CS7650 Problem Set 1 (Spring 2022)

January 9, 2022

Problem Set 1 is a review of mathematical concepts in probability, linear algebra and calculus. The questions represent material students should be familiar with in order to succeed in the course. CS 7650 covers deep learning and other machine learning methods for natural language processing. This will involve a lot of math and programming to implement the models discussed in lecture. In addition to this assignment, we also recommend looking at the first programming assignment (Project 1)[1] to determine if you have the pre-requisite background to succeed in the course.

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 $P(A, B)$:

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

(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 $P(A = B)$?

[1] https://colab.research.google.com/drive/1ulvIvI5LWWGrk1P0gkcMY2X3tmJfUcZN

[2] https://www.gradescope.com/courses/344493/assignments/1753023/submissions

2 Independence

(2 points) Assume X is conditionally independent of Y given Z. Which of the following statements are always true? Note that there may be more than 1 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 10% chance that a thunderstorm is approaching at any given moment. You own a dog that has a 75% chance of barking when a thunderstorm is approaching and only a 25% chance of barking when there is no thunderstorm approaching. If your dog is currently barking, how likely is it that a thunderstorm is approaching?

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:

		y = 0	y = 1
x	0	0	1/5
	1	2/5	2/5

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

(b) (**1 point**) Evaluate $\mathrm{H}[y]$

(c) (**1 point**) Evaluate $\mathrm{H}[x, y]$

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., $E(x)$.

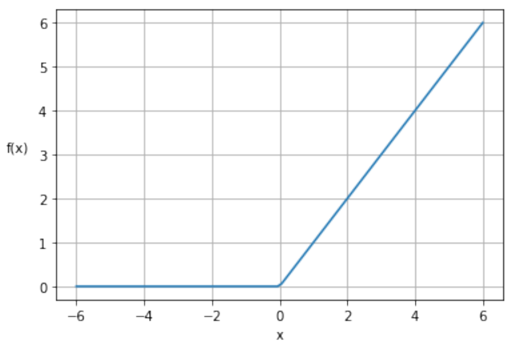
(b) (**2 points**) Three locks are randomly matched with three corresponding keys. What is the probability that at least one lock is matched with the right key?

6 Calculus Review

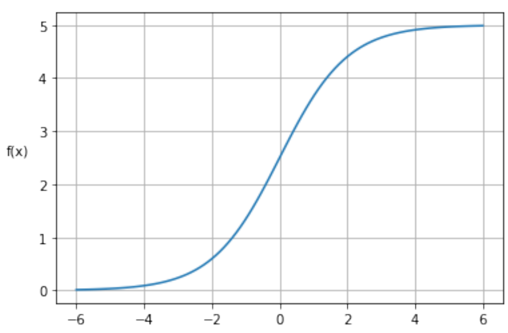
Consider the following function (often referred to as the *logistic* or *sigmoid* function):

$$f(x) = \frac{1}{1 + e^{-x}}$$

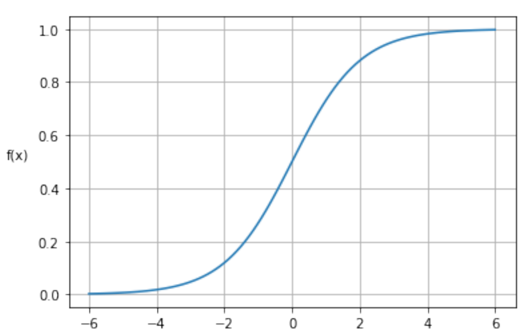
(1 point) Select the plot that matches the given equation.



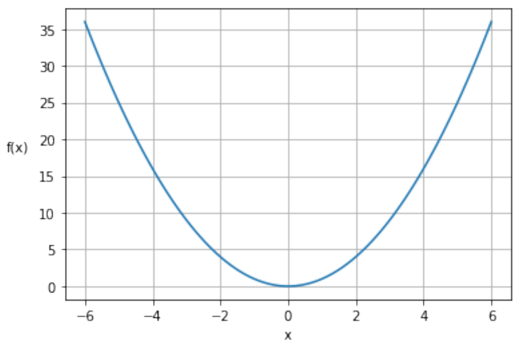
(a)



(b)



(c)



(d)

(**1 point**) What are the maximum and minimum values of this function? Why might this function be useful when considering probability? (1-2 sentences)

(**1 point**) Show that the derivative of $f(x)$ can be written simply in terms of the function's value like so:

$$\frac{df(x)}{dx} = f(x)(1 - f(x))$$

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?