ISTD 50.570 (Machine Learning): Assignment 2

March 7, 2016

Grading Policy and Due Date

- You are required to submit: 1) a report that summarizes your experimental results and findings, based on each of the following question asked; 2) your implementation (source code) of the algorithms.
- You are free to choose any programming language you prefer.
- Submit your assignment report and code to the Dropbox folder shared with you.
- This assignment is an individual assignment. Discussions amongst yourselves are allowed and encouraged, but you should write your own code and report.
- Submit your assignment to the Dropbox folder by 11:59 PM on Tuesday 22 March 2016. The last modified time of your most recently submitted file will be regarded as your submission time. This is a hard deadline. Late submissions will be heavily penalized (20% deduction per day).

Task 1: SVM and Logistic Regression

We would like to implement SVM using quadratic programming discussed in class. Please first download the optimizer by following this link: `http://tinyurl.com/50-570-optimizer`. You may need to obtain a free Academic License in order to use the optimizer. Please also download `a2-data.zip` which contains the data needed for this task. There are three datasets (data A, data B and data C). Each dataset consists of a training and a test set. You are required to report results on the three datasets separately.

1. (10 pts) Implement a simple linear SVM that deals with the case where the data is slightly not linearly separable (*i.e.*, there are slack variables) in the primal form and in the dual form. Write a prediction function that can be used to predict an input point. Run your code on the training data to build the model, with $C = 1.0$ (note $C = 1/\lambda$), and evaluate your model's performance on the training and test data. Report your results and findings clearly using tables.

   Hint: You should be able to verify the correctness of your implementations by comparing the results obtained from the two implementations.

2. (10 pts) Discuss the behavior of the performance on the training and test set as we change the value of C (or $1/\lambda$; see the notes). What happens to the margin as C increases? What happens to the number of support vectors as C increases? Use graphs and tables to show such results clearly.

3. (10 pts) Typically, a different value of C corresponds to a different model for classification. What is a reasonable way to select C? Do you believe we can optimize the value of C by maximizing the margin on the training set? If so, give your argument and provide empirical support for your argument. If not, explain why not, and propose and implement a different method to select a good value of C so as to obtain a good performance on the test set. Report your performance on the training and test set with the C selected by your method.

4. (10 pts) Extend your implementation to support kernels. (Please note that only minimal modifications will be required.) Try to use the Gaussian (RBF) kernel and another kernel of your choice and perform experiments on the three data sets. Use tables to clearly present your results.

5. (10 pts) Now, implement the regularized logistic regression using stochastic gradient descent. Perform experiments using the same data used in the previous questions and compare the performance with that of SVM in the previous questions. Try setting different values for λ (*e.g.*, 0.001, 0.01, 0.1, 1, 10, 100, 1000) and report the different results obtained. Also report the learning rate you used and the final objective value obtained when the algorithm converges.