Juliette Woodrow
CS 109

Section #8
March 5, 2026

Section 8

Warmup!

(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) (Optional) 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) (Optional) 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.

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:

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

$\sigma(z) = 1/(1 + e^{-z})$ 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.

Vision Test Logistic Regression

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.

Decoding Movement for a Brain-Controlled Prosthetic Leg (Optional)

Engineers are designing a brain-controlled prosthetic ankle that infers a user's intended movement from electrical activity in their leg muscles. To train the system, the user performs known movements while electrodes measure muscle activity (EMG), producing labeled data that link muscle signals to intended actions.

At each time step, two muscle sensors are recorded:

- S_{TA}: tibialis anterior ("lift up" muscle),
- S_{GA}: gastrocnemius ("press down" muscle).

Each sensor reading is labeled as either `Active` (A) or `Quiet` (Q). The user can intend one of three movements:

$$U = \text{lift foot up}, \qquad D = \text{press foot down}, \qquad N = \text{neutral/relax}.$$

During calibration, the engineers measured how often each sensor fired while the user intended each movement. They found:

- When the user intends Up, sensor TA is Active 90% of the time and sensor GA is Active 20% of the time.
- When the user intends Down, sensor TA is Active 10% of the time and sensor GA is Active 85% of the time.
- When the user intends Neutral, sensor TA is Active 10% of the time and sensor GA is Active 10% of the time.

Engineers also found that when walking, a user spends about 30% of the time intending "Up," about 30% intending "Down," and the remaining 40% in "Neutral." Engineers model the two muscle sensors as independent once the users intended movement is specified.

We observe S_{TA} = Active and S_{GA} = Active, which intended movement $M \in \{U, D, N\}$ is most likely?