ELEC 378 – Spring 2023
Homework 8

Due: Friday March 24, 5PM

1 Multiple Logistic Regression on Paper

Let $\sigma(z) = \frac{1}{1+e^{-z}}$ denote the sigmoid function. A multiple logistic regression model with parameters $\mathbf{w} \in \mathbb{R}^p$ and $b \in \mathbb{R}$ predicts the label $y_i \in \{0, 1\}$ of a data point $\mathbf{x}_i \in \mathbb{R}^p$ by computing $\hat{c}(\mathbf{x}_i) = \sigma(\langle \mathbf{x}_i, \mathbf{w} \rangle + b)$ and rounding to the nearest integer

$$\hat{y}_i = \begin{cases} 1, & \hat{c}(\mathbf{x}_i) \geq 1/2 \\ 0, & \hat{c}(\mathbf{x}_i) < 1/2, \end{cases}$$

where $\mathbf{w}$ and b are found by minimizing the objective function

$$L(\mathbf{w}, b) = \sum_{i=1}^{n} -y_i \log(\hat{c}(\mathbf{x}_i)) - (1 - y_i) \log(1 - \hat{c}(\mathbf{x}_i)).$$

a) Prove that

$$\hat{y}_i = \begin{cases} 1, & \langle \mathbf{w}, \mathbf{x}_i \rangle \geq -b \\ 0, & \langle \mathbf{w}, \mathbf{x}_i \rangle < -b, \end{cases}$$

i.e., that logistic regression can be thought of as thresholded linear regression.

b) Prove that $\frac{d}{dz}\sigma(z) = \sigma(z)(1 - \sigma(z))$.

c) Prove that $\nabla_{\mathbf{w}} \hat{c}(\mathbf{x}_i) = \hat{c}(\mathbf{x}_i)(1 - \hat{c}(\mathbf{x}_i))\mathbf{x}_i$.

d) Prove that $\nabla_b \hat{c}(\mathbf{x}_i) = \hat{c}(\mathbf{x}_i)(1 - \hat{c}(\mathbf{x}_i))$.

e) Derive a gradient step towards optimal $\mathbf{w}^*$, that is, find $\nabla_{\mathbf{w}} L(\mathbf{w}, b)$ in the gradient descent algorithm $\mathbf{w}^{t+1} = \mathbf{w}^t + \mu^t \nabla_{\mathbf{w}} L(\mathbf{w}, b)$.

f) Derive a gradient step towards optimal b^*, that is, find $\nabla_b L(\mathbf{w}, b)$ in the gradient descent algorithm $b^{t+1} = b^t + \mu^t \nabla_b L(\mathbf{w}, b)$.

2 Multiple Logistic Regression in Practice

Construct a labeled dataset containing all samples from the MNIST handwritten digits corresponding to eights and nines. Fit an `sklearn` logistic regression model on the first 288 samples, and use this model to predict the labels of the remaining samples. For each incorrectly predicted sample, display the probabilities predicted by the model, as well as the image itself. How does the model perform?

3 Linear Prediction: Least Squares vs LMS

a) Show that the Wiener filter acquired via least squares produces more accurate predictions than those produced by its LMS algorithm estimate using the provided SIVB stock price data.

b) When is it desirable or necessary to use the LMS algorithm to estimate the Wiener filter via stochastic gradient descent instead of computing it directly via least squares? Describe such a scenario in terms of stock price data.

Submission Instructions

Every student must submit their work in PDF format, providing intermediate and final results as well as any necessary code. Submit your homework on Gradescope.

Collaboration Policy

Collaboration both inside and outside class is encouraged. You may talk to other students for general ideas and concepts, but individual write-ups must be done independently.

Plagiarism

Plagiarism of any form will not be tolerated. You are expected to credit all sources explicitly.