CSCI 57800 ML Fall 2023 Homework 2

October 4, 2023

Instructions

We will be using Canvas to collect your assignments. Please read the following instructions to prepare your submission.

1. Submit your solution in a pdf file and a zip file (<yourLastName_FirstName>.pdf/zip). Your write-up must be in pdf. Your code must be in the zip file.
2. In your pdf file, the solution to each problem should start on a new page.
3. Latex is strongly encouraged to write your solutions, e.g., using Overleaf (https://www.overleaf.com/). Neither scanned handwritten copies nor hard copies are acceptable.
4. You need to add screen captures of your code and the output in your write-up.
5. You may discuss the problems and potential directions for solving them with another student. However, you need to write your own solutions and code separately, and not as a group activity. Please list the students you collaborated with on your submission.

Problem 1 (12 points)

Consider the matrix A and the vectors x and y below.

$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} 7 \\ 8 \end{bmatrix}$

Answer the following questions. Show all your work.

(a) (3 pts) What is the inner product of the vectors $\mathbf{x}$ and $\mathbf{y}$?

(b) (3 pts) What is the product of $\mathbf{Ax}$?

(c) (3 pts) Is $\mathbf{A}$ invertible? If so, give the inverse, and if not, explain why not.

(d) (3 pts) What is the rank of $\mathbf{A}$?

Problem 2 (25 points)

Answer the following questions. Show all your work.

(a) (5 pts) If $y = x^5 + x^2 + 10$, then what is the derivative of y with respect to x?

(b) (5 pts) What is the derivative of $f(x) = \exp(-\frac{3}{5}x^2)$?

(c) (5 pts) What is the derivative of $f(x) = \log(\frac{1}{2}x^2 - x + 1)$?

(d) (5 pts) If $f(x_1, x_2) = x_2 sin(x_1) e^{-\frac{1}{2}x_2}$, what is the gradient $\nabla f(x)$ of f?
Recall that $\nabla f(x) = \begin{pmatrix} \partial_{x_1} f \\ \partial_{x_2} f \end{pmatrix}$.

(e) (5 pts) Compute the partial derivative $\partial_b J$, where $J(m, b) = \sum_{i=1}^{n}((mx_i + b) - y_i)^2$.

Problem 3 (21 points)

Here are training samples.

x_1	x_2	y
2	1	4
3	5	5
4	6	7
9	10	9

We are going to learn a linear regression model using the gradient descent method. The objective function is $J(\theta) = \frac{1}{m}\sum_{i=1}^{m}(y - \hat{y})^2$ when m is the number of training samples.

Assuming that learning rate is $\alpha = 0.1$ and parameter θ has been initialized to $[0, 0, 0]^T$. Note that the first element is a bias, θ_0. We will perform one iteration of gradient descent.

(a) (8 pts) What is the gradient of the objective function $J(\theta)$ with respect to θ: $\nabla_\theta J(\theta)$ in the first iteration? Show all your work.

(b) (8 pts) What is θ after one iteration of gradient descent? Show all your work.

(c) (5 pts) How could we pick which value of α to use if we weren't given the learning rate?

Problem 4 (8 points)

In Logistic Regression, the conditional probability of $y^{(i)}$ given $\mathbf{x}^{(i)}$ is

$$p(y^{(i)}|\mathbf{x}^{(i)},\theta) = \begin{cases} \sigma(\theta^T\mathbf{x}^{(i)}) & y^{(i)} = 1 \\ 1 - \sigma(\theta^T\mathbf{x}^{(i)}) & y^{(i)} = 0 \end{cases} \tag{1}$$

We can rewrite this as a single statement below.

$$p(y^{(i)}|\mathbf{x}^{(i)},\theta) = \sigma(\theta^T\mathbf{x}^{(i)})^{y^{(i)}}(1 - \sigma(\theta^T\mathbf{x}^{(i)}))^{(1-y^{(i)})} \tag{2}$$

Can you show why this single statement is true?

Problem 5 (34 points, Programming Involved)

Consider the following function f. We will write a program to numerically minimize this function.

$$f(w_1, w_2) = (x - w_1)^2 + y(w_2 - w_1^2)^2 \tag{3}$$

(a) (8 pts) Derive the partial derivatives for the function f. Show all your work in your write-up.

(b) (10 pts) Write code to minimize f using gradient descent with a fixed learning rate. Attach the screenshot of the code here.

(c) (6 pts) Write code for the following experiment. You will run your code using $x = 1, y = 100$ from a few different initial values for (w_1, w_2) and a few different learning rates, and repeat this process using $x = 5, y = 50$. Attach the screenshot of the code here.

(d) (10 pts) Write up your results using prose, graphs, and tables as needed to explain the optimal values (w_1^*, w_2^*) found for each run. Discuss the impact of the starting point and the learning rate.

Three bonus points will be given if your homework is easy to review.