

# Carnegie Mellon University in Qatar

## AI for Medicine

15-182/282 - Spring 2022

### Assignment 4

Name: \_\_\_\_\_

Andrew ID: \_\_\_\_\_

**Due on:** April 14, 2022 by midnight

#### Instructions:

- This assignment has a maximum score of 92 points for the 15-182 students and 100 points for the 15-282 students.
- You should submit your solution through Gradescope.

Question	Points	Score
Inference Using Logistic Regression	10	
Learning Using Logistic Regression	40	
AI and Employment	50	
Total:	100	

**Problem 1: Inference Using Logistic Regression (10 Points)**

Suppose you have collected data for past students in the "AI for Medicine" course with features  $x_1 = \text{"Hours Studied"}$  and  $x_2 = \text{"Cumulative GPA"}$ , and label  $y = \text{"Received an A"}$  or  $y = \text{"Did Not Receive an A"}$ . Assume also that you learnt a logistic regression model that fits this data. The model ended up with a parameter vector  $\theta = [-6, 0.05, 1]$ .

Answer the following questions (show all your work):

5pts

- (a) Estimate the probability of a student getting "A" if they study for 40 hours and have a cumulative GPA of 3.5.

5pts

- (b) How many hours would the student in part (a) need to study to have a 50% chance of getting an "A" in the class?

*Assignment continues on the next page(s)*

## Problem 2: Learning Using Logistic Regression (40 Points)

The following training dataset assumes two classes, "1" and "0", and obeys the rule that the examples of class "1" all have vectors whose components sum to 10 or more, while the sum is less than 10 for the examples of class "0".

([3, 4, 5], 1) ([2, 7, 2], 1) ([5, 5, 5], 1)

([1, 2, 3], 0) ([3, 3, 2], 0) ([2, 4, 1], 0)

Answer the following questions (show all your work):

4pts

(a) Propose a parameter vector  $\theta$  such that the hypothesis function defined by  $\frac{1}{1+e^{-\theta^T x}}$  renders a good classifier for the "1" and "0" examples.

22pts

(b) Starting with your answer to part (a), use gradient descent to find the "optimal"  $\theta$  in less than 15 rounds, assuming a learning rate of 0.5. In this problem, we consider "optimum" to be the case when *every* example in the training dataset is predicted correctly by the hypothesis function. When this happens, you can assume that gradient descent has converged and stop training. (*Hint*: your choice in part (a) impacts the number of rounds gradient descent will take to converge; hence, you may want to revisit part (a) if gradient descent in this part does not converge in less than 15 rounds).

6pts

(c) Use the parameter vector learnt in part (b) to infer the classes of several new (say, 5 to 10) different feature vectors of your choice, some whose components sum to less than 10 and some whose components sum to 10 or more. How many of these vectors were classified correctly and why in your opinion some has been misclassified, if any?

8pts

(d) **This part is only for 15-282:** Write Python Code to train a logistic regression model using the above training dataset. How many rounds did your code take to converge to the "optimal" parameter vector (as defined in part (b))?

*Assignment continues on the next page(s)*

### Problem 3: AI and Employment (50 Points)

Now that you understand AI in its technical context, let us try to understand one of the societal concerns surrounding it, namely, job security, and develop a related response.

To begin with, productivity gains in certain tasks through AI may entail that fewer (or no) humans are required for producing the same output in the same tasks. It is argued that this does not necessarily imply a loss of overall employment because wealth increases with increased productivity, which can increase demand sufficiently to counteract the productivity gains (Vincent C. Muller 2019). In the long run, higher productivity in industrial societies has led to more wealth overall.

Major labour market disruptions have occurred in the past. For instance, farming employed over 60% of the workforce in Europe and North America in 1800, while by 2010 it employed around 5% in the EU, and even less in the wealthiest countries (European Commission 2013). In the 20 years between 1950 and 1970 the number of hired agricultural workers in the UK was reduced by 50% (Zayed and Loft 2019). In short, some of these disruptions have led to more labour-intensive industries moving to places with lower labour cost. This seems to be an ongoing process (Vincent C. Muller 2019).

There is one caveat though. Some argues that classic automation replaces human muscle while digital automation replaces human thought or information-processing—and unlike physical machines, digital automation is very cheap to duplicate (Bostrom and Yudkowsky 2014). It may thus mean a more radical change in the job market!

Conduct research on the subject of "automation and employment" and present your opinion on the matter in 1-3 pages. In your write-up, make sure to cite all the references that you have gone through during your research. The main questions that you may want to try to form opinions about are:

- (a) Will the effects of automation driven by AI be different this time on employment?
- (b) Do you think AI will fully replace certain jobs? If so, name and discuss one or two, if any.
- (c) Do you think AI will create new kinds of jobs? If so, name and discuss one or two, if any.
- (d) Do you think AI will replace medical doctors; not replace medical doctors but make a shift on how they work; or not replace medical doctors and not make any change on how they work? Discuss.
- (e) Is there a need in your opinion to come up with codes of ethics for AI like the codes of ethics for medical doctors? Discuss.