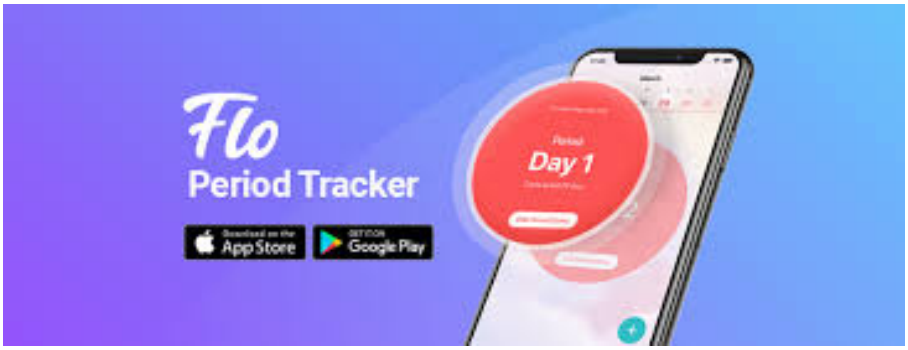
Chris Gregg
CS 109

Tutorial #7
November 16, 2026

Tutorial 7: Logistic Regression

1 MLE Flo



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:

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

1. 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.

2. 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.

3. 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 for 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.

2 Vision Test

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.

3 Recalibrating an Uncalibrated Model

You have an uncalibrated binary classification model that outputs values $\hat{p} \in [0, 1]$. These outputs are meant to be the probability that $Y = 1$. However, the outputs from this model are not well-calibrated. For instance, among all examples where $\hat{p} = 0.9$, it was the case that Y was 1 only 70% of the time. To recalibrate the models outputs you decide to use Platt Recalibration, where the corrected probability that $Y = 1$ is:

$$P(Y = 1 \mid \hat{p}) = \sigma(a \cdot \hat{p} - 0.5)$$

where $\sigma(z) = 1/(1 + e^{-z})$ is the sigmoid function and a is the parameter of the recalibration model. Here is the partial derivative of the Platt Recalibration model with respect to a:

$$\frac{\partial}{\partial a}\sigma(a \cdot \hat{p} - 0.5) = \sigma(a \cdot \hat{p} - 0.5) \cdot [1 - \sigma(a \cdot \hat{p} - 0.5)] \cdot \hat{p}$$

1. For a new datapoint the uncalibrated model outputs $\hat{p}$ of 0.9. If you use Platt Recalibration with $a = 2$ what is the recalibrated probability that $Y = 1$?

2. You are given a training dataset with n outputs from the uncalibrated model $(\hat{p}^{(i)}, y^{(i)})$ where $\hat{p}^{(i)}$ is the uncalibrated output and $y^{(i)} \in \{0, 1\}$ is the true binary outcome. Explain how you could estimate the value of a that makes the $y^{(i)}$ values as likely as possible. Solve for any and all partial derivatives required by your answer.