CS4650 Problem Set 0 (Fall 2022)

August 21, 2022

Problem Set 0 is a brief review of mathematical concepts necessary to succeed in this class. CS 4650 will cover deep learning and other machine learning methods for natural language processing. These will be discussed in a level of mathematical detail that is commonly understood by today's NLP practitioners. We will do our best to make this material easy to follow, but there is a certain level of mathematical background required, which is not possible to cover within a single semester. To succeed in the course, it is important for students to be familiar with the concepts and notation from probability, linear algebra and calculus. If you see any symbols or concepts in this assignment you don't recognize, this is a sign you probably need another math course before registering for CS 4650 - please feel free to reach out to the course staff to discuss this.

Similarly, students who do not have sufficient programming experience are likely to struggle with the homework assignments. The course will involve a significant programming component to implement the models discussed in lecture. We expect students are comfortable programming in Python and Numpy, or are able to learn a new programming language and environment very quickly. In addition to this assignment, we also recommend looking at the first programming assignment (Project 0)[1] to determine if you have the pre-requisite background to succeed in the course. Please be aware that the programming projects will require some amount of computation, and require access to GPUs to complete the assignments. Once you start working on the assignments that make use of PyTorch, you will need to subscribe for a Colab Pro account, which costs $10/month. This subscription is not needed for Project 0, but will be necessary for the assignments afterward.

Collaboration is **NOT** allowed. All questions represent material that students are expected to be familiar with before taking this class. Please show your work and write clearly. We will not be able to give credit for answers that are not legible.

Please submit your solutions on Gradescope.[2]

1 Joint and Marginal Probabilities

Assume the following joint distribution for $P(A, B)$:

$$\begin{aligned} P(A=0, B=0) &= 0.2 \\ P(A=0, B=1) &= 0.5 \\ P(A=1, B=0) &= 0.1 \\ P(A=1, B=1) &= 0.2 \end{aligned}$$

[1] https://colab.research.google.com/drive/1Iw73xF8KLH8afGGnYTR_H_IH9RNTQ7sq?usp=sharing

[2] https://www.gradescope.com/courses/417826

(a) **(1 point)** What is the marginal probability of $P(A = 0)$?

(b) **(1 point)** What is $P(B = 0|A = 1)$?

(c) **(1 point)** What is $P(A = B)$?

2 Independence

(2 points) Assume X is conditionally independent of Y given Z. Which of the following statements are always true? $\mathcal{X}_{\mathcal{Z}}$ represents the set of all possible values of random variable Z. Note that there may be more than one correct answer.

(a) $P(X, Y) = \sum_{c \in \mathcal{X}_Z} P(X, Y, Z = c)$

(b) $P(X, Y, Z) = P(X) + P(Y) + P(Z = c), c \in \mathcal{X}_Z$

(c) $P(X, Y|Z) = P(X|Z)P(Y|Z)$

(d) $P(X, Y, Z) = P(X) + P(Y) - P(Z)$

(e) $P(X, Y) = P(X)P(Y)$

3 Bayes Rule

(2 points) There is a 20% chance that a thunderstorm is approaching at any given moment. You own a dog that has a 60% chance of barking when a thunderstorm is approaching and only a 40% chance of barking when there is no thunderstorm approaching. If your dog is currently barking, how likely is it that a thunderstorm is approaching? *Hint*: write down the relevant probabilities and apply Bayes' Rule to answer this question.

4 Entropy

The entropy of a random variable x with a probability distribution $p(x)$ is given by:

$$\mathrm{H}[x] = -\sum_x p(x) \log_2 p(x)$$

Consider two binary random variables x and y having the joint distribution:

$$\begin{array}{cc|cc} & & \multicolumn{2}{c}{\text{y}} \\ & & 0 & 1 \\ \hline \text{x} & 0 & 0 & 1/5 \\ & 1 & 2/5 & 2/5 \end{array}$$

(a) (**1 point**) Evaluate $\mathrm{H}[x]$

(b) (**1 point**) Evaluate H[y]

(c) (**1 point**) Evaluate H[x, y]

5 Probability

(a) A probability density function is defined by

$$f(x) = \begin{cases} Ce^{-x} & \text{if } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

(i) (**1 point**) Find the value of C that makes $f(x)$ a valid probability density function.

(ii) (**1 point**) Compute the expected value of x, i.e., $E(x)$.

(b) (**2 points**) Three locks are randomly matched with three corresponding keys. What is the probability that at least one lock is matched with the right key?

6 Calculus Review

Consider the following function:

$$f(x) = x^2 \log_e(x)$$

(1 point) For what range of x is $f(x)$ differentiable and continuous?
(**2 point**) What is the minimum value of $f(x)$ (approximated to 3 decimal points)?
(**2 point**) Evaluate the following expression:

$$\lim_{x \to 0^+} f(x)$$

Hint: Use L'Hopital's Rule.

7 Multivariate Calculus

(a) (**2 points**) The number of members of a gym in Midtown Atlanta grows approximately as a function of the number of weeks, t, in the first year it is opened: $f(t) = 100(60 + 5t)^{2/3}$. How fast was the membership increasing initially (i.e., what is the gradient of $f(t)$ when $t = 0$)?

(b) Let $\mathbf{c}$ be a column vector. Let $\mathbf{x}$ be another column vector of the same dimension.

(i) (**1 point**) Consider a linear function $f(\mathbf{x}) = \mathbf{c}^\top \mathbf{x}$. Compute the gradient $\frac{\partial}{\partial \mathbf{x}} f(\mathbf{x})$.

(ii) (**1 point**) Consider a quadratic function $g(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top \mathbf{H}\mathbf{x}$. Compute the gradient $\frac{\partial}{\partial \mathbf{x}} g(\mathbf{x})$.

(iii) (**2 points**) Let $h(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{H}\mathbf{x} + \mathbf{c}^T\mathbf{x}$, where $\mathbf{H} = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$ and $\mathbf{c} = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$. When the gradient $\frac{\partial}{\partial \mathbf{x}} h(\mathbf{x}) = 0$, what is $\mathbf{x}$ =? Is it a local minimum, maximum or saddle point?