CS109 Lecture 26 Worksheet: Beyond Classification

Final-Exam Machine Learning: Logistic Regression, MLE for a New Distribution, and Recalibration

Summer 2026

Lecture 26 Worksheet: Beyond Classification

Topics: the machine-learning recipe applied end to end — writing a likelihood, taking its log, differentiating, and climbing the gradient — exercised on three full final-exam problems: logistic regression, maximum likelihood for the Gamma distribution, and recalibrating a miscalibrated classifier.

How to use this sheet. These are real problems from prior CS109 final exams. Every problem today is the *same three-step recipe* in disguise — assume a distribution for $Y \mid X$ (or for the data), write the log-likelihood, take derivatives, and follow the gradient. Once you see that Problems 1, 2, and 3 are the same machine wearing different costumes, you are ready for the final.

Useful facts. $\sigma(z) = \dfrac{1}{1+e^{-z}}$, and $\dfrac{d}{dz}\sigma(z) = \sigma(z)\big(1-\sigma(z)\big)$. For a distribution with PDF/PMF f, the log-likelihood of i.i.d. data is $LL = \sum_i \log f(\text{datapoint}_i)$. Gradient ascent steps in the direction of the gradient: $\theta \leftarrow \theta + \eta\, \nabla_\theta LL$.

Problem 1: Logistic Regression *(problem from the Spring 2026 Final)*

A logistic regression model predicts a binary label $Y \in \{0, 1\}$ from a feature vector. For a datapoint with features x_1 and x_2, the model predicts

$$P(Y = 1 \mid \mathbf{x}) = \sigma(\theta_0 + \theta_1 x_1 x_2)\,, \qquad \sigma(z) = \frac{1}{1+e^{-z}}.$$

(a) Suppose the trained weights are $\theta_0 = -1$ and $\theta_1 = 2$. For a new datapoint with $x_1 = 1$ and $x_2 = 1$, write an expression for $P(Y = 1 \mid \mathbf{x})$ and simplify it as far as you can.

(b) You are given a training set of n datapoints $\big(\mathbf{x}^{(i)}, y^{(i)}\big)$ with $y^{(i)} \in \{0, 1\}$. Use maximum likelihood estimation to define the optimal estimates for θ_0 and θ_1. Write the log-likelihood of the training labels, derive the partial derivatives of the log-likelihood with respect to θ_0 and θ_1, and state the gradient-ascent update rules (with learning rate η) used to find the optimal values.

Problem 2: It's Gamma Time *(problem from the Winter 2026 Final)*

The Gamma distribution is a continuous distribution that generalizes the Exponential. If $X \sim \text{Gamma}(\alpha, \beta)$ then it has the following PDF:

$$f(x \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \qquad \text{for } x > 0.$$

You have n i.i.d. samples from a Gamma: $[0.347,\ 0.811,\ 1.523,\ 0.194,\ 0.612,\ \ldots]$. Let x_i be the ith value.

The **digamma function**, $\psi(x)$, gives the derivative of $\log \Gamma(x)$ with respect to x:

$$\psi(x) = \frac{d}{dx} \log \Gamma(x).$$

The digamma function does not have an inverse function, so it is not possible to compute $\psi^{-1}(\alpha)$.

Explain how you would use MLE to estimate the parameters of this distribution and provide any necessary derivatives. You may assume you have access to the function `digamma(a)` which returns $\psi(a)$.

Problem 3: Recalibrating an Uncalibrated Model *(problem from the Fall 2024 Final)*

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 $\hat{p} \approx 0.9$, it was the case that Y was 1 only 70% of the time. To recalibrate the model's outputs you decide to use **Platt Recalibration**, where the corrected probability that $Y = 1$ is:

$$P(Y = 1 \mid \hat{p}) = \sigma(a \cdot \hat{p} - 0.5), \qquad \sigma(z) = \frac{1}{1 + e^{-z}},$$

and a is the parameter of the recalibration model. Here is the partial derivative of the Platt Recalibration model with respect to a:

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

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

(b) You are given a training dataset with n outputs from the uncalibrated model $\left(\hat{p}^{(i)}, y^{(i)}\right)$ 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.