CS7052-Machine Learning

Workshop 3: k-NN classification and Regression

You will learn:

- To practice k-NN classification and Regression with several datasets
- Complete task 2 from last week's workshop and answer test 2

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

Follow the instructions on pages 31-46 in chapter 2.

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 2 (02-supervised-learning.ipynb) in your Jupyter Notebook.

Make sure you understand the meaning of each line of code, make some changes to improve your understanding.

Answer following questions and try suggested changes to the above code, upload your answers on test 3 under week 3 folder.

W3.1 Which of the datasets you used today are synthetic datasets? And what is the meaning of a synthetic dataset?

W3.2 Draw decision boundaries for 2, 5 and 8 neighbours in figure 2-6

W3.3 Where is the 'sweet spot' in the cancer dataset?

W3.4 Draw the prediction plot for two and six-nearest-neighbours regression on the wave dataset.

W3.5 What are the prediction scores of two and six-nearest-neighbours regression on the wave dataset?

W3.6 Draw the plot of figure 2.10 for 2, 4 and 6 neighbours. What are the train and test scores in for 2, 4 and 6 neighbours?

W3.7 Show the output to your tutor when you are done.