CSE 5819 Assignment #4

by: Aayushi Verma (uef24001)

This assignment is organized as follows:

1. ChatGPT - pg 1–4
2. Problems - pg 5–7
3. Coding - attached as a separate .ipynb file

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.

1.1 What is overfitting? What is underfitting?

1.2 How does logistic regression handle overfitting?

1.3 How does least squares handle overfitting?

1.4 What regularizers can be used to regularize machine learning algorithms?

1.5 How to compute gradient of the least squares loss?

1.6 How to compute gradient of the logistic loss?

2 Problems

2.1 Question One

[Gradient Descent]

[20pts] You have a univariate function you wish to minimize, $f(w) = 5(w - 11)^4$. Suppose you wish to perform gradient descent with constant step size $\alpha = 1/40$. Starting with $w_0 = 13$, perform 2 steps of gradient descent. What are $w_0, ..., w_2$?

2.2 Question Two

[Logistic Regression]

In this problem, we are going to assume the same notation setup in class. For logistic regression, we model the class probability by

$$P(y|\mathbf{x}_i) = \sigma(y(\mathbf{w}^T\mathbf{x}_i))$$

where we define the *logistic function* $\sigma : \mathbb{R} \to \mathbb{R}$ as

$$\sigma(s) = \frac{1}{1 + e^{-s}}$$

(Note: We dropped the bias term b since we can always absorb the bias into $\mathbf{w}$ by representing $\mathbf{x}$ in homogeneous coordinates.)

1. [20 pts] Prove that the given σ is a properly defined probabilistic model. In other words, prove that $P(y = +1|\mathrm{x_i}) + P(y = -1|\mathrm{x_i}) = 1$

   (Hint: use the relationship between $\sigma(s)$ and $\sigma(-s)$.)

2.3 Question Three

2. [20 pts] Show that the gradient of the log likelihood function, namely, $\nabla_w \log P(\mathbf{y}|X, \mathbf{w})$ is

$$\sum_{i=1}^{n}(1 - \sigma(y_i(\mathbf{w}^T\mathbf{x}_i)))y_i\mathbf{x}_i$$

where y_i is a scalar, and x_i is a vector of descriptors, and $y = [y_1, y_2, \cdots, y_n]^T$ is a vector, and $X = [x_1, x_2, \cdots, x_n]$ is a matrix indicating all data.