

# SIT384 Cyber security analytics

## Pass Task 6.1P: kNN classification

### Task description:

KNN is a non-parametric learning algorithm. Its purpose is to use a database in which the data points are separated into several classes to predict the classification of a new sample point.

You are given a dataset task6\_1\_dataset.csv. The first column is index, x1 and x2 can be treated as the coordinates of a point, and y is the class the point (x1,x2) belongs to.

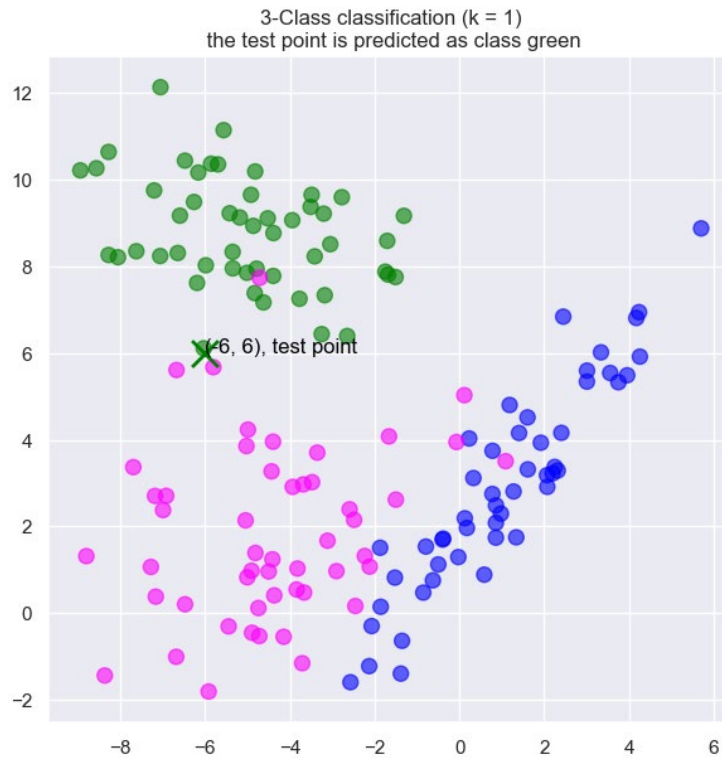
	x1	x2	y
0	1.68223	7.81035	0
1	7.20088	9.754777	0
2	6.47379	10.44102	0
3	5.98175	8.025655	0
4	4.52135	9.10985	0
5	3.94231	9.067026	0
...	...	...	...

(The above data is for demonstration purposes only. Please download the full version of task6\_1\_dataset.csv.)

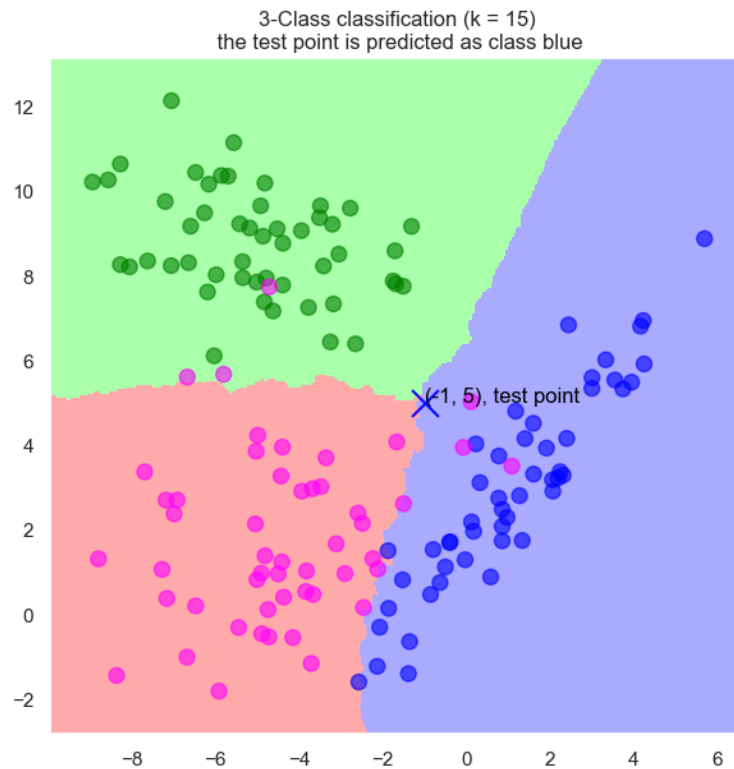
You are asked to :

- train the above dataset with k=1
- test the kNN classifier with [-6, 6] (1<sup>st</sup> testing point)
- visualize data and the 1<sup>st</sup> testing point using scatter plot.
  - X axis is x1
  - Y axis is x2
  - Testing point's marker is "x", with a text label in color black.
  - plt.subplots(figsize=(7, 5), dpi=100)
  - scale = 75
  - alpha = 0.6
  - colors = ['green', 'blue', 'magenta'] representing class 0, 1 and 2; or colors of your choice
  - For a training data point, its color is the y class color; for a testing point, color is the predicted class color
  - set the plot title according to the screenshot below
  - other settings of your choice

Sample output as shown in the following figure is for demonstration purposes only.



- train the above dataset with k=15
- test the kNN classifier with [-1, 5] (2<sup>nd</sup> testing point)
- visualize data and the 2<sup>nd</sup> testing point using scatter plot.
  - Create color maps/decision boundaries  
`cmap_light = ListedColormap(['#AAFFAA', '#AAAAFF', '#FFAAAA'])`
  - X axis is x1
  - Y axis is x2
  - Testing point's marker is "x", with a text label in color black.
  - `plt.subplots(figsize=(7, 5), dpi=100)`
  - For a training data point, its color is the y class color; for a testing point, color is the predicted class color
  - set the plot title according to the screenshot below
  - other settings of your choice



#### Submission:

Submit the following files to OnTrack:

1. Your program source code (e.g. task6\_1.py)
2. A screen shot of your program running

Check the following things before submitting:

1. Add proper comments to your code