Statistical Learning (MT7049) - Project 3

Instructions: This project consists of 2 tasks that should be solved individually. You are free to use any programming package in this project.

The solution should be submitted at the course webpage in a single .pdf file with your source code attached as appendices. Your source code should include clear comments and documentations to describe what you are evaluating.

TASK 1

This task allows you to see the connection between the SVM optimization problem in terms of the slack variables (given in Eq. 12.8 in the course book), and the SVM optimization problem as a penalization method (given in Eq. 12.25 in the course book). In your answer, please write concisely and to-the-point.

a) Show that the optimization problems in Eq. 12.8 and Eq. 12.25 are the same. Some take-home messages from this exercise: I) Eq.12.25 does not involve the slack variables as in Eq. 12.8, so complicated optimization methods (i.e., nonlinear programming described in Sec 12.2.1) can be avoided. In fact, simple gradient descent methods can be used to minimize Eq. 12.25; II) Although we are working on classification problem, the formulation of Eq. 12.25 closely resembles that of ridge regression. Therefore, concepts that we learned before in regularization and shrinkage also apply to SVM.

b) Referring to the Table 12.1 in the course book, derive the minimizing functions for the "binomial deviance" and the "SVM hinge loss", i.e., the first two rows of the table.

c) Consider the Gaussian kernel in Eq. 12.22, discuss how the change of γ could affect the bias-variance tradeoff and suggest a way to choose γ.

TASK 2

This task allows you to experience and explore the use of SVM to perform binary classification with kernel trick. This task uses the "Spam" data in the course book as in Project 2. To download the data, go to the webpage of the course book (https://web.stanford.edu/~hastie/ElemStatLearn/), and then click "Data" on the left panel to find the data and follow the instruction to use them.

a) Pick up the correct function/package in R, Python, etc., to perform binary classification using SVM with kernel trick. Describe which function and package are picked. Explain which kernel is used and why it is chosen for the task. List and explain the choices of all other parameters and settings required by the function to train the SVM.

b) Write a code to construct the SVM with the chosen kernel. Attach your source code at the end of the report.

c) Construct the cross-validation (CV) plot, i.e., CV error versus λ or C (see Eq. 12.25 in course book) with 10-fold CV. Standard error bars should be included in the plot so that you can use the one-standard-error rule to identify the optimal model.

d) Based on part c) above and part d) in the Task 2 of Project 2, compare the performances of gradient boosting trees and SVM with kernel trick in classifying the Spam data. Reason why one method performs better than another.
