Chris Piech
CS 109

Section #8
March 3, 2022

Section 8: Machine Learning

1. **Why Boba Cares About MAP**:

   You don't understand why there's no boba place within walking distance around campus, so you decide to start one. In order to estimate the amount of ingredients needed and the time you will spend in the business (you still need to study), you want to estimate how many orders you will receive per hour. After taking CS109, you are pretty confident that incoming orders can be considered as independent events and the process can be modeled with a Poisson.

   Now the question is - what is the λ parameter of the Poisson? In the first hour of your soft opening, you are visited by 4 curious students, each of whom made an order. You have a prior belief that $f(\Lambda = \lambda) = K \cdot \lambda \cdot e^{-\frac{\lambda}{2}}$. What is the MLE estimate? What is inference of λ given the observation? What is the Maximum-A-Posteriori (MAP) estimate of λ? Through your process try to identify what is a point-estimate, and what is a distribution.

2. **Vision Test MLE**

   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:

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

   Explain how you could estimate a user's vision score (θ) based on their 20 responses $(f^{(1)}, y^{(1)}) \ldots (f^{(20)}, y^{(20)})$, 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 any and all partial derivatives required by the approach you describe in your answer.

3. **Multiclass Bayes**

   Note: we don't expect folks to get to this problem in section. It is here just for students who would like some extra review!

In this problem we are going to explore how to write Naive Bayes for multiple output classes. We want to predict a single output variable Y which represents how a user feels about a book. Unlike in your homework, the output variable Y can take on one of the *four* values in the set $\{\text{Like}, \text{Love}, \text{Haha}, \text{Sad}\}$. We will base our predictions off of three binary feature variables $X_1, X_2,$ and X_3 which are indicators of the user's taste. All values $X_i \in \{0, 1\}$.

We have access to a dataset with 10,000 users. Each user in the dataset has a value for X_1, X_2, X_3 and Y. You can use a special query method **count** that returns the number of users in the dataset with the given *equality* constraints (and only equality constraints). Here are some example usages of **count**:

count($X_1 = 1, Y = \text{Haha}$)	returns the number of users where $X_1 = 1$ and $Y = \text{Haha}$.
count($Y = \text{Love}$)	returns the number of users where $Y = \text{Love}$.
count($X_1 = 0, X_3 = 0$)	returns the number of users where $X_1 = 0$, and $X_3 = 0$.

You are given a new user with $X_1 = 1$, $X_2 = 1$, $X_3 = 0$. What is the best prediction for how the user will feel about the book (Y)? You may leave your answer in terms of an argmax function. You should explain how you would calculate all probabilities used in your expression. Use **Laplace estimation** when calculating probabilities.