***Assessed Coursework 1 Report***

***Support Vector Machines***

*Group 1*

*Minhao Li, Housu Zhang, Yijun Yang, Shiqi Xin, Chenqi Zhao*

***Introduction***

This report contains the findings when implemented four tasks for this lab. The datasets we chosen are Iris and Wine Quality: Iris for Classification and Wine Quality for Regression. Since this lab is focusing on binary classification, so the first category of Iris dataset is removed so that only two categories remain. And both dataset we used have high data quality and no missing data, and the data distribution is normal, so we don’t preprocess data there.

***Task 1: Train SVMs with the linear kernel***

Different values of the parameter ‘Epsilon’

Findings

***Task 2: Train SVMs with Gaussian RBF and Polynomial kernels***

Each model we trained

RBF:

Polynomial kernels

***Task3: Method Evaluation***

Linear:

Gaussian:

Polynomial kernels:

Classification rate

RMSE

***Task4: Additional Questions***

**Question 1:** What does the kernel parameter of the Gaussian RBF kernel signify (sigma)? What happens when you increase its value?

**Answer:**

**Question 2:** Explain what happens when a hard-margin SVM is fit to a dataset of two classes with overlapping features. What value do you need to set C (the slack-variable hyper-parameter) to attain a hard-margin SVM?

**Answer:**

**Question 3:** Explain why do you need to use both inner cross-validation and an outer k- fold cross-validation (known as nested cross-validation) in the machine learning process?

**Answer:**

***Conclusion***