Chris Piech
CS 109

Section #8
March 5, 2025

Section 8: Machine Learning

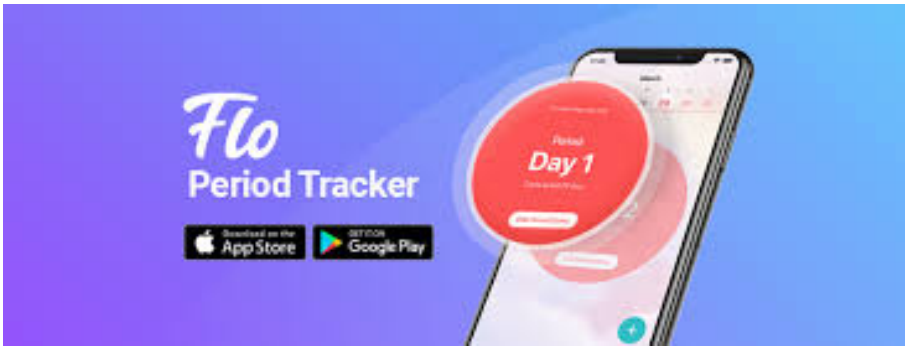
1. **Warmup! (Short Answer)**

   (a) True or False: The log likelihood function that is used to estimate p of a Bernoulli, for a set of observations, must always be 0 or smaller. Briefly explain why.

   (b) When implementing logistic regression, a student decides to add a second intercept value. To do so they add an extra feature with value 0 to each datapoint. How will this impact training?

   (c) When implementing logistic regression, a student decides to incorporate an additional term that represents a non-linear relationship or "interaction" between the features. Their dataset has two features, x_1 and x_2, and a corresponding label y for each datapoint. They add the interaction term $x_1 \cdot x_2$, so that the full model is $P(Y = y|X = x) = \sigma(\theta_0 + \theta_1 \cdot x_1 + \theta_2 \cdot x_2 + \theta_3 \cdot x_1 \cdot x_2)$. Explain briefly how this change impacts model training.

2. **Flo. Tracking Menstrual Cycles**



Let X represent the length of a menstrual cycle: the number of days, as a continuous value, between the first moment of one period to the first moment of the next, for a given person. X is parameterized by α and β with probability density function:

$$f(X = x) = \beta \cdot (x - \alpha)^{\beta - 1} \cdot e^{-(x-\alpha)^2}$$

a. For a particular person, $\alpha = 27$ and $\beta = 2$. Write an expression for the probability that they have their period on day 29. In other words, what is $P(29.0 < X < 30.0)$? You do not need to simplify.

b. For a particular person, $\alpha = 27$ and $\beta = 2$. How many times more likely is their cycle to last **exactly** 28.0 days than exactly 29.0 days? Simplify your expression.

c. A person has recorded their cycle length for 12 cycles stored in a list: $m = [29.0, 28.5, \ldots, 30.1]$ where m_i is the recorded cycle length for cycle i. Use MLE to estimate the parameter values α and β. Assume that cycle lengths are IID.

You don't need a closed-form solution. Derive any necessary partial derivatives and write up to three sentences describing how a program can use the derivatives in order to choose the most likely parameter values. Here is the PDF again:

$$f(X = x) = \beta \cdot (x - \alpha)^{\beta - 1} \cdot e^{-(x-\alpha)^2}$$

3. **Vision Test MLE**

> **Why is this question helpful?** In this problem you will be practicing all the mechanics and math behind logistic regression. The problem has all the same structural features (including a similar assumption), but it is not itself, logistic regression. Wahoo!

You decide that the vision tests given by eye doctors would be more precise if we used an approach inspired by logistic regression. In a vision test a user looks at a letter with a particular font size and either correctly guesses the letter or incorrectly guesses the letter.

You assume that the probability that a particular patient is able to guess a letter correctly is:

$$p = \sigma(\theta + f)$$

Where θ is the user's vision score and f is the font size of the letter. This formula uses the sigmoid function:

$$\sigma(z) = \frac{1}{1 + e^{-z}} \qquad \frac{\partial \sigma(z)}{\partial z} = \sigma(z)[1 - \sigma(z)]$$

Explain how you could estimate a user's vision score (θ) based on their 20 responses, $(f^{(1)}, y^{(1)})$ to $(f^{(20)}, y^{(20)})$, where $y^{(i)}$ is an indicator variable for whether the user correctly identified the ith letter and $f^{(i)}$ is the font size of the ith letter. Solve for all partial derivatives necessary.