
        1            1      apple  granny_smith   180    8.0     6.8         0.59
        2            1      apple  granny_smith   176    7.4     7.2         0.60
        3            2   mandarin      mandarin    86    6.2     4.7         0.80
        4            2   mandarin      mandarin    84    6.0     4.6         0.79

In [5]: fruits.tail()

Out[5]:    fruit_label fruit_name fruit_subtype  mass  width  height  color_score
        54           4      lemon       unknown   116    6.1     8.5         0.71
        55           4      lemon       unknown   116    6.3     7.7         0.72
        56           4      lemon       unknown   116    5.9     8.1         0.73
        57           4      lemon       unknown   152    6.5     8.5         0.72
        58           4      lemon       unknown   118    6.1     8.1         0.70

In [6]: fruits.shape

Out[6]: (59, 7)

In [7]: fruits.describe()

Out[7]:        fruit_label        mass      width     height  color_score
        count    59.000000   59.000000  59.000000  59.000000    59.000000
        mean      2.542373  163.118644   7.105085   7.693220     0.762881
```

```
std       1.208048   55.018832   0.816938   1.361017     0.076857
min       1.000000   76.000000   5.800000   4.000000     0.550000
25%       1.000000  140.000000   6.600000   7.200000     0.720000
50%       3.000000  158.000000   7.200000   7.600000     0.750000
75%       4.000000  177.000000   7.500000   8.200000     0.810000
max       4.000000  362.000000   9.600000  10.500000     0.930000
```

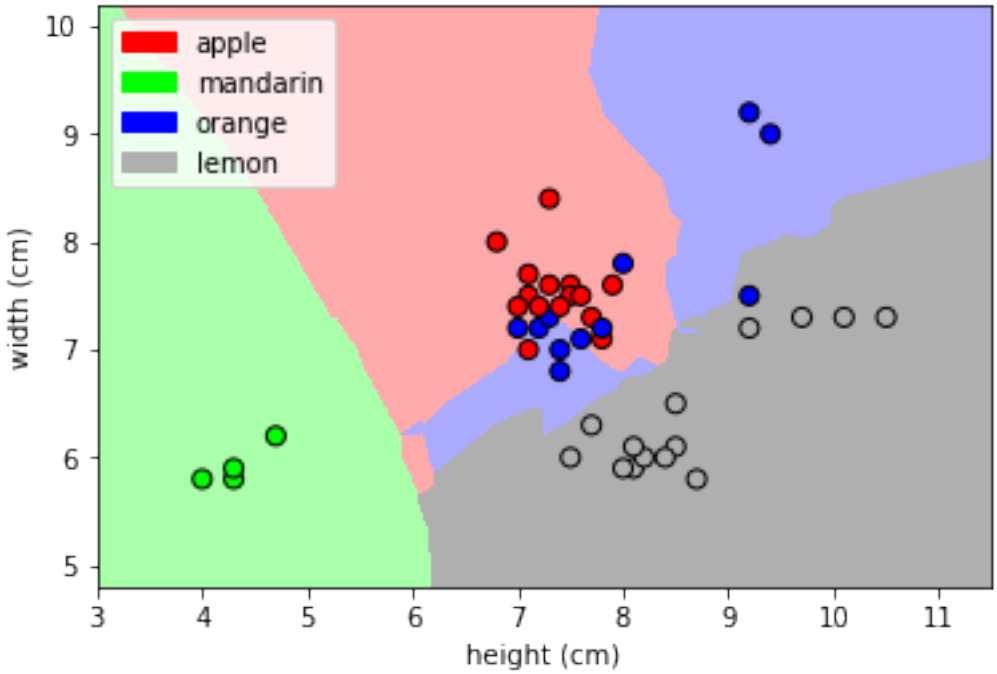
```
In [8]: #finding the independent and dependent variables

In [9]: fruits.head()
```

Out[9]:

	fruit_label	fruit_name	fruit_subtype	mass	width	height	color_score
0	1	apple	granny_smith	192	8.4	7.3	0.55
1	1	apple	granny_smith	180	8.0	6.8	0.59
2	1	apple	granny_smith	176	7.4	7.2	0.60
3	2	mandarin	mandarin	86	6.2	4.7	0.80
4	2	mandarin	mandarin	84	6.0	4.6	0.79

```
In [10]: look_up_fruit_name= dict(zip(fruits.fruit_label.unique(), fruits.fruit_name.unique()))

In [11]: look_up_fruit_name

Out[11]: {1: 'apple', 2: 'mandarin', 3: 'orange', 4: 'lemon'}

In [12]: #seperating into X and y

In [13]: X=fruits[['mass','width','height','color_score']]
         y=fruits['fruit_label']

In [14]: #splitting part where we split into train, test

In [15]: X_train,X_test,y_train,y_test= train_test_split(X,y,random_state=0)

In [16]: #create a classifier object

In [17]: from sklearn.neighbors import KNeighborsClassifier

In [18]: knn=KNeighborsClassifier(n_neighbors=5) #instantiation the KNeighborsClassifer

In [19]: knn.fit(X_train,y_train)   #training the classifier using the train data

Out[19]: KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
                    metric_params=None, n_jobs=None, n_neighbors=5, p=2,
                    weights='uniform')

In [20]: knn.score(X_test,y_test) #estimating the accuracy of this model

/home/subarna1/.local/lib/python2.7/site-packages/sklearn/externals/joblib/parallel.py:268: Dep
  ' removed in 0.13', DeprecationWarning)
```

```
Out[20]: 0.5333333333333333

In [21]: # Prediction TIme

In [22]: fruit_predict=knn.predict([[20,4.3,5.5,0.32]])
         fruit_predict

/home/subarna1/.local/lib/python2.7/site-packages/sklearn/externals/joblib/parallel.py:268: Dep
  ' removed in 0.13', DeprecationWarning)


Out[22]: array([2])

In [23]: look_up_fruit_name[fruit_predict[0]]

Out[23]: 'mandarin'

In [24]: fruit_predict=knn.predict([[100,5.3,9,0.34]])
         look_up_fruit_name[fruit_predict[0]]

/home/subarna1/.local/lib/python2.7/site-packages/sklearn/externals/joblib/parallel.py:268: Dep
  ' removed in 0.13', DeprecationWarning)


Out[24]: 'lemon'

In [25]: #plot the decision boundaries of the KNN classifier

In [26]: cd Downloads

/home/subarna1/Downloads/Downloads


In [34]: from adspy_shared_utilities import plot_fruit_knn

In [35]: plot_fruit_knn(X_train,y_train,4,'uniform')

adspy_shared_utilities.py:172: FutureWarning: Method .as_matrix will be removed in a future ver
  X_mat = X[['height', 'width']].as_matrix()
adspy_shared_utilities.py:173: FutureWarning: Method .as_matrix will be removed in a future ver
  y_mat = y.as_matrix()
/home/subarna1/.local/lib/python2.7/site-packages/sklearn/externals/joblib/parallel.py:268: Dep
  ' removed in 0.13', DeprecationWarning)
```


apple
mandarin
orange
lemon
width (cm)
height (cm)
10
9
8
7
6
5
3
4
5
6
7
8
9
10
11

In [ ]: