CS5691 : Pattern Recognition and Machine Learning

Programming Assignment 2 **Deadline: 16th April 2019, 23:55 hrs**

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 12 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.

This assignment is divided into two parts, Part A on Logistic Regression, Part B on Perceptron and Part C on SVM.

Datasets for each team can be found here. Use your smail account to access the link. Divide the dataset randomly for training(65%), and testing(35%).

1 Part A

Implement the following logistic regression models and apply it to different datasets assigned to your team.

1.1 Logistic Regression

You need to implement:

1. Two-class logistic regression on "Dataset_4".
2. Multi-class logistic regression on "Dataset_1".

(You can refer to Sections 4.3.2 and 4.3.4 of [1] for the theory.)
Specify the parameters used in the implementation of logistic regression. In particular, the initial weight, and the learning rate used. Also, report the number of iterations performed before collecting the results.
The report should include the following results for both models: (1+1+2+1 marks)

1. Table of classification accuracy on the train and test data.
2. Confusion matrix on the test data.
3. Decision boundary plots with superposed training data.
4. Interpret your results.

1.2 Regularized logistic regression

You need to build:

1. Two-class regularized logistic regression model on "Dataset_2".
2. Two-class regularized logistic regression model with feature mapping on "Dataset_2".

Theory for this experiment:

1. For regularized logistic regression you can add $\frac{\lambda}{2}\|\mathbf{w}\|^2$ term in the cost function of logistic regression.
2. Feature mapping: To fit the data better, you can create more features from each data point $x = [x_1, x_2]^T$. You can map the features into all polynomial terms of x_1 and x_2 up to six degree.

$$\text{mapfeatures(x)} = [1, x_1, x_2, x_1^2, x_2^2, x_1x_2, \dots]^T$$

As in the previous question, report all the parameter values used for the experiments.
The report should include the following for both models: (2+2+1+3+2 marks)

1. Try different regularization parameter (λ) between 0 and 1, and report the best $\lambda, say \lambda^*$.
2. Table of classification accuracy on the train and test data, for $\lambda = \{0, \lambda^*, 100\}$.
3. Confusion matrix on the test data, for $\lambda = \{0, \lambda^*, 100\}$.
4. Decision boundary plots with superposed training data, for $\lambda = \{0, \lambda^*, 100\}$. Provide the decision boundary in the original feature space, as well as the space of polynomial features.
5. Explain your findings and correlate the observed results with the available theory.

1.3 Kernel logistic regression

Build logistic regression model on "Dataset_2", "Dataset_3" and "Dataset_4" using the following kernels:

1. Linear kernel;
2. Polynomial kernel; and
3. Radial basis function kernel.

(You can refer to Chapter 6 of [1] for a theoretical background on these topics.)

The report should include the following: (3+5+2 marks)

1. Table of classification accuracy on the train and test data for all the three kernels on all the above three datasets.
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?

2 Part B

Implement Perceptron based classifier on "Dataset_2", "Dataset_4" and "Image_Dataset" assigned to your team, and specify the parameters, for e.g., learning rate, initial weight, used in the implementation. Note the number of iterations of Perceptron, for obtaining the results.

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

1. Table of classification accuracy on the train and test data.
2. Decision boundary plots with superposed training data.
3. Confusion matrix on the test data.
4. Compare the performance of Perceptron across the three datasets and explain your observations. In particular, compare the convergence behaviour as a function of the number of iterations. Discuss if Perceptron separated the classes? If not, are the classes linearly separable? Compare the number of iterations of Perceptron against the theoretical bound. (Refer to the convergence analysis in Section 1.3 of [2]).

3 Part C

Implement the following Support Vector Machine (SVM) models using your preferred library and apply it to different datasets assigned to your team.

3.1 Support Vector Machine

Implement SVM based classifier on "Dataset_4" and "Dataset_5". Try for different values of hyper-parameter C and report the following on the best value of C for both the datasets: (0.5+0.5+2+1 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. Interpret your results.

3.2 Kernel Support Vector Machine

Build SVM model on "Dataset_2", "Dataset_3" and "Dataset_4" using the following kernels:

1. Linear kernel;
2. Polynomial kernel; and
3. Radial basis function kernel.

(You can refer to Chapter 6 of [1] for a theoretical background on these topics.)

Try for different values of hyper-parameters. The report should include the following results for all models on best value of hyper-parameters: (1+3+2 marks)

1. Table of classification accuracy on the train and test data for all the three kernels on all the above three datasets.
2. Plot decision boundary along with margin and support vectors 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?

References

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

[2] Simon S. Haykin. *Neural networks and learning machines*. Pearson Education, Upper Saddle River, NJ, third edition, 2009.