School of Computer Science
The University of Adelaide

Introduction to Statistical Machine Learning
Assignment 1

Semester 2 2015
Due 11:55pm, 31 August 2015

Instructions and submission guidelines:

- The assignment consists of a report and a matlab implementation of binary class linear Support Vector Machines (SVMs).
- Explain the key points in the report.
- Make sure that your writing is legible and clear, and the mathematical symbols are consistent.
- Make sure that your Matlab scripts are well commented and can be executed directly.
- You must sign an assessment declaration coversheet to submit with your assignment. The assessment declaration coversheet is included in the zip file.
- Submit your report and all your Matlab code via the MoodleForum on the course web page.

1 Reading

We have briefly covered soft margin binary SVMs in Lecture 2. Please read the SVM tutorial [1] and the guide [3] for more details to complete the assignment.

2 Report

Please write down your understanding of binary class linear SVMs (within 4 pages) and experiment comparison of your code and an existing implementation of SVMs, libsvm [2] (within 4 pages) in the report. So in total, you have at most 8 pages for the report. This is no strict format of the report (rather than the page limit). The purpose of the report is to show me what you have understood about SVMs and what you have done to (at least try to) make your code correct. How to do this is up to you.

The report should at least cover the following key points (not limited to)

- The primal form and its dual form for both hard margin and soft margin case;
- Concept of support vectors;
- Why max margin is good;
- Concepts of generalisation error, and generalisation bounds;
- Concepts of duality gap, weak duality and strong duality;
- Experimental comparison with libsvm.

3 Code

- Please implement **soft margin** binary class linear SVMs by
  - solving the **primal** problem.
  - solving the **dual** problem.
- Please use the matlab optimisation tool cvx `http://cvxr.com/cvx/` to solve the above two optimisation problems (the primal and dual)
- You can use any binary classification datasets on libsvm.
- Self-check the correctness of your code. Please compare it with libsvm[2].

Note that the formulations in the libsvm and yours might be slightly different, thus you will need to figure out the difference and correspondence.

4 Marking criteria

Total score has 100 points with the following breakdown:

1. Define variables before you use them [5 points]
2. Primal and dual forms of hard and soft margin SVMs [20 points]
3. Concept of support vectors (two types) [10 points]
4. Why max margin is good [5 points]
5. Generalisation error and bounds [10 points]
6. Duality gap, weak and strong duality [10 points]
7. Experiment
   (a) compare your $\mathbf{w}, b$ obtained via solving the primal problem, with the $\mathbf{w}, b$ reconstructed by the dual variables obtained via solving the dual problem (both the results and the reconstruction formulation) [5 points]
   (b) check duality gap of yours (both the result and the formulation) [5 points]
   (c) compare your $\mathbf{w}, b, \alpha$ with those of libsvm [5 points]
   (d) compare training and testing errors of your code and libsvm [5 points]
   (e) code [20 points]

Please note that responses/answers to all above checkpoints (except the code itself) should be included in the report (not in the code).

$\sim\sim\sim$ Good luck $\sim\sim\sim$
by Javen Qinfeng Shi, 2015

References

[1] Christopher J.C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2:121–167, 1998.

[2] C.C. Chang and C.J. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[3] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin. *A Practical Guide to Support Vector Classification*, 2010.