Assignment 2: Support Vector Machines

Started: 14 Aug at 20:05

Quiz instructions

Please read the following instruction carefully before you complete assignment 2

Instructions and submission guidelines:

- This is not a group assignment. Everyone is asked to complete the assignment individually.
- The assignment consists of a mix set of tasks, including derivation, question answering, coding and result analysis. You need to provide your solutions to the corresponding sections in this quiz.
- You need to also upload your code. We will check the code to ensure that your results can be generated from your code. If your result does not match with your code, you will get 0 mark for the relevant sections.
- You should use Python to finish this assignment. Juptyer notebook is allowed and encouraged.

Data:

Please download the data that will be used in this assignment here: **train.csv** (https://myuni.adelaide.edu.au/courses/75049/files/11310716/download?download_frd=1) **test.csv** (https://myuni.adelaide.edu.au/courses/75049/files/11310715/download?download_frd=1)

There are a "train.csv" file and a "test.csv" file. They are the training data and testing data, respectively. Inside the CSV file, the first column is the class label, the remaining columns are the features.

The label (y) can take one of two classes - 0, 1;
In total, there are 200 features (X);
Trainset consists of 8500 samples, testset has 1500;

Please use **the first 4000 samples** in the "train.csv" file as the training set and the remaining samples as validation set. All the samples in the "test.csv" file will be used as test samples.

Third-party Libraries,

You can use any third-party libraries to read and process data.

You can use numpy in any question.

Please follow the instruction of each question to complete the relevant parts.

Question 1 — 10 pts

Please write down the primal form of the hard-margin SVM and derive its dual formation. You are required to upload a PDF file (preferably in one page) to show your derivation of the dual formulation.

Upload Choose a file

Question 2 — 25 pts

Please implement the training and testing algorithms of soft-margin Linear Support Vector Machine from its primal form, that is,

$$\min_{\mathbf{w},b,\{\xi_i\}} \frac{1}{2}\|\mathbf{w}\|_2^2 + \frac{C}{N}\sum_{i=1}^{N}\xi_i$$
$$s.t.\ \ y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - \xi_i,\ \ \forall i$$
$$\xi_i \geq 0$$

by using CVX. Your implementation should follow the following I/O format:

svm_model = svm_train_primal (data_train , label_train , regularisation_para_C)

test_accuracy = svm_predict_primal (data_test , label_test , svm_model)

Please copy the code snippet for the implementation of those two functions (You can also do this by attaching the screenshot of your code.).

By setting C = 100, run your implementation, please report the solution of b and sum of all dimensions of $\mathbf{w}$ solution, e.g., np.sum(w). (For a quick check of the correctness of your code)

Edit View Insert Format Tools Table

12pt Paragraph B I U A T²

p 0 words </>

Question 3 20 pts

Please implement the training algorithm of the soft-margin Linear Support Vector Machine from its dual form, that is,

$$\max_{\alpha_i} \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j < \mathbf{x}_i, \mathbf{x}_j >$$
$$s.t. \quad 0 \leq \alpha_i \leq \frac{C}{N}$$
$$\sum_i \alpha_i y_i = 0$$

by using CVX. Your implementation should follow the following I/O format:

svm_model = svm_train_dual (data_train , label_train , regularisation_para_C)

Please copy the code snippet for the implementation of this function (You can also do this by attaching the screenshot of your code.).

By setting C = 100, run your implementation, please report the sum of all dimensions of the optimal α, e.g., np.sum(alpha). (For a quick check of the correctness of your code)

Edit View Insert Format Tools Table

12pt Paragraph

p 0 words

Question 4

10 pts

Write code to obtain the primal problem solution $\mathbf{w}^*, b^*$ from its dual solution α^*. Please copy the code snippet for the implementation (You can also do this by attaching the screenshot of your code.).

Edit View Insert Format Tools Table

12pt Paragraph B I U A T²

p 0 words </>

Question 5

5 pts

Write code to find the support vectors from the primal problem solutions. Please attach the code snippet for the implementation.

Edit View Insert Format Tools Table

12pt Paragraph B I U A T²

p 0 words </>

Question 6

5 pts

Write code to find the support vectors from the dual problem solutions. Please copy the code snippet for the implementation

Edit View Insert Format Tools Table

12pt Paragraph B I U A T²

p 0 words </>

Question 7

10 pts

Write code to choose C by using the validation set. Please copy the code snippet for the implementation. Report the test accuracy you get by using the optimal C found in the validation set.

Hint: you can search C from the range $\{2^{-10}, 2^{-8}, 2^{-6}, \cdots, 2^{10}\}$

Question 8

15 pts

Please study one of the following packages and perform classification with linear SVM (with optimal C searched in the validation set) on the assignment dataset

1. Libsvm

https://www.csie.ntu.edu.tw/~cjlin/libsvm/ (https://www.csie.ntu.edu.tw/~cjlin/libsvm/)

2. Scikit-learn SVM

https://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html#sklearn.svm (https://scikit-

learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html#sklearn.svm.LinearS

Please copy your code snippet and report your test accuracy.

Question 9

0 pts

Please upload the entire code of your code.

Upload Choose a file

Saved at 8:23 Submit quiz