CS 4232/5232 – Fall 2023 – Homework 3
Assigned: 10/25/2023
Due: 11/01/2023 at 11:59 p.m.
Maximum Grade: 100 pts.

Objectives: The objectives of this homework are the following:

- Learn how to compute and plot the gradient of a multivariate function.
- Learn how to implement gradient descent.
- Learn how to fit linear regression models with the normal equations and gradient descent.
- Learn how to fit decision trees to solve classification problems.

Instructions:

- This homework is to be done individually.
- Submit **a single HTML file** named: yourlastname_hw3.html **and also the source R markdown file** (with a suffix of .Rmd) with the answers to your homework. To generate this HTML file, you **must use R markdown**. Check the video here for more information https://rmarkdown.rstudio.com/lesson-1.html.

Activity 1: (10 pts.) (Gradient Concept) Perform the following tasks related to the fundamentals of gradient descent.

a) (5 pts.) Compute the gradient of the function $f(x, y) = 18x^2y^3 + 7y^2$ by hand, showing each step of your calculation. Show all your work.
b) (5 pts.) Find and plot the values of the gradient of the above function at the points $(0,1)$ and $(1, -1)$. You can draw the plot by hand and scan it.

Activity 2: (20 pts.) (Gradient Descent Basics) Run by hand and in a step-by-step fashion the gradient descent algorithm for three iterations of k, where k is the variable shown in slide 52 of Lecture Topic 4 – Linear Regression (start from x_0 and compute up to and including x_2). Use $\lambda = 0.1$, x_0= [0.5, 0.1, 0.2], and H and z the same as the ones we used in the car example of class, the example where we seek to find the initial position x_0, velocity v_0, and constant acceleration a.

Activity 3: (15 pts.) (Implement Gradient Descent) Using R, reimplement the Matlab algorithm for gradient descent provided in the slide titled "Gradient Descent for Linear Least Squares."

Activity 4: (15 pts.) (Fitting Linear Regression Models) Do the following tasks in R using the same H and z as in the car example in class (not the horsepower and displacement but the one where we seek to find the initial position x_0, velocity v_0, and constant acceleration a):

a) (4 pts.) Using the normal equations, find the optimum value for x = (x_0, v_0, a). Do not use gradient descent. Show your work step by step.
b) (3 pts.) Using the lm command in R, find the optimum value for x = (x_0, v_0, a). Do not use gradient descent and do not use the normal equations.

c) (5 pts.) Using your R implementation of gradient descent, find the optimum value for x = (x_0, v_0, a). You may choose any values for lambda and the initial guess for the solution. There are two different elements called x_0 in this exercise; one is a number representing the initial position of the car, and the other is a 3x1 vector representing the initial guess for the solution.
d) (3 pts.) According to the models you fit in Activity a–c, where will the vehicle be 102 seconds after departure? Explain why your procedure is correct.

Activity 5: (20 pts.) (Optimizing with Gradient Descent) Using R, perform the activities below:

a) (4 pts.) Compute by hand and in a step-by-step fashion the gradient of $f(x, y) = (2.12 - x)^2 + 3(y - x^2)^2$. Show your work step by step.
b) (2 pts.) Modify your implementation of gradient descent of Activity 3 to find a local minimum for the function f.
c) (3 pts.) Explain in detail the reasons behind the changes you made in part b.
d) (4 pts.) Choose five different values (of different magnitudes) of lambda (learning rate) and plot in the same figure (they must all share the same x- and y-axis) the norm of the error $\|x_{k+1} - x_k\|_2$ as a function of k for each lambda. Your single plot must have labeled axes and a title. Show the code used to plot the figure.
e) (3 pts.) Find an R package implementing an optimizer and run it to find a local minimum for that function.
f) (4 pts.) How many steps did your algorithm perform until convergence? How many did the R optimizer do? What type of optimizer does R use in the background? Explain the reason for difference the number of steps (iterations) in detail. Show the code you used to find the number of steps performed. Cite the source that you use to figure out the type of optimizer that R uses.

Activity 6: (20 pts.) Consider the following dataset:

id	Home Owner	Marital Status	Income	Defaulted?
1	N	Married	300,000	**N**
2	N	Divorced	20,000	**Y**
3	Y	Married	100,000	**N**
4	N	Divorced	50,000	**Y**
5	N	Divorced	100,000	**N**
6	Y	Single	20,000	**N**
7	N	Married	300,000	**N**
8	Y	Single	50,000	**Y**
9	Y	Married	100,000	**N**
10	Y	Single	20,000	**N**

Run Hunt's algorithm by hand and in a step-by-step fashion, as explained in class, to build a decision tree that predicts if a person defaulted on their debt. Use Gini as the impurity measure. To get full credit, you must show all your work, including the calculation of all possible splits. To consider all splits for the quantitative attribute, follow the table procedure explained in class.

Important Notes:

- **You cannot have all your code in a single R markdown chunk**. You need to break down your markdown file into meaningful chunks. For example, one chunk can be a function definition, another a function call, another the definition of a set of related variables, etc.