Carnegie Mellon University in Qatar

AI for Medicine

15-182 - Spring 2023

Assignment 4

Name: _____

Andrew ID: _____

Due on: March 30, 2023 by midnight

Instructions:

- This assignment consists of two problems. Solve them both.
- Submit your solution through Gradescope.

Question	Points	Score
Is this tumor benign or malignant	35	
Predicting heart disease	40	
Total:	75	

Problem 1: Is this tumor benign or malignant (35 Points)

A medical researcher wants to classify tumor samples as either benign or malignant based on the size of the tumor and the age of the patient. They collect a dataset of 8 tumor samples, 4 of which are benign and 4 are malignant (see the table below). For each tumor, the researcher records the size of the tumor in centimeters and the age of the patient in years.

Tumor Size (cm)	Age (years)	Label (0=benign, 1=malignant)
3	50	0
2	40	0
3	20	0
4	70	0
6	80	1
5	75	1
3	55	1
7	85	1

Using logistic regression, the researcher builds a binary classification model to predict whether a tumor sample is benign or malignant based on the tumor size and patient age features. The logistic regression model is given by:

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}}$$

where θ is the vector of parameters and x is the vector of features.

You can assume a learning rate of 0.1 for this whole problem.

25pts (a) Find the optimal parameter values for θ_0, θ_1, and θ_2.

5pts (b) Using the parameter values above, calculate the predicted probability of malignancy for a tumor sample with size 4 cm and age 60 years. Round your answer to two decimal places.

5pts (c) Conceptually, how would increasing and decreasing the learning rate affect the training process?

Assignment continues on the next page(s)

Problem 2: Predicting heart disease (40 Points)

Suppose you have the following data on 10 patients who have either heart disease (1) or not (0):

Age (years)	Gender (0=female, 1=male)	Exercise (hours per week)	Heart disease (0=no, 1=yes)
50	1	2	1
45	0	3	0
60	1	1	1
55	1	2	1
50	0	0	0
40	0	1	0
35	0	3	0
30	0	2	0
65	1	4	1
55	1	1	1

5pts (a) Calculate the odds ratio of having heart disease for males compared to females (**Note**: you shall research and read more about *odds ratio* before answering this question).

5pts (b) Calculate the odds ratio of having heart disease for individuals who exercise 3 or more hours per week compared to those who exercise less than 3 hours per week (**Note**: you shall research and read more about *odds ratio* before answering this question).

25pts (c) Determine the coefficients of the logistic regression model for predicting the probability of having heart disease based on age, gender, and exercise.

5pts (d) Use the logistic regression model to predict the probability of having heart disease for a 50-year-old female who exercises 2 hours per week.