CS480/680: Introduction to Machine Learning

Homework 4

Due: 11:59 pm, June 8, 2020, submit on LEARN.

Include your name and student number!

Submit your writeup in pdf and all source code in a zip file (with proper documentation). Write a script for each programming exercise so that the TAs can easily run and verify your results. Make sure your code runs!
[Text in square brackets are hints that can be ignored.]

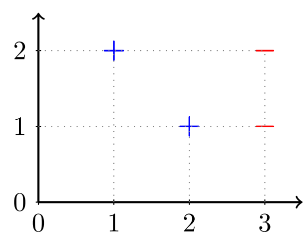
Exercise 1: Squared Hinge SVM (10 pts)

Let us consider the soft-margin SVM with the *squared* hinge loss:

$$\min_{\mathbf{w}\in\mathbb{R}^d, b\in\mathbb{R}} \frac{1}{2}\|\mathbf{w}\|_2^2 + C\sum_{i=1}^n \max\{1 - y_i(\mathbf{w}^\top \mathbf{x}_i + b), 0\}^2, \tag{1}$$

where $\mathbf{x}_i \in \mathbb{R}^d$, $y_i \in \{\pm 1\}$, and $\|\mathbf{w}\|_2 := \sqrt{\sum_{j=1}^d w_j^2}$ is the Euclidean norm.

1. (1 pt) Find the hard-margin SVM optimal solution $(\mathbf{w}^*, b^*)$ for the following dataset (where coordinates of all positive and negative examples are given):



   No need to prove the optimality of your $(\mathbf{w}^*, b^*)$, but please include and explain all intermediate steps.

2. (2 pts) Derive the Lagrangian dual of the squared hinge SVM (1). Please include intermediate steps so that you can get partial credits.

   In the following you will complete and implement the following gradient algorithm for solving the squared hinge SVM in Equation (1):

   Algorithm 1: SGD for SVM.

   Input: $X \in \mathbb{R}^{n\times d}$, $\mathbf{y} \in \{-1, 1\}^n$, $\mathbf{w} = \mathbf{0}_d$, $b = 0$, $\mathsf{max\_pass} \in \mathbb{N}$, step size η

   Output: $\mathbf{w}, b$

   for $t = 1, 2, \ldots, \mathsf{max\_pass}$ **do**
      for $i = 1, 2, \ldots, n$ **do**
         choose step size η
         if $y_i(\langle \mathbf{x}_i, \mathbf{w}\rangle + b) \leq 1$ **then**
            $\mathbf{w} \leftarrow$ // $\mathbf{x}_i$ is the i-th row of X
            $b \leftarrow$
         $\mathbf{w} \leftarrow$ // the proximal step to maximize margin

3. (2 pts) Compute the gradient w.r.t. $\mathbf{w}$ and b for each second term in Equation (1):

$$C(1 - y_i\hat{y}_i)_+^2, \quad \text{where} \quad \hat{y}_i = \mathbf{w}^\top \mathbf{x}_i + b. \tag{2}$$

4. (1 pt) Find the closed-form solution of the following proximal step:

$$\mathsf{P}^{\eta}(\mathbf{w}) = \operatorname*{argmin}_{\mathbf{z}} \ \frac{1}{2\eta}\|\mathbf{z} - \mathbf{w}\|_2^2 + \frac{1}{2}\|\mathbf{z}\|_2^2 \tag{3}$$

5. (3 pts) Implement Algorithm 1. You should use Ex 1.3 to complete lines 5-6, and Ex 1.4 for line 7. [FYI: you may want to contrast Algorithm 1 with the perceptron algorithm in Lecture 01.]

6. (1 pt) Run your implementation on the toy dataset in Ex 1.1. Can you use a large C to recover the hard-margin SVM solution you derived in Ex 1.1? Report your C value and the solution $\mathbf{w}, b$ found by your algorithm.