CAP 6610 Machine Learning, Spring 2020

Homework 2

Due 2/17/2020 11:59PM

Helpful reading: Boyd & Vandenberghe, *Convex Optimization.*

1. (10 points) What is the distance between two parallel hyperplanes $\{x \in \mathbb{R}^n \mid a^\top x = b_1\}$ and $\{x \in \mathbb{R}^n \mid a^\top x = b_2\}$? *Hint.* Let $a^\top x_1 = b_1$, $a^\top x_2 = b_2$, and minimize $\|x_1 - x_2\|^2$.

2. (20 points) Let x be a real-valued random variable with sample space $\{a_1, \dots, a_k\}$ where $a_1 \leq a_2 \leq \cdots \leq a_k$. This can be view as a categorical random variable with each catogory assigned a real value. Let $\Pr[x = a_i] = p_i$, then the vector p satisfies $p \geq 0$ and $\mathbf{1}^\top p = 1$, i.e., it lies in the probability simplex Δ. For each of the following functions of p on the probability simplex, determine if the function is convex, concave, or neither.

    (a) $\mathrm{E}[x]$

    (b) $\Pr[x > \alpha]$

    (c) $\Pr[\alpha < x < \beta]$

    (d) $-\sum_{i=1}^{k} p_i \log p_i$, the entropy of this distribution

    (e) $\mathrm{var}(x)$

3. (10 points) Show that the following two convex problems are equivalent. Carefully explain how the solution of (b) is obtained from the solution of (a).

    (a) The robust least squares problem

    $$\underset{\boldsymbol{\theta}}{\text{minimize}} \quad \sum_{i=1}^{n} h(\boldsymbol{\phi}_i^\top \boldsymbol{\theta} - \psi_i),$$

    where $h : \mathbb{R} \to \mathbb{R}$ is the Huber function defined (with a constant M) as

    $$h(t) = \begin{cases} t^2 & |t| \leq M \\ M(2|t| - M) & |t| > M. \end{cases}$$

    (b) The quadratic program

    $$\underset{\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{v}}{\text{minimize}} \quad \sum_{i=1}^{n} (u_i^2 + 2M v_i)$$

    $$\text{subject to} \quad -\boldsymbol{u} - \boldsymbol{v} \leq \boldsymbol{\Phi}\boldsymbol{\theta} - \boldsymbol{\psi} \leq \boldsymbol{u} + \boldsymbol{v}$$

    $$0 \leq \boldsymbol{u} \leq M\mathbf{1}, \qquad \boldsymbol{v} \geq 0.$$

4. (30 points) We test the performance of three regression methods on the wine data set `http://archive.ics.uci.edu/ml/datasets/Wine+Quality`. We will only consider the red wine data set, with 1599 samples. We use the first 1400 samples for training, and the last 199 samples for testing. The goal is to build a linear model of the first 11 features (together with a constant term) to predict the quality of the wine. All models are trained by solving the following optimization problem

$$\underset{\boldsymbol{w},\beta}{\text{minimize}} \quad \sum_{i=1}^{n} \ell(\boldsymbol{x}_i^\top \boldsymbol{w} + \beta - y_i),$$

where the loss functions are

- least squares loss $\ell(t) = t^2$
- Huber loss defined in the previous problem, with $M = 1$
- hinge (deadzone-linear) loss

$$\ell(t) = \begin{cases} 0 & |t| \le 0.5 \\ |t| - 0.5 & |t| > 0.5 \end{cases}$$

The least squares loss can be directly solved by the command `Phi\y` for some properly defined `Phi`. For the latter two, you will use the `cvx` package found on Prof. Boyd's website `https://web.stanford.edu/~boyd/software.html`. Report their prediction performance on the test set using a different metric, mean absolution error (MAE), defined as $(1/n) \sum_{i=1}^{n} |y_i - \hat{y}_i|$.

5. (30 points) We test the performance of three classification methods on the ionosphere data set `https://archive.ics.uci.edu/ml/datasets/ionosphere`. There are 351 samples. We use the first 300 samples for training, and the last 51 samples for testing. The goal is to build a linear model of the 34 features (together with a constant term) to predict the binary (± 1) outcome. All models are trained by solving the following optimization problem

$$\underset{\boldsymbol{w},\beta}{\text{minimize}} \quad \sum_{i=1}^{n} \ell(\boldsymbol{x}_i^\top \boldsymbol{w} + \beta, y_i),$$

where the loss functions are

- least squares loss $\ell(t, y) = (yt - 1)^2$
- logistic loss $\ell(t, y) = \log(1 + \exp(-yt))$
- hinge loss $\ell(t, y) = \max(0, 1 - yt)$

Again, you will use the backslash command to solve for the first model, and `cvx` to solve for the latter two. Report their prediction accuracy on the test set.