1) Which of the following methods do we use to best fit the data in Logistic Regression?
   A) Least Square Error
   B) Maximum Likelihood
   C) Both A and B

2) You are given a coin whose probability of landing on Heads is p. We toss a coin 10 times and get 7 Heads. What is the most-likely value of p?
   A) 7/10
   B) 5/10
   C) 3/10

3) Which of the following evaluation metrics does not make sense if applied to logistic regression output to compare with target?
   A) Accuracy
   B) Log loss
   C) Mean-Squared-Error

4) What can you say about feature normalization?
   A) It is a good practice but is not required to run logistic regression
   B) It is required to run logistic regression
   C) It is a bad practice and should not be performed to run a logistic regression
   D) None of the above

5) Consider a sample of independently and identically distributed (I.I.D.) random variables $x_1, x_2, ..., x_m$ that each have Geometric distributions. In other words, $x_i \sim \text{Geo}(p)$ for all $1 \leq i \leq m$.
   a) Derive the Maximum Likelihood Estimate for the parameter p of the Geometric distribution. Give a numeric answer. Explicitly show all the steps in your derivation.
   b) Say that we have a sample of five such I.I.D. Geometric variables with the following values: $x_1 = 4, x_2 = 3, x_3 = 4, x_4 = 2, x_5 = 7$. What value of p in the Geometric distribution would maximize the likelihood of these observations? Give a numeric answer.

6) In this problem we will simultaneously estimate the difficulty of problem set questions and the skill level of each student. Consider a set of 200 students and 10 questions where each student answers each question. Let $S_{ij}$ be an indicator variable which is 1 if student i answered question j correctly. You observe all $S_{ij}$.

   We are going to make the assumption that the probability ($p_{ij}$) that student i answers question j correctly is $p_{ij} = \sigma(a_i - d_j)$ where:
   - $\sigma$ is the sigmoid function,
   - $a_i$ is a parameter which represents a student’s ability
   - $d_j$ is a parameter which represents a question’s difficulty

   Use MLE to estimate the values for all parameters.
a) Write the log likelihood for a single response \( S_{ij} \) in terms of \( p_{ij} \) (hint logistic regression also assumes that its output is a probability of a binary event)

b) What is the partial derivative of LL for a single response \( S_{ij} \) with respect to \( a_i \)?

c) What is the partial derivative of LL for a single response \( S_{ij} \) with respect to \( a_j \)?

d) Explain briefly how you can estimate parameters given derivatives of log likelihood with respect to those parameters.

7) Suppose you train a logistic regression classifier and your hypothesis function \( h \) is

\[
f_{w,b}(x) = g(w_1 x_1 + w_2 x_2 + b) \text{ where } b = 6, w_1 = 0, w_2 = -1.\]

Which of the following figures will represent the decision boundary as given by the above classifier?

A) ![Figure A]

B) ![Figure B]

C) ![Figure C]

D) ![Figure D]

8) If you replace coefficient of \( x_1 \) with \( x_2 \) what would be the output figure?

A) ![Figure A (replaced)]]

B) ![Figure B (replaced)]

C) ![Figure C (replaced)]

D) ![Figure D (replaced)]

Similar to 7.

If \( W \) is a constant, what is the value of the slope of the logistic function at \( x = 0 \)?

You can assume that the model is just \( g(wx) \)

A) \( w \)

B) \( w/4 \)

C) \( 1/4 \)

D) \( w^2 \)
9) Below are two different logistic models with different values for \( b \) and \( w_1 \). Which of the following statement(s) is true about \( b \) and \( w_1 \) values of two logistics models (Green starts on top)?

\[ Y = g \left( w_1 X + b \right) \]

Note: consider \( Y = g \left( w_1 X + b \right) \). Here, \( b \) is intercept and \( w_1 \) is coefficient.

A) \( w_1 \) for Green is greater than Black
B) \( w_1 \) for Green is lower than Black
C) \( w_1 \) for both models is same
D) Can’t Say

10) Can a Logistic Regression classifier do a perfect classification on the data shown below?

Note: You can use only X1 and X2 variables where X1 and X2 can take only two binary values(0,1).

A) True
B) False
C) Can’t say
D) None of these

11) What can you say about regularized logistic regression vs. non-regularized logistic regression?

A) It will perform better on the training set
B) We can expect it to perform better on the training set
C) It will perform better on the testing set
D) We can expect it to perform better on the testing set