

CS 3346A / CS 9146A Assignment 3

Due: Monday, Dec. 7th, 2020 (midnight)

Total Weight: 15%; must be individual effort

Silent Policy: A silent policy will take effect 24 hours before this assignment is due, i.e. no question about this assignment will be answered, whether it is asked on the discussion board, via email or in person.

Submission

Please submit your final answers in a **PDF** file through the OWL.

Please note you will not receive grades if TA can not recognize your handwriting, please finalize your solutions clearly and neatly and transform it to be a PDF version.

Q1: [Probability] (20 pts; adapted from the textbook)

Given the full joint distribution shown in the below table, calculate the following

- $P(\text{toothache})$
- $P(\text{Cavity})$
- $P(\text{Toothache} \mid \text{cavity})$
- $P(\text{Cavity} \mid \text{toothache} \vee \text{catch})$

	toothache		~toothache	
	catch	~catch	catch	~catch
cavity	0.108	0.012	0.072	0.008
~cavity	0.016	0.064	0.144	0.576

*[Hint] The main point of this exercise is to understand the various notations of **uppercase versus lowercase variable names**.*

Q2: [Probability and Bayes rule] (10 pts)

Consider two medical tests, A and B, for a virus. Test A is 95% effective at recognizing the virus when it is present. But has a 10% false positive rate (indicating that the virus is present, when it is not). Test B is 90% effective at recognizing the virus, but has a 5% false positive rate. The two tests use independent methods of identifying the virus.

The virus is carried by 1% of all people. Say that a person is tested for the virus using only one of the tests, and that test comes back positive for carrying the virus. Which test returning positive is more indicative of someone really carrying the virus? Justify your answer mathematically.

Q3: [Probability and Bayes theorem] (20 pts)

After conducting a blood test, the doctor told you that you were tested positive for a fatal disease. Worse, the test is quite accurate: the probability of false-positive (one is tested positive without the disease) is 0.05, and the probability of false-negative (one is tested negative with the disease) is 0.02. Seeing that you are desperate, the doctor told you that overall the disease is rather rare, only 1 in 10,000 people.

- (a). What is the chance now that you have the disease?
- (b). Naturally, the doctor orders a retest on you. The result of the second independent test is still positive. What is now your chance of having the disease?

Q4: [Classifier] (60 pts)

Given the training data in the below table.

Home Owner	Marital Status	Job Experience (1-5)	Defaulted
Yes	Single	3	No
No	Married	4	No
No	Single	5	No
Yes	Married	4	No
No	Divorced	2	Yes
No	Married	4	No
Yes	Divorced	2	No
No	Married	3	Yes
No	Married	3	No
Yes	Single	2	Yes

Predict if Bob will default his loan or not

	Home Owner	Marital Status	Job Experience (1-5)
Bob	No	Married	3

- Using Naive Bayes Classifier
- Using Decision Tree Classifier
- Using KNN Classifier ($k = 3$)
- Run k-Means (with $k = 2$) on the whole training data using all attributes (Home Owner, Marital Status, Job Experience, and Defaulted) to see what two clusters would be produced.

Note: Use the algorithms above “by hand” (not by using any machine learning software). The purpose is to help you understand the algorithms well. You can use the principles of the algorithms and detailed calculations may not be needed. For example, in b above, if you can easily decide which attribute should have maximum information gain, you don’t need to use the entropy formula (but you can, if you wish). In d above, choose random starting points by yourself, and decide how to calculate the “distance” for each attribute - there is usually not “correct” way anyway. If there are ties, break the tie randomly. Write down your thinking process so we can see you do understand how the algorithms work.

Q5: [Deep Neural Network] (20 pts)

A deep neural network with given weights and biases is shown in the figure. The activation function is the ReLU function (see the insert; also video DNN_1 at about 28 min). When $(x_1, x_2) = (0, 1)$ and $(1, -1)$, what will the outputs (z_1, z_2) be, respectively? (Note: to show your work, also write down the hidden layer output (y_1, y_2) , and (y_3, y_4) in your solutions).

