CS7052-Machine Learning

Workshop 7: Decision Trees

You will learn:

- To practice Decision Trees for Regression and classification with several datasets

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

Follow the instructions on pages 72- 84 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 and answer the following questions:

W7.1. There is a sentence on page 77 of the book as follows: "the accuracy on the training set is 100%, because the leaves are pure". What is the meaning of this sentence.

W7.2. Is this an example of overfitting or underfitting or none, please explain.

W7.3. Please explain how the complexity of the model (decision tree) is reduced on cell **In[57].**

W7.4. How many leaves can you see in Figure 2.27 and how many of them are pure leaves?

W7.5. Can you explain the meaning of colors (orange and blue) in Figure 2.27?

W7.6. Train and display a decision tree model for the problem statement displayed on slide number 14 of the week 7 lecture slides. (buys computer). Upload your code and the visualization of the tree.

Show the output to your tutor when you are done.