6.034 Quiz 3

November 9, 2020

You will have 5 hours total to complete and upload the quiz (4 hours to complete the quiz, 1 hour to upload) in a 36-hour submission window starting 6am EST, November 9, 2020 through 6pm EST, November 10, 2020. Quizzes should take on average about 1 hour.

Note: If you open the quiz with less than 5 hours remaining in the submission window, you will only have the remaining time until 6pm, November 10, 2020 to complete and upload the quiz.

This quiz is open book and open notes. You may NOT use the Internet **(including the recitation videos)** or other people. You can resubmit the quiz as many times as you want on Gradescope before the quiz closes.

Updates and clarifications will be posted to a pinned Piazza post for this quiz. **Send private Piazza posts** to the instructors to ask for clarification. Do not post publically or make follow-up discussions on the pinned post—they will be deleted.

Instructions and tear-off sheets are pages 1-4. The actual quiz is from page 5.

Gradescope

Quiz 3 submissions work exactly like Quiz 2, on Gradescope.

You will download a single PDF template of the quiz (with optional tear-offs). Your quiz submission will be a single PDF or multiple camera images, submitted on Gradescope. **You must then designate which pages correspond to which questions on Gradescope.**

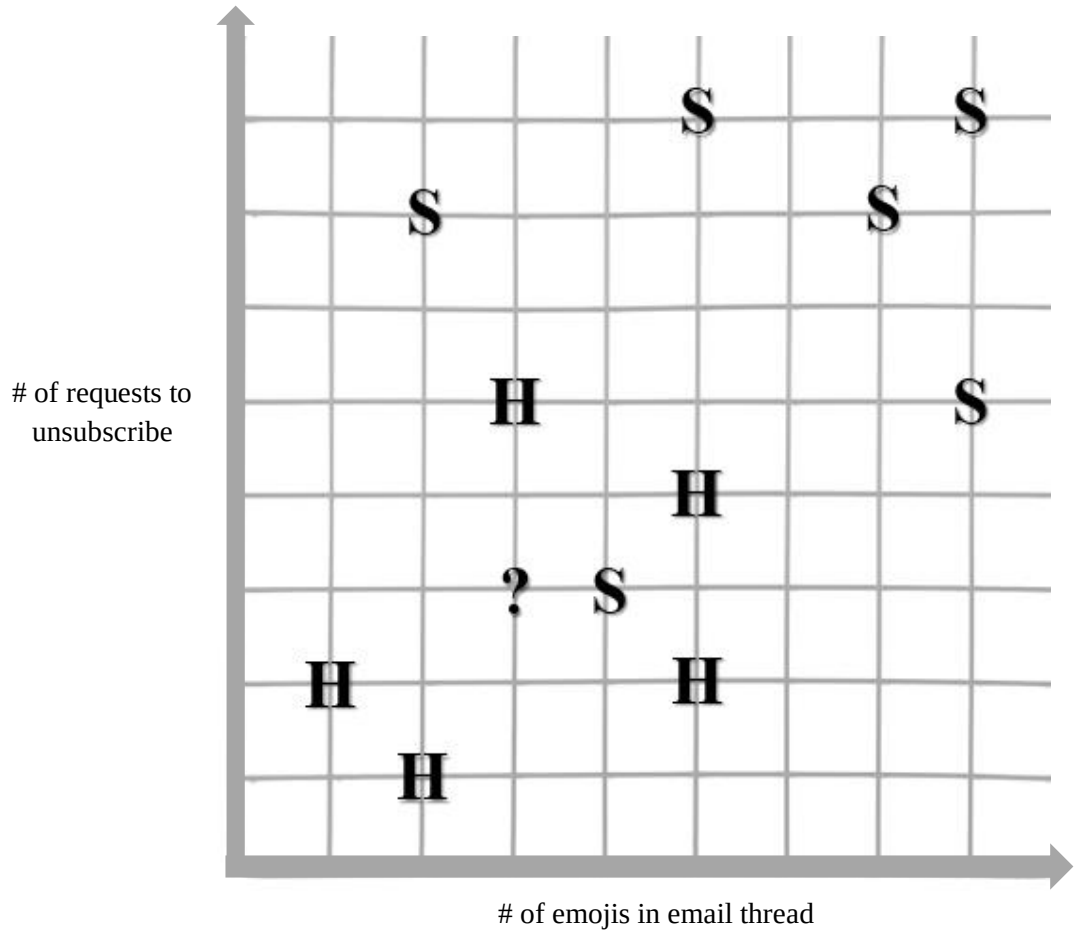
You will be given one extra-credit point if you correctly designate your submission's pages to the corresponding questions. Only designate the pages that have your written work i.e. the pages that need to be graded. Do not designate any instruction pages or tear-off sheets.

Here are three methods to submit the quiz:

1. Edit the PDF electronically e.g. Adobe Acrobat, Preview (OS X), Notability (OS X/iOS). These are just examples--you are free to use whatever PDF editing tool you want. Acrobat and Notability are free for MIT students.
2. Print out the PDF, write your answers, and scan/take pictures.
3. Write your answers on a blank sheet of paper, making sure to label your answers clearly. Scan/take pictures.

Tear-off sheet. We do not collect tear-off sheets, so please show your work on the quiz pages, not the tear-off sheet.

Email Classifications for Problem 1 (k-NN) Part B



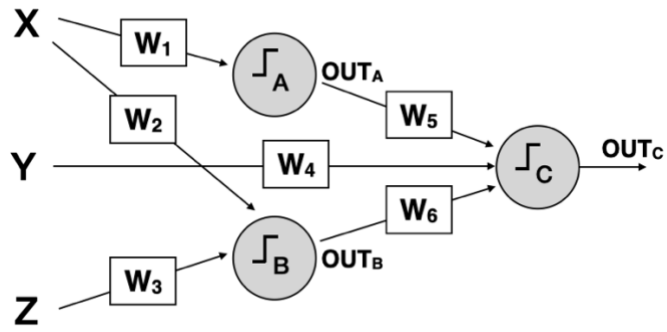
Tear-off sheet. We do not collect tear-off sheets, so please show your work on the quiz pages, not the tear-off sheet.

Feature Table for Problem 2 (ID Trees)

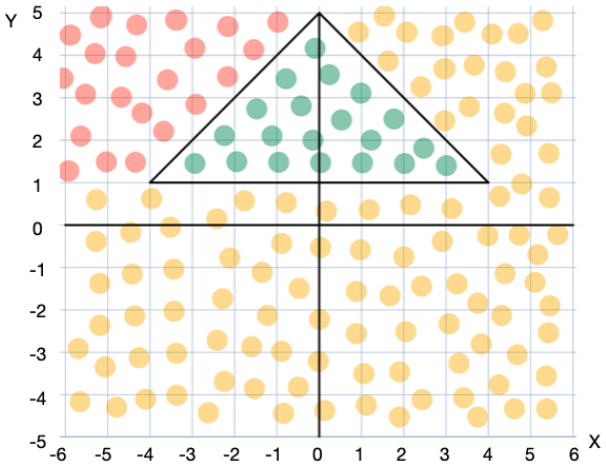
Name	Popularity	Type of Milk	Firm	Age (months)	Mold
Parmigiano	**High**	Cow	Yes	24	No
Gorgonzola	**Low**	Cow	No	4	Yes
Roquefort	**Low**	Sheep	No	5	Yes
Sharp Cheddar	**High**	Cow	Yes	12	No
Comté	**Medium**	Cow	No	24	No
Manchego Curado	**Medium**	Sheep	Yes	12	No
Pecorino Romano	**Medium**	Sheep	Yes	5	No
Oaxaca	**High**	Cow	No	1	No

Tear-off sheet. We do not collect tear-off sheets, so please show your work on the quiz pages, not the tear-off sheet.

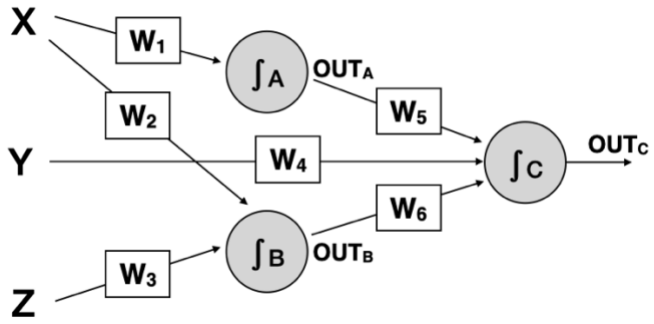
Neural Net for Problem 3 (Neural Nets) Part A



Classifications for Problem 3 (Neural Nets) Part A



Neural Net for Problem 3 (Neural Nets) Part B



Problem 1: k-Nearest Neighbors (k-NN) (25 points)

Part A: Doofenshmirtz Evil, Incorporated (8 points)

Dr. Doofenshmirtz is back at it again with a dastardly plot to wreak havoc on the Tri-State area. He's tired of his archnemesis, Perry the Platypus, constantly foiling his plans, so he decides to use *the power of Machine Learning* (namely k-Nearest Neighbors) to help him succeed. He enlists you, his trusty assistant, to help him choose the right distance metric for each k-NN model.

First, Dr. Doofenshmirtz wants to predict which of his future inventions will most likely be classified as Chaos-Inducing or Not Chaos-Inducing. He collects a series of different data points based on his past inventions, which have the following features and values. Each feature is restricted to the values in the below table.

Feature	*Values*
Invention name ends in "-inator"	Yes/No
Color	Red/Green/Blue
Size	Small/Medium/Large

1.1 (2 points) Circle the best distance metric that Dr. Doofenshmirtz should use for his kNN algorithm.

A. Euclidean Distance

B. Manhattan Distance

C. Hamming Distance

D. Cosine Distance

In 1 sentence, briefly explain your answer.

Based on the previous k-NN model and inspired by his newfound quarantine hobbies, Dr. Doofenshmirtz choses to use the "Bread-inator" to spew bread all over the city. However, he realizes he needs at least 30 minutes to carry out his nefarious plan and fill the air with the deliciously evil smell of sourdough. In order to successfully set up the "Bread-inator", Dr. Doofenshmirtz decides to once again use k-NNs to help him choose a location where Perry the Platypus is expected to take longer than 30 minutes to arrive.

He draws out a map of the Tri-State area, plots all the places from where he's conducted his attacks in the past, and labels them with whether Perry the Platypus took greater than 30 minutes to arrive (> 30 min), or less than 30 minutes to arrive (< 30 min). He has secret intel that Perry the Platypus's hang glider is broken and thus will be forced to use the roads, which are laid out in a uniform grid, to travel.

1.2 (2 points) Given this new information above, circle the best distance metric that Dr. Doofenshmirtz should use for his kNN algorithm.

A. Euclidean Distance

B. Manhattan Distance

C. Hamming Distance

D. Cosine Distance

In 1 sentence, briefly explain your answer.

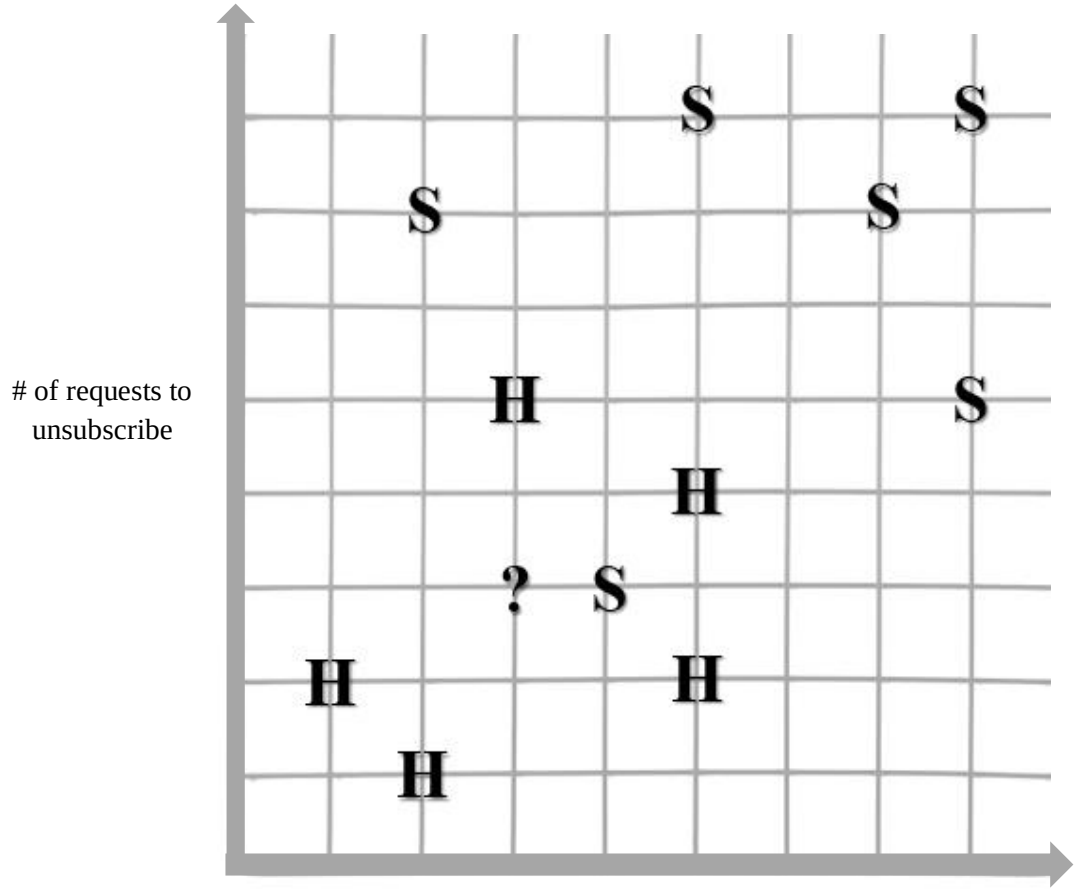
1.3 (2 points) Dr. Doofenshmirtz fears that if his model isn't accurate enough, Perry the Platypus will foil his plans yet again. He chooses a random value of k and decides to evaluate his model before deploying it. Given a dataset with known classifications, **in at most 1 sentence**, explain how he can test how accurate his model is.

1.4 (2 points) The value of k Dr. Doofenshmirtz chose in A3 turns out to yield a low accuracy. **In at most 2 sentences,** explain how he can find the optimal value of k to use given the dataset defined in A3.

Part B: Update: Tetazoo Glounge (10 points)

You are sick and tired of flamewars from dormspam[1] clogging up your email inbox. You couldn't care less about how someone's "plont" [2] is doing, and certainly don't want to receive 10,000+ word copypastas of the Bee Movie script. However, you don't want to unsubscribe entirely from your dorm's mailing list, so using your 6.034 knowledge, you build a model using k-NN to classify new incoming emails as spam based on the number of emojis in the email thread and the number of requests to unsubscribe.

You decide to evaluate how different values of k will fair in classifying your emails as Spam **(S)** and Ham **(H)**. ("Ham" is email that is *not* spam.) You give your model an unknown point **(?)** to see what it will be classified as.



[1] MIT undergraduate dorm mailing lists that are generally used for announcing events.

[2] Earlier this semester, someone, adopting the moniker "Plont", decided to email the entire undergraduate population with updates on their plant's health. Of course, in true MIT fashion, this spurred what became a flamewar consisting of plant puns, requests to unsubscribe, and many, many well wishes for the plant to recover soon.

2.1 (4 points) Using the graph on the previous page, for each value of k circle what the unknown point **(?)** will be classified as. If there is not enough information to determine a classification, circle "CAN'T TELL". Use Euclidean distance for your distance metric.

k=1	S	H	CAN'T TELL
k=2	S	H	CAN'T TELL
k=5	S	H	CAN'T TELL
k=11	S	H	CAN'T TELL

2.2 (4 points) Using the same graph and still using Euclidean distance for your distance metric, circle whether the model will be overfitting or underfitting the data for each value of k given below. Circle "NEITHER" if the model neither overfits nor underfits the data.

k=1 OVERFITTING UNDERFITTING NEITHER

Briefly explain your answer in one sentence or show relevant work.

k=2 OVERFITTING UNDERFITTING NEITHER

Briefly explain your answer in one sentence or show relevant work.

k=5 OVERFITTING UNDERFITTING NEITHER

Briefly explain your answer in one sentence or show relevant work.

k=11 OVERFITTING UNDERFITTING NEITHER

Briefly explain your answer in one sentence or show relevant work.

2.3 (2 points) You realize that two of the emails that were originally classified as Spam are actually Ham. Circle whether this re-classification results in overfitting or underfitting the data.

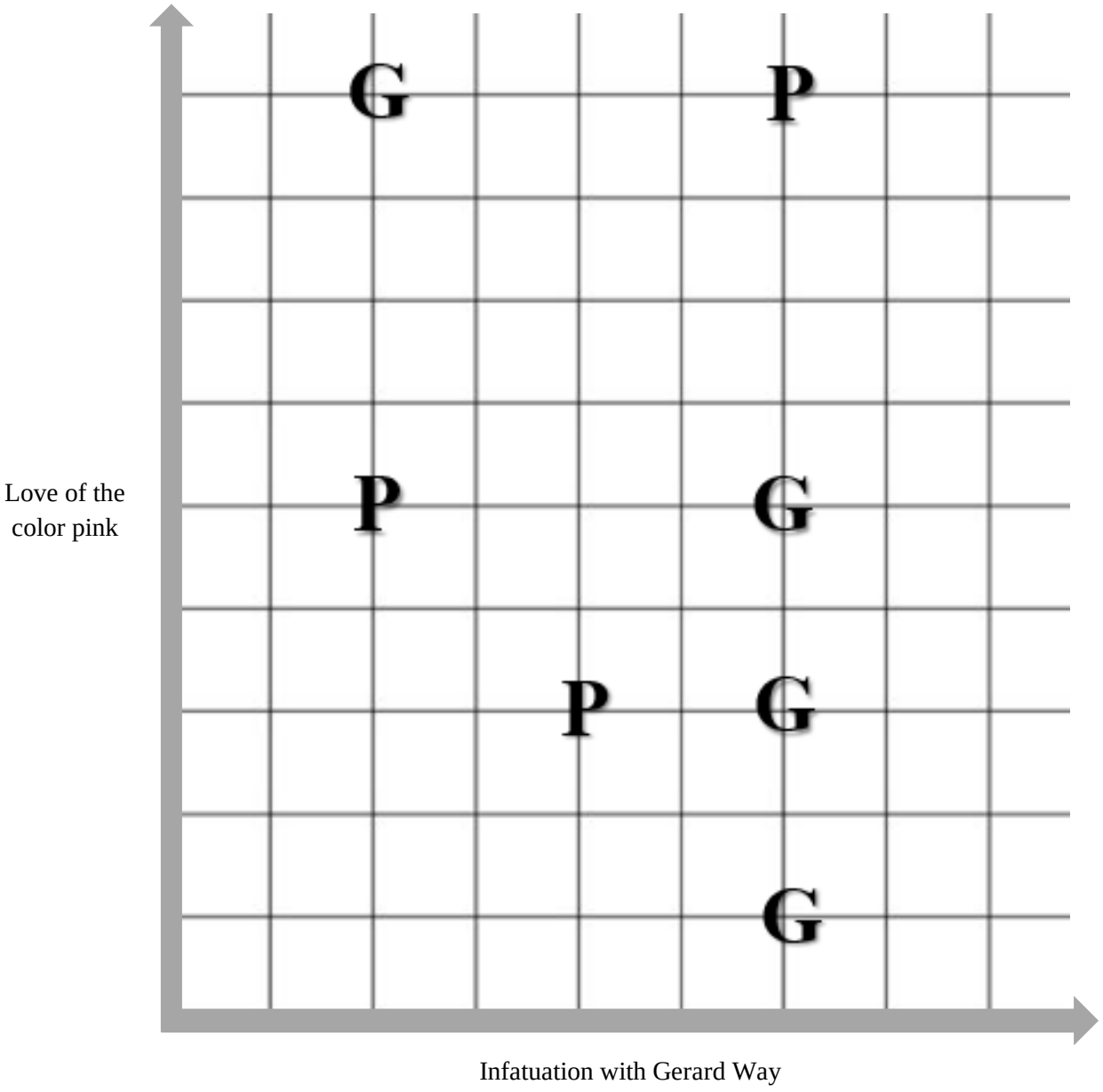
k=11 OVERFITTING UNDERFITTING NEITHER

Has your answer for k=11 changed from your answer in B2 (the previous question)? Why or why not?

Part C: Sorting First-Years (7 points)

Ebony Dark'ness Dementia Raven Way, of the acclaimed literary composition *My Immortal*, is trying to predict which of the incoming first-year Hogwarts students will be classified as "preps" **(P)** and which will be classified as "goffs" **(G)**. Being the wise headmaster he is, Albert Dumblydore suggests she gather data on the current students and use k-Nearest Neighbors to help her in her goal. Ebony collects some preliminary datapoints and wants to visualize her results.

3. (7 points) Use k-NNs where k=1 to draw the classification/decision boundaries on the graph below. Use Euclidean distance as the distance metric. If you are hand-copying this graph on a blank sheet of paper, you must include the grid lines.



Problem 2: Identification (ID) Trees (35 points)

You are elected president for the College of Cheese (CoC) and are tasked with ordering cheese for all the social events this year. There is a quickly approaching Tent Party that will take place in the brand new Artificial Indoors (AI) Tent, designed to mitigate Covid transmission to the same degree as hosting the event outdoors. The Cheese Science and Artificial Indoors Laboratory (CSAIL) is really proud to be hosting this event, and you don't want to cause any trouble by ordering too much or too little of each cheese variety.

In order to do this, you need to classify each cheese variety based on its **popularity** level (Low, Medium, or High). Knowing nothing about cheese, you gather information about each cheese variety's features and come up with the following dataset:

Name	Popularity	Type of Milk	Firm	Age (months)	Mold
Parmigiano	**High**	Cow	Yes	24	No
Gorgonzola	**Low**	Cow	No	4	Yes
Roquefort	**Low**	Sheep	No	5	Yes
Sharp Cheddar	**High**	Cow	Yes	12	No
Comté	**Medium**	Cow	No	24	No
Manchego Curado	**Medium**	Sheep	Yes	12	No
Pecorino Romano	**Medium**	Sheep	Yes	5	No
Oaxaca	**High**	Cow	No	1	No

Part A: Oh Cheezus, the Disorder! (10 points)

4. (10 points) Compute the average test disorder for each of the following two feature tests. You are encouraged to **express your answer with log terms**; you do NOT need to simplify these to a numerical value to obtain full credit.

Feature Test	Average Test Disorder (expressed in log terms)
A1.1: Type of Milk (5 points)	Answer: Calculations:
A1.2: Firm (5 points)	Answer: Calculations:

Part B: I wish Cheese grew on Trees (20 points)

5.1 (16 points) Construct a **greedy, disorder-minimizing ID tree** to correctly classify all of the cheeses as having **Low, Medium, or High popularity** using the four feature tests: **Type of Milk, Firm, Age,** and **Mold.**

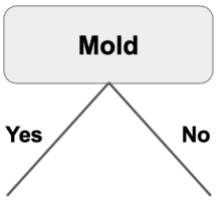
Constraints:

Your friendly neighborhood Foodie requires that

1. You start your ID tree with **Mold** as the first feature test.
2. You treat the **Age** feature as having only **two values** (Low = less than 12 months, High = at least 12 months).

If there are any ties between classifiers, choose the one that comes first in alphabetical order. A full credit tree must include: clearly labeled feature test nodes, branches, and leaves. There is space on the next page to show work for possible partial credit.

Draw your ID tree here.



5.1 Partial Credit

Show your work and/or disorder calculations here for partial credit:

5.2 (4 points) The day before you have to place the order for cheeses, President Rafael Reif requests that you also include two of his favorite cheeses in the order. Using the ID tree you constructed in B1 above, determine how popular these cheese varieties are:

Name	Popularity	Type of Milk	Firm	Age (months)	Mold
Wagassi	**??**	Cow	No	2	No
Kanterkomijnekaas	**??**	Cow	Yes	Unknown	No

5.2.1 (2 points) Circle the classification for Wagassi:

Popularity = **High** **Medium** **Low** **Can't Tell**

5.2.2 (2 points) Circle the classification for Kanterkomijnekaas:

Popularity = **High** **Medium** **Low** **Can't Tell**

Part C: Vegan Cheese: Is it Worth it? (5 points)

Your friendly neighborhood butcher suggests that you also take into account whether each of the cheeses is vegan in order to determine their popularity. You add a new feature, **Vegan (Yes/No)**, to your dataset as follows:

Name	Popularity	...	Vegan
Parmigiano	**High**	**...**	No
Gorgonzola	**Low**	**...**	No
Roquefort	**Low**	**...**	No
Sharp Cheddar	**High**	**...**	No
Comté	**Medium**	**...**	No
Manchego Curado	**Medium**	**...**	No
Pecorino Romano	**Medium**	**...**	No
Camembert	**High**	**...**	No

6. (5 points) Is the Butcher's suggestion helpful? In other words, would having a **Vegan** feature impact the ID tree you constructed in Question 2?

Circle one: **YES** **NO** **CAN'T TELL**

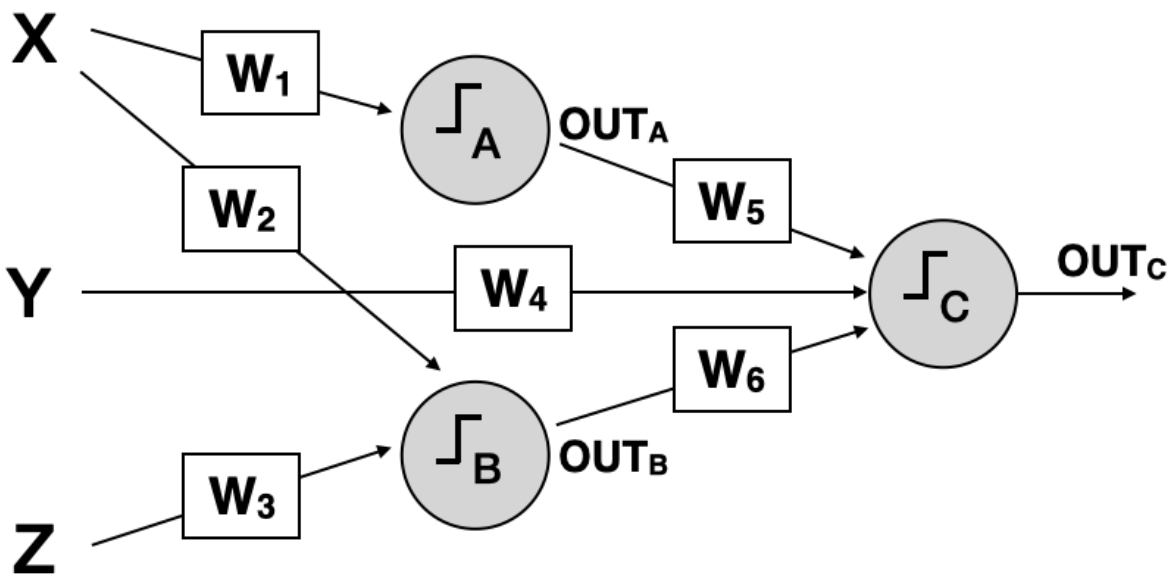
Why or why not?

Problem 3: Neural Networks (40 points)

Bradley Ericford has to deal with a whole new set of challenges this fall with online classes. There are so many decisions to make when he logs into Zoom that it really stresses him out, so he turned to neural networks for assistance.

Part A: Forward Propagation (20 points)

The following neural network helps Bradley Ericford decide whether he should turn his camera on in lecture. It takes as input certain factors that go into this decision. For example, **X** measures his current mood, **Y** measures how brushed his hair is, and **Z** measures his current internet bandwidth. The network outputs a 1 if he should turn on his camera and a 0 if he should not. All neurons use the stairstep threshold function.



7.1 (7 points) Write the expressions for the outputs of neurons **A-C**, in terms of the inputs (**X, Y, Z**), the unknown weights (**W_{1-6}**) and the unit step threshold function u(x), shown below**.**

$$u(x) = \begin{cases} 0 & x < T \\ 1 & x \geq T \end{cases}$$

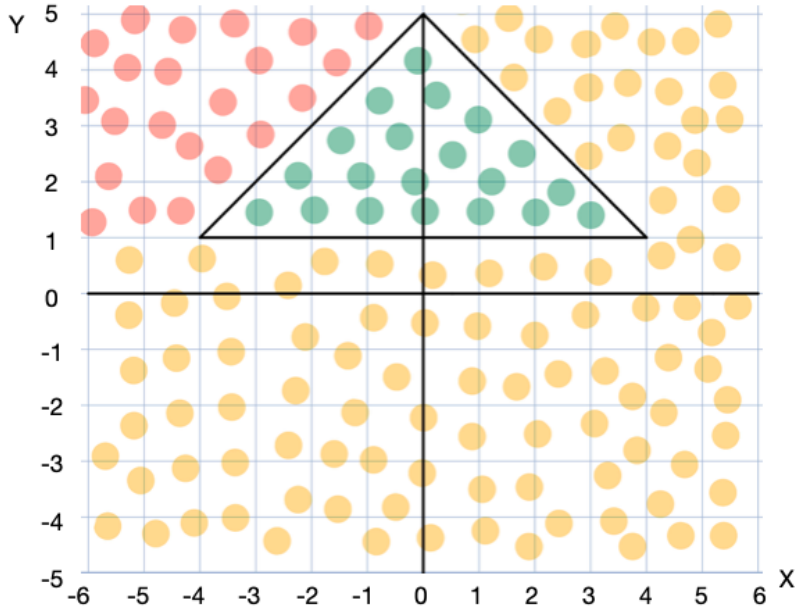
OUT_A =

OUT_B =

You may use **OUT_A** and **OUT_B** in your answer for **OUT_C**.

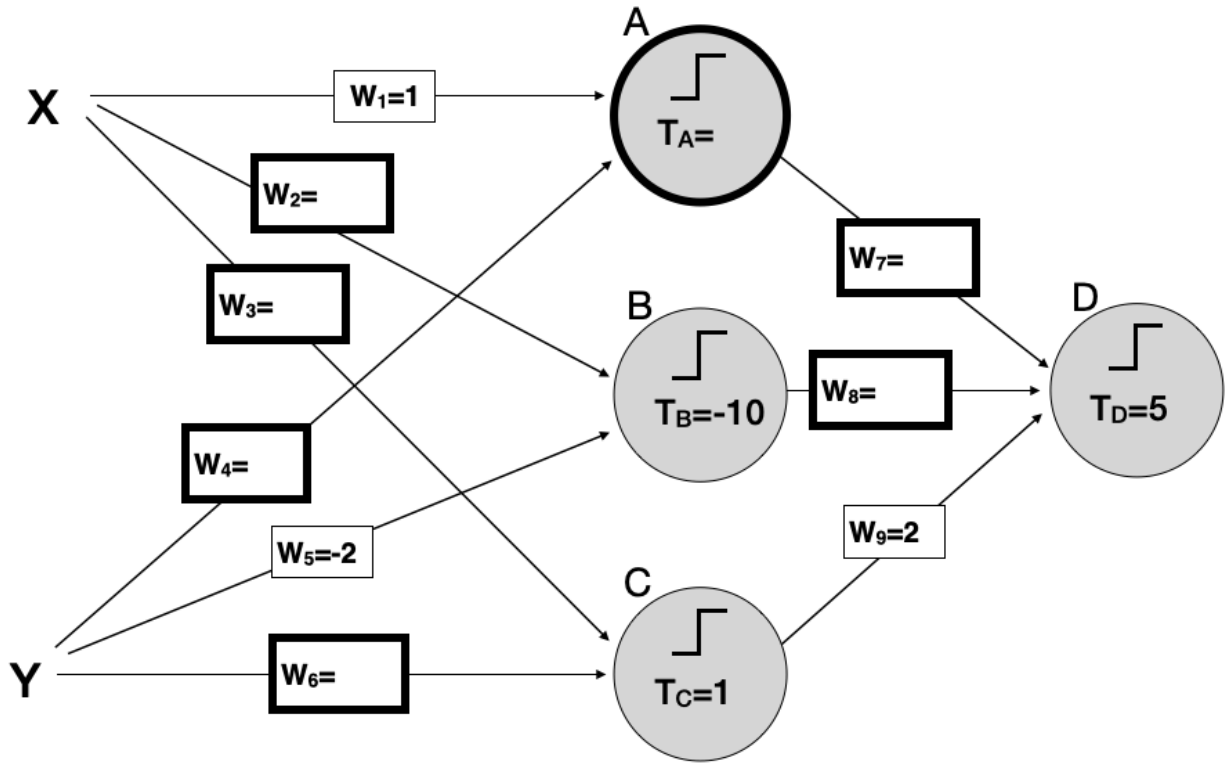
OUT_C =

A map of Mr. Ericford's room is shown below, centered around (0, 0). Mr. Ericford is trying identify where he has a Wi-Fi connection in his room so he can attend Zoom lecture. The green spots have a Wi-Fi connection, while the yellow and red spots do not have a Wi-Fi connection.



He makes a neural net and assigns it some weights and thresholds. The network should output a 1 where there is a Wi-Fi connection, and a 0 where there is no connection.

7.2 (7 points) In the neural net diagram below, fill in the **bolded** missing weights and missing threshold using integers, so that it correctly classifies all points in the graph. All neurons use the stairstep threshold function.



Show your work for partial credit.

7.3 (6 points) Mr. Ericford bought an additional router and placed it such that the red spots now have a Wi-Fi connection, so he needs to change the neural network. He wants to update the network to classify both the green spots and the red spots as a 1, and still classify the yellow dots as 0.

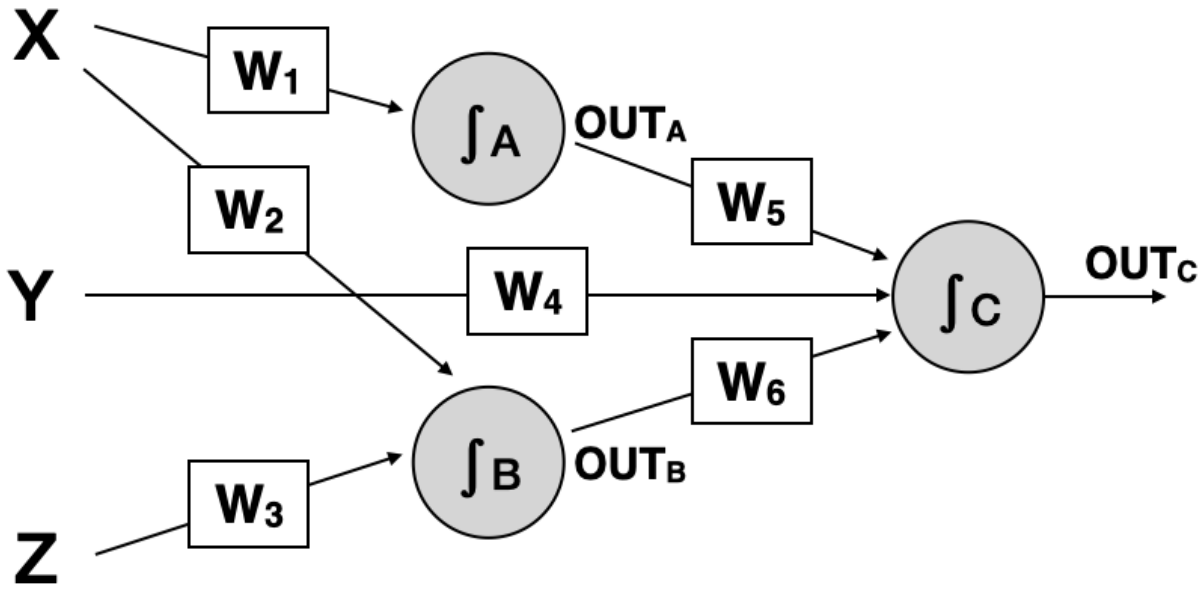
Is it possible to compute the new desired output by only deleting neurons and/or weights from his network? Circle **YES**, **NO**, or **CAN'T TELL**.

YES **NO** **CAN'T TELL**

If yes, list the neurons and/or weights that should be deleted. If no, write NONE.

Part B: Backward Propagation (12 points)

Mr. Ericford's classmates told him that they want to see his smiling face at more lectures, so he needs to update the neural net he made to help decide whether or not to turn on his camera. His network will now use sigmoid activation functions for the neurons (depicted below). He will do a single step of backpropagation to update $\mathbf{W_2}$.



8.1 (4 points) List the inputs (**X, Y, Z**), outputs ($\mathbf{OUT_{A\text{-}C}}$), and weights ($\mathbf{W_{1\text{-}6}}$) from the network above that are needed in order to compute an update of weight $\mathbf{W_{2,\ new}}$.

8.2 (8 points) Calculate the weight update for $\mathbf{W_{2,\ new}}$, in terms of the inputs (**X, Y, Z**), outputs ($\mathbf{OUT_{A\text{-}C}}$), weights ($\mathbf{W_{1\text{-}6}}$), learning rate **r,** desired output **OUT*,** and $\mathbf{W_{2,\ old}}$.

Part C: Neural Concepts (8 points)

Mr. Ericford notices that in Zoom classes, it's hard to know when it's acceptable to leave the meeting so he starts designing a neural net that helps him decide whether or not he should wave goodbye one more time before leaving the Zoom meeting.

In designing his network, he encounters some problems along the way. He thinks they have to do with the learning rate, **r**, but has been skipping 6.034 recitations so he's unsure.

9.1 (4 points) For each of the following four problems he encounters, circle which change (increase **r,** decrease **r**) might fix his problem, or **None** if neither change will fix his problem.

1. His training is taking too long.	increase **r**	decrease **r**	None
2. His model's chosen weights are causing its neurons to output only zeroes.	increase **r**	decrease **r**	None
3. His weights are unstable.	increase **r**	decrease **r**	None
4. His model's chosen weights are converging on a local maximum.	increase **r**	decrease **r**	None

Brad is a logical guy sometimes.

9.2 (4 points) Circle *all* logic functions that can be computed by the neural network architecture below (the weights and thresholds can be any real numbers and can be different for each function). Circle **None** if none of the functions can be computed by this architecture.

AND(X, Y) **XOR(X, Y)** **OR(X, Y)** **NOT(X)** **None**