CS7052-Machine Learning

Workshop 10: Unsupervised Learning, Clustering

You will learn:

- To practice parameter tuning and comparison of performances of different supervised machine learning models
- To practice unsupervised learning models, mainly clustering with several datasets

Task 1

Download the Jupyter Notebook file in week 10 folder. Download the image labeled as KNN.

W10.1. Complete the code in week 10 folder to display the expected output in the image file (knn plot.png); write a snippet of code to draw a plot displaying training and test accuracies of KNNClassifier for different values of K from 1 to 20.

W10.2. Apply logistic regression on the same model and compare the performance.

W10.3. Apply decision tree classifier and compare the performance.

W10.4. Apply MLP classifier and compare the performance.

W10.5. Which model is outperforming the rest?

Task 2

Open Muller & Guido's book from the reading list. Open your Jupyter notebook.

Follow the instructions on pages 170-178 and 184-189 and 194-196 in chapter 3.

Muller and Guido's book comes with accompanying code, which you can find on https://github.com/amueller/introduction_to_ml_with_python.

You can download the code and then open corresponding file to chapter 3 (03-unsupervised-learning.ipynb) in your Jupyter Notebook. Make sure you understand the meaning of each line of code, make some changes to improve your understanding and answer the following questions:

W10.6. Change number of clusters on page 174 to 3 and then 4 and upload cluster assignment figures.

W10.7. Which of the four assignments in Figure 3040 on page 196 has the best score? Which one do you think is the best assignment according to the figure?

Show the output to your tutor when you are done.