CS5691 : Pattern Recognition and Machine Learning

Programming Assignment 2 **Deadline:** 28^{th} **February 2046**

Instructions:

1. You have to turn in the well-documented code along with a detailed report of the results of the experiments.
2. **Any sort of plagiarism/cheating will be dealt very strictly. Acknowledge any source used for performing the experiments.**
3. Plot your data and analyze before proceeding.
4. Be precise with your explanations. Avoid verbosity. Place relevant results that bolster your conclusions. Report should be within 15 pages in a single column format, and with 11pt font. Results/conclusions beyond the page limit will be ignored during evaluation.
5. You can use any programming language for this assignment. However, we recommend Python and MATLAB.
6. Create a folder named "TeamNumber_TeamMember1RollNo_TeamMember2RollNo" (for e.g. "1_CS17S016_CS17S011"). In this folder, you should have your report and a subfolder "codes" which should have all your codes. Upload this folder(.zip) on Moodle. Please follow the naming convention strictly.
7. Please make only one submission for the team. No emailed reports will be accepted.

1 Part A

1.1 Kernel Ridge Regression

Build a ridge regression model on "Regression_dataset" using the following kernels:

1. Linear kernel; and
2. Polynomial kernel.

(You can refer to Section 11.3.2 of [2] for a theoretical background on these topics.)

The report should include the following: (2+2+1 marks)

1. Table of classification accuracy on the train and test data for both the kernels.
2. Plot decision boundary plots in the original input space and transformed space (You can refer to Figure 4.12 of [1]).
3. Interpret your results. Which was the best kernel for each dataset? Why?

1.2 Kernel Logistic Regression and Kernel SVM

Build a logistic regression model and an SVM model on "Dataset 3" using the following kernels:

1. Linear kernel; and
2. Polynomial kernel.

(You can refer to Chapter 6 of [1] for a theoretical background on these topics.)
The report should include the following: (2+4+1 marks)

1. Table of classification accuracy on the train and test data for both the kernels on both the above datasets for both the models.
2. Plot decision boundary plots for both models as well as the support vectors for the SVM model.
3. Interpret your results. Which was the best kernel for each dataset? Why?

2 Part B

Implement a soft margin SVM-based classifier on "Dataset 2". Try for different values of hyper-parameter C and report the following on the best value of C: (2+1+4+2+1+4 marks)

1. Table of classification accuracy on the train and test data.
2. Confusion matrix on the test data.
3. Decision boundary plots along with margin and support vectors.
4. How many of the support vectors lie on the margin hyperplanes?
5. Interpret your results.
6. In the standard two-group classification, errors on positive or negative points are treated in the same manner. Suppose, however, that we wish to penalize an error on a negative point (false positive error) $k > 0$ times more than an error on a positive point, where k is an integer. Apply the modified SVMs to the classification task previously examined and compare with your previous SVMs results for $k = 2, 4, 8, 16$.

3 Part C

Implement Perceptron-based classifier on "Dataset 1" and "Dataset 3". Specify the parameters, for e.g., learning rate, initial weight, used in the implementation and note the number of iterations of Perceptron, for obtaining the results using the following kernels:

1. Linear kernel; and
2. Polynomial kernel.

(You can refer to Section 8.3.1 of [2] for a theoretical background on these topics.)
The report should include the following: (2+2+4+1+2 marks)

1. Table of classification accuracy on the train and test data for both the kernels on both the above datasets.
2. Decision boundary plots with superposed training data.
3. Compare the performance of Perceptron for both kernels across the two datasets and explain your observations. In particular, compare the convergence behaviour as a function of the number of iterations. Compare the number of iterations of Perceptron against the theoretical bound.
4. Interpret your results. Which was the best kernel for each dataset? Why?
5. Compare the results obtained from Perceptron for both kernels against that of hard margin SVMs across the two datasets.

References

[1] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, 1 edition, 2007.

[2] Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of Machine Learning*. MIT Press, Second Edition, 2018.