CS480/680: Introduction to Machine Learning

Homework 1

Due: 11:59 pm, May 18, 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: Perceptron Implementation (5 pts)

Convention: All algebraic operations, when applied to a vector or matrix, are understood to be element-wise (unless otherwise stated).

Algorithm 1: The perceptron algorithm.

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

```
for t = 1, 2, ..., max_pass do
    mistake(t) ← 0
    for i = 1, 2, ..., n do
        if y_i(⟨x_i, w⟩ + b) ≤ 0 then
            w ← w + y_i x_i                      // x_i is the i-th row of X
            b ← b + y_i
            mistake(t) ← mistake(t) + 1
```

Implement the perceptron in Algorithm 1. Your implementation should take input as $X = [\mathbf{x}_1^\top, \ldots, \mathbf{x}_n^\top]^\top \in \mathbb{R}^{n\times d}$, $\mathbf{y} \in \{-1, 1\}^n$, an initialization of the hyperplane parameters $\mathbf{w} \in \mathbb{R}^d$ and $b \in \mathbb{R}$, and the maximum number of passes of the training set [suggested $\mathsf{max\_pass} = 500$]. Run your perceptron algorithm on the spambase dataset (available on course website), and plot the number of mistakes (y-axis) w.r.t. the number of passes (x-axis).

Ans:

Exercise 2: Linear Regression (5 pts)

Recall that ridge regression refers to the following ℓ_2 norm regularized linear regression problem:

$$\min_{\mathbf{w}} \ \tfrac{1}{2}\|X\mathbf{w} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{w}\|_2^2, \tag{1}$$

where $X \in \mathbb{R}^{n\times d}, \mathbf{y} \in \mathbb{R}^n$ and $\mathbf{w} \in \mathbb{R}^d$. Obviously, setting $\lambda = 0$ we recover ordinary linear regression.

1. (3 pts) Prove that ridge regression with any $\lambda > 0$ is equivalent to ordinary linear regression after performing the following data augmentation:

$$X \leftarrow \begin{bmatrix} X \\ \sqrt{2\lambda} I_{d\times d} \end{bmatrix} \tag{2}$$

$$\mathbf{y} \leftarrow \begin{bmatrix} \mathbf{y} \\ 0_d \end{bmatrix}, \tag{3}$$

where $I_{d\times d}$ is the $d \times d$ identity matrix and 0_d is the d-dimensional zero vector.

Ans:

2. (2 pts) Explain the data augmentation step (2)-(3). [E.g., what kind of data are we adding to the training set? how many of them? their dimension? what effect are they bringing to the weight vector $\mathbf{w}$?]

Ans: