CSE 5819 Assignment #2

by: Aayushi Verma (uef24001)

This assignment is organized as follows:

1 Part 1: ChatGPT Self-Learning (20pts)

You can start from the following prompts, but you need to create subsequent questions that attempt to understand the basic concepts of machine learning.

2 Problems

[Convex Optimization, Mean/Medium/Mode, Preparation for Learning Loss Functions]

Take the training data set $D = \{(x_i, y_i): 1 \le i \le n\}$. Now consider the following hypothesis class (i.e., the set of constant functions)

$$H = \{h_z(x) = z, \forall\, x \in R\}$$

We will choose a model from this hypothesis class in our machine learning task.

2.1 Question One

1. [20 pts] Let us use L_2 vector norm to measure the discrepancy between the observed y and the model output $h_z(x)$. In other words, the squared loss is defined as

$$L_{sq}(h) = \frac{1}{n}\sum_{i=1}^{n}(h(x_i) - y_i)^2$$

Prove that if we set the constant z equal to the mean of y

$$m = \frac{1}{n}\sum_{i=1}^{n} y_i$$

this constant function $h_m(\cdot)$ minimizes the squared loss, and in other words,

$$h_m = \arg\min_{\mathrm{h}\in\mathrm{H}} L_{sq}(h)$$

2.2 Question Two

2. [20 pts] Let us use the L_1 vector norm to measure the discrepancy between the observed y and the model output $h_z(x)$. In other words, the absolute loss is defined as

$$L_{abs}(h) = \frac{1}{n}\sum_{i=1}^{n}|h(x_i) - y_i|$$

Prove that if we set the constant z equal to the median of y

$$m = median(\{y_i, \quad 1 \le i \le n\})$$

this constant function $h_m(\cdot)$ minimizes the absolute loss, and in other words,

$$h_m = \arg\min_{\mathrm{h}\in\mathrm{H}} L_{abs}(h)$$

2.3 Question Three

3. [20 pts] Let us use the L_0 vector norm to measure the discrepancy between the observed y and the model output $h_z(x)$. In other words, the binary loss is defined as

$$L_{bin}(h) = \frac{1}{n}\sum_{i=1}^{n} 1\{h(x_i) \neq y_i\}$$

where $1\{h(x_i) \neq y_i\}$ returns 1 if $h(x_i) \neq y_i$; or otherwise return 0. In other words, this loss counts how many times the model output $h(x_i)$ is not equal to the observed y_i. Then normalize the sum by n .

Prove that if we set the constant z equal to the mode of y

$$m = mode(\{y_i, \quad 1 \leq i \leq n\})$$

where the mode is the most common number in the observed y's,

this constant function $h_m(\cdot)$ minimizes the binary loss, and in other words,

$$h_m = \arg\min_{\mathrm{h} \in \mathrm{H}} L_{bin}(h)$$