**Assignment University of Canberra  
Faculty of Science and Technology**

**Pattern Recognition and Machine Learning**

**Assignment 2 Part B**

**Group No.:**

**Project Team Members (maximum of 2):**

In Assignment 2 Part B of the assessment you (in your team) are asked to study the Support vector machine strategy of learning and its application to the given Cancer Detection Dataset to be covered in Tutorial 7 (Weeks 7-9) – the code, data and documentation is available in the Assignment 2 Part B documentation [here.](https://uclearn.canberra.edu.au/courses/14458/modules#module_185862)

The exercise is to study the code supplied for the Decision Tree model (DTM) for the breast cancer dataset and the SVM model for the given face recognition dataset and use the given SVM code for face recognition to develop the code for a SVM classifier for the breast cancer dataset (SVM-BC) as used in DTM. Your SVM-BC classifier needs to be tuned for model and hyperparameters to obtain optimal accuracy. Then you are required to spot-check the DTM and the SVM-BC to determine the more suited model for the breast cancer dataset based on accuracy analysis and other evaluation for the given breast cancer dataset. (An example that may be useful – an implementation of SVM on breast-cancer dataset and its analysis is given [here](https://towardsdatascience.com/breast-cancer-classification-using-support-vector-machine-svm-a510907d4878). Use the Tut7 dataset for this exercise – it is included in the sklearn.datasets package, breast\_cancer\_dataset.

You will generate a similar classification report to the one generated using DecisionTree for analysis. You may use Jupyter Notebook to explain the various sections and report the outcomes analysing and comparing the performances. You may also use Spyder to run the code also and develop a report responding to the key points as below separately.

Your report should cover the following (you may use the sample implementation as a reference at <https://towardsdatascience.com/breast-cancer-classification-using-support-vector-machine-svm-a510907d4878>).

1. **Decision Trees:**
2. How are decision trees used for regression tasks different from those used for classification?
3. How does a decision tree make predictions for new data points?
4. **SVMs:**
5. What is the criterion being optimized by SVMs?
6. How do SVMs handle the case of non-linearly separable data? What is the ‘kernel trick’, and why is it used in SVMs?
7. How can SVMs be extended to handle multi-class classification problems?
8. **Breast Cancer detection/prediction problem:**
9. Dataset description – breast-cancer dataset (refer to the sklearn dataset on breast-cancer)
10. What model parameters and hyperparameters are relevant to SVM-BC and used in the sklearn package SVM? How can they be tuned for optimisation?
11. What are relevant kernel functions and how are they used in SVM learners? Which one have you selected for your implementation of SVM through sklearn? How have you decided the best kernel?
12. What do the parameters C, decision\_function\_shape=’ovr’, degree, gamma and kernel mean in the following SVM classifier call?

Text, letter

Description automatically generated

1. Implement the SVM-BC classifier for the breast-cancer dataset.
2. Perform an accuracy analysis using a classification report for both the DecisionTree model and the SVM-BC model?
3. Which model is better? Why?

**Submission:**For Assignment 2 Part B you need to submit a report answering the above questions in 3-5 pages as a pdf file, or a ipynb file, or a py file or a combination of these but compiled as 1 zipped file and uploaded through canvas by the submission deadline.