CS7052-Machine Learning

Workshop 9: Model Evaluation and Cross-Validation

You will learn:

- To practice model evaluation and cross-validation with several datasets

Task 1

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

Follow the instructions on pages 257-292 in chapter 5.

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 5 (05-model-evaluation.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 following questions:

W9.1. Calculate cross validation score of KNN for Iris dataset. Hint: you can make changes to the code displayed on page 259. Upload your code and answers.

W9.2. Looking at cross validation scores of logistic regression for Iris dataset on page 259, compare them with cross validation scores of kNN for the same dataset.

W9.3. Looking at the heatmap of Fig 5-8, what is the lowest accuracy scores?

W9.4. Looking at the heatmap of Fig 5-8, what is the highest accuracy scores?

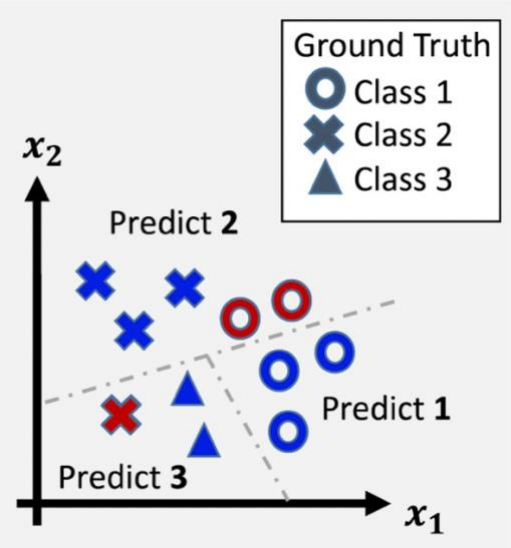
W9.5. Looking at the heatmap of Fig 5-8, what is the best value for parameter gamma?

W9.6. Looking at the heatmap of Fig 5-8, what is the best value for parameter C?

W9.7. Do you think the SVC associated with Fig 5-8 is sensitive to train-test split? Justify your answer.

Task 2

W9.8 Draw confusion matrix for this classification problem and calculate recall, precision and f-score for all classes.



Show the output to your tutor when you are done.