CSCE 421: Fall 2020 Homework 1

Assigned Wed, Aug 26, due on Monday, Sep 7, by 11:59 PM.

Submit PDF and zip file (of code + latex) on Canvas. No late submissions accepted unless you have slip days remaining.

For your submission please name the files with *last name, first name, hw1, csce421.*

A Few Notes:

- You may submit a single jupyter notebook if it is convenient to both compose the written portions and the code together.
- Coding assignments are primarily designed for Python. You may, however, find R to be useful as well. If you wish to use any other language
- Please start early! This includes learning how to use Latex!
- if you need to use a different editor to write the solutions - please contact the TA and Instructor first
- This is an individual assignment. While you are welcome to discuss general concepts together and on the discussion board your solutions must be yours and yours alone.
- **SHOW YOUR WORK.**

Problem 1: Gradient Calculation. NOTE: This is not a programming assignment, so you may NOT use programming tools to help solve this problem. Show your work.

In this question you are required to calculate gradients for 2 scalar functions.

(1) Calculate the gradient of the function $f(x, y) = x^2 + ln(x) + xy + y^3$. What is the gradient value for $(x, y) = (1, -1)$?

(2) Calculate the gradient of the function $f(x, y, z) = tanh(x^3y^3) + sin(z)$. What is the gradient value for $(x, y, z) = (-1, 0, \pi/2)$?

Solution.
(1)

$$\nabla f(x,y) = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}) \tag{1}$$

$$\frac{\partial f}{\partial x} = 2x + \frac{1}{x} + y \tag{2}$$

$$\frac{\partial f}{\partial y} = x + 3y^2 \tag{3}$$

$$\nabla f(x,y) = (2x + \frac{1}{x} + y, x + 3y^2). \tag{4}$$

$$\nabla f(1,-1) = (2 + 1 - 1, 1 + 3) = (2, 4). \tag{5}$$

(2)

$$\nabla f(x,y,z) = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}) \tag{6}$$

$$\frac{\partial f}{\partial x} = \frac{3x^2y^3}{cosh^2(x^3y^3)} \tag{7}$$

$$\frac{\partial f}{\partial y} = \frac{3x^3y^2}{cosh^2(x^3y^3)} \tag{8}$$

$$\frac{\partial f}{\partial z} = cos(z) \tag{9}$$

$$\nabla f(x,y,z) = (\frac{3x^2y^3}{cosh^2(x^3y^3)}, \frac{3x^3y^2}{cosh^2(x^3y^3)}, cos(z)) \tag{10}$$

$$\nabla f(-1,0,\frac{\pi}{2}) = (\frac{3 \times 0}{cosh^2(-1 \times 0)}, \frac{-1 \times 0}{cosh^2(-1 \times 0)}, cos(\frac{\pi}{2})) = (0,0,0). \tag{11}$$

Problem 2: Matrix Multiplication. NOTE: This is not a programming assignment, so you may NOT use programming tools to help solve this problem. Show your work.

In this question you are required to perform matrix multiplication.

(1)

$$\begin{bmatrix} 1 & -1 & 6 & 7 \\ 9 & 0 & 8 & 1 \\ -8 & 5 & 2 & 3 \\ 10 & 4 & 0 & 1 \end{bmatrix} \begin{bmatrix} 6 & 2 \\ 0 & -1 \\ -3 & 0 \\ 11 & 4 \end{bmatrix} = \begin{bmatrix} 65 & 31 \\ 41 & 22 \\ -21 & -9 \\ 71 & 20 \end{bmatrix}$$

Work: Calculating the individual matrix coefficients:
$c_{1,1} = (1*6) + (-1*0) + (6*-3) + (7*11) = 65$
$c_{1,2} = (1*2) + (-1*-1) + (6*0) + (7*4) = 31$
$c_{2,1} = (9*6) + (0*0) + (8*-3) + (1*11) = 41$
$c_{2,2} = (9*2) + (0*-1) + (8*0) + (1*4) = 22$
$c_{3,1} = (-8*6) + (5*0) + (2*-3) + (3*11) = -21$
$c_{3,2} = (-8*2) + (5*-1) + (2*0) + (3*4) = -9$
$c_{4,1} = (10*6) + (4*0) + (0*-3) + (1*11) = 71$
$c_{4,2} = (10*2) + (4*-1) + (0*0) + (1*4) = 20$

(2)

$$\begin{bmatrix} 10 \\ 4 \\ -1 \\ 8 \end{bmatrix} \begin{bmatrix} 7 & 3 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 10 \begin{bmatrix} 7 & 3 & 0 & 1 \end{bmatrix} \\ 4 \begin{bmatrix} 7 & 3 & 0 & 1 \end{bmatrix} \\ -1 \begin{bmatrix} 7 & 3 & 0 & 1 \end{bmatrix} \\ 8 \begin{bmatrix} 7 & 3 & 0 & 1 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} 70 & 30 & 0 & 10 \\ 28 & 12 & 0 & 4 \\ -7 & -3 & 0 & -1 \\ 56 & 24 & 0 & 8 \end{bmatrix}$$

(3)

$$\begin{bmatrix} 9 & -3 & 1 & 6 \end{bmatrix} \begin{bmatrix} -3 \\ 4 \\ -9 \\ 0 \end{bmatrix} = \begin{bmatrix} (9 * -3) + (-3 * 4) + (1 * -9) + (6 * 0) \end{bmatrix} = \begin{bmatrix} -48 \end{bmatrix}$$

Problem 3: Vector Norms. NOTE: This is not a programming assignment, so you may NOT use programming tools to help solve this problem. Show your work.

Consider these two points in the 3-dimensional space:

$$\mathbf{a} = \begin{bmatrix} 5 \\ 0 \\ -1 \\ 4 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 7 \\ 9 \\ 5 \\ 2 \end{bmatrix}, \textbf{(a-b)} = \begin{bmatrix} -2 \\ -9 \\ -6 \\ 2 \end{bmatrix}$$

Calculate their distance using the following norms:

(1) $\ell_0 : \|a - b\|_0 = 0$

(2) $\ell_1 : \|a - b\|_1 = \| \begin{bmatrix} -2 \\ -9 \\ -6 \\ 2 \end{bmatrix} \|_1 = |-2| + |-9| + |-6| + |2| = 19$

(3) $\ell_2 : \|a - b\|_2 = \sqrt{(-2)^2 + (-9)^2 + (-6)^2 + (2)^2} = \sqrt{125} = 5\sqrt{5}.$

(4) $\ell_\infty : \|a - b\|_\infty = 9$

Problem 4: Probability Calculation. NOTE: This is not a programming assignment, so you may NOT use programming tools to help solve this problem. Show your work.

Consider a problem where we are rolling 2 dices where each dice has 6 faces numbered from 1 to 6. Answer the following questions:

(1) What is the sample space? (Written in the form {first roll, second roll}): $\{(1, 1), (1, 2), (1, 3), ..., (1, 6), (2, 1), (2, 2), ..., (2, 6), ..., (6, 6)\}$

(2) If the event we are interested in is the sum being 10, what would be the probability of observing such an event? All outcomes where the sum is 10: $\{(4, 6), (5, 5), (6, 4)\}$. The size of this set is 3. Number of total outcomes: $6 \times 6 = 36$. Probability of rolling a 10-sum: $\frac{3}{36} = \frac{1}{12} = 8.\overline{3}\%$.

(3) If the event we are interested in is the sum being 6, what would be the probability of observing such an event? All outcomes where the sum is 6: $\{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}$. The size of this set is 5. Number of total outcomes: $6 \times 6 = 36$. Probability of rolling a 6-sum: $\frac{5}{36} = 13.\overline{8}\%$.

Problem 5: Mean/Variance Calculation. NOTE: This is not a programming assignment, so you may NOT use programming tools to help solve this problem. Show your work.

Assume we have a random variable X with a Uniform probability density function. Uniform probability density is defined as:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & o.w. \end{cases}$$

(1) What is the mean of X?
$E(x) = \int_a^b \frac{xdx}{b-a} = \frac{1}{b-a}\int_a^b (x\ dx) = \frac{1}{b-a} \times [\frac{b^2}{2} - \frac{a^2}{2}] = \frac{1}{b-a} \times \frac{b^2-a^2}{2} = \frac{1}{b-a} \times \frac{(b+a)(b-a)}{2} = \frac{a+b}{2}$. (2) What is the standard deviation of X?
The equation for the variance of the normal distribution is $\frac{(b-a)^2}{12}$.
So by taking the square root of that, we get the standard deviation: $\frac{b-a}{\sqrt{12}}$.

Problem 6: Classification Quality Metric Computation. NOTE: This is not a programming assignment, so you may NOT use programming tools to help solve this problem. Show your work.

Following up on the example presented in the class about Taqueria El Tio, assume Martin and Jose have decided to take it to the next level and they have bought a microwave avocado detector to detect tacos with no avocados inside. Here is the confusion matrix of the microwave avocado detector:

		ground truth	
		avocado	no avocado
Avocado detector	avocado	37	23
	no avocado	45	55

(1) What is the accuracy of the detector? $\frac{A+D}{A+B+C+D} = \frac{37+55}{37+23+45+55} = \frac{92}{160} = 0.575$.

(2) What is the balanced accuracy of the detector? Avocado accuracy: $\frac{37}{37+23} = 0.6167$. No Avocado accuracy: $\frac{55}{55+45} = 0.55$. Balanced accuracy $= \frac{0.6167+0.55}{2} = 0.5834$

(3) What is the precision of the detector? $\frac{TP}{TP+FP} = \frac{37}{37+45} = 0.4512$.

(4) What is the recall of the detector? $\frac{TP}{TP+FN} = \frac{37}{37+23} = 0.6167$.

(5) What is the F-measure of the detector? $2 \times \frac{Precision \times Recall}{Precision+Recall} = \frac{2 \times 0.4512 \times 0.6167}{0.4512+0.6167} = 0.5211$.

Problem 7: ROC Computation. NOTE: This is not a programming assignment, so you may NOT use programming tools to help solve this problem. Show your work.

In Problem 6, assume that their microwave avocado detector does not give a binary output regarding the existence of avocados inside the taco. Alternatively, it outputs a probability of such an event. Jose, a CS sophomore who wants to put his knowledge to practice, wants to approximate the AUROC of the detector using 5 points as candidate thresholds: $\{0, 0.25, 0.5, 0.75, 1\}$. In a few tests that they ran, the probabilities and their corresponding ground truths were as follows:

predicted	ground truth
10%	0
5%	0
70%	1
50%	0
90%	1
65%	1
35%	1
60%	0
15%	1
20%	0

Please help him by computing the following:

(1) What would be the ROC value for threshold = 0? (1, 1)

(2) What would be the ROC value for threshold = 0.25? (0.4, 0.8)

(3) What would be the ROC value for threshold = 0.5? (0.4, 0.6)

(4) What would be the ROC value for threshold = 0.75? (0, 0.2)

(5) What would be the ROC value for threshold = 1? (0, 0)

(6) What would be the AUROC approximation using the above results? (HINT: remember Riemann sum): The best Riemann sum I can give for this curve is: $\frac{1}{5} \times (0.2 + 0.6 + 0.8 + 0.9 + 1) = \frac{3.5}{5} = 0.7$.

First, plot the outputs on varying thresholds using values (0, 0.25, 0.5, 0.75, 1):

Predicted	Ground Truth	Threshold: 0	Threshold: 0.25	Threshold: 0.5	Threshold: 0.75	Threshold: 1
10%	0	1	0	0	0	0
5%	0	1	0	0	0	0
70%	1	1	1	1	0	0
50%	0	1	1	1	0	0
90%	1	1	1	1	1	0
65%	1	1	1	1	0	0
35%	1	1	1	0	0	0
60%	0	1	1	1	0	0
15%	1	1	0	0	0	0
20%	0	1	0	0	0	0

Then, use these values to calculate the confusion matrix values for each threshold, and calculate TPR and FPR:

-------------	Threshold: 0	Threshold: 0.25	Threshold: 0.5	Threshold: 0.75	Threshold: 1
TP 11	5	4	3	1	0
FP 01	5	2	2	0	0
TN 00	0	3	3	5	5
FN 10	0	1	2	4	5
TPR	1	0.8	0.6	0.2	0
FPR	1	0.4	0.4	0	0

I used these calculations to arrive at the results above.

Problem 8: Coding K-NN.

This is a coding assignment. Throughout the course, you will have several coding assignments. You are free to choose any language you prefer but our preference, and our hints, are directed towards Python. This can be a good place to get you going with Python if you haven't already.

In this assignment we will implement k-NN. More specifically, we are interested in seeing the effect of varying k on the performance.

The dataset we will use in this assignment is named *Smarket* (can be downloaded from https://github.com/jcrouser/islr-python/blob/master/data/Smarket.csv).

Your submission should include a script which can be run seamlessly and performs all the following steps one after another. Any submission with a runtime error would result in lost points.
You may use libraries and you do not need to implement anything from scratch.

(1) Download and read the data. For Python, you may use *pandas* library and use *read csv* function

(2) Print the data. How does the data look like? (For Python, you may use *head()* function in *pandas* library)

(3) Print the shape of the data. Shape means the dimensions of the data. (In Python, *pandas* dataframe instances have a variable *shape*)

(4) Extract the features and the label from the data. The features we are interested in are *Lag1* and *Lag2* and the label is *Direction*.

(5) Split the data into a train/test split. (In Python, you can use *train test split* from *sklearn* library.)

(6) Apply k-NN to the data. (In Python, you can use the *KNeighborsClassifier* function from *sklearn* library.)

(7) Plot the accuracy of your implementation for $k \in 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$.