Midterm Exam

Name:

Instructions: **Read all of these instructions before starting the exam:**

1. All of the answers submitted by you must be your work. If you have questions ask the instructor or the TA.
2. Submit online via Collab through the tests and quizzes page. You must also submit a PDF of responses with work shown as an attachment.
3. Write and sign the honor pledge.
4. Points for each question are listed in parentheses by each question. The exam is worth 100 points total.

Due Date: Friday, March 8 12:00 PM

Part 1 (50 Points) Short Answer

1. **[6 Points]** Suppose you have just designed three different types of machine learning systems that can predict whether a stock will fall or rise the next day given as input a set of attribute-value pairs. The systems are as follows:

   System 1- Instance-based methods such as k-NN
   System 2- Id3 Decision-tree
   System 3- Propositional-rule learning system

   Assume all systems have the same classification accuracy on test data. You are trying to convince a trader to buy your system. Which system will you recommend for each of the following criterion and why?

   (a) Ease of understanding

   (b) Speed of predictions

   (c) Easily update models with new data.

2. **[3 Points]** In inductive learning, often times, there will be multiple consistent hypotheses. Explain how we choose among these hypotheses.

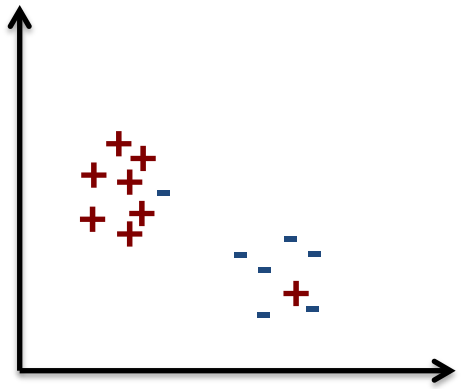
3. **[3 Points]** Explain why the size of the hypothesis space grows with increased number of attributes and unique values. Describe in general terms how the number of possible instances and hypotheses grow with the addition of a new attribute A that takes on k possible values.

4. **[3 Points]** What is bias in machine learning and is it possible to have an unbiased learner?

5. **[3 Points]** Suppose you have a noisy data set. What would happen if the candidate elimination algorithm was used? What if you used decision tree learning? Explain.

6. **[3 Points]** Some patient attributes are expensive to collect (e.g. MRI) while others are not (e.g. blood pressure, temperature). Therefore, we have decided to first ask our classification algorithm to predict whether a patient has a disease with some confidence and only then would pursue additional examinations. In this case, which classification method(s) do you recommend and why?

7. **[2 Points]** If you wanted to *perfectly* classify the data in the figures which method(s) could you use?



8. **[6 Points]** Suppose you have the following mystery data set:

A	B	C	Class
1	0	1	1
1	1	0	0
0	0	0	0
0	1	1	1
1	0	1	1
0	0	1	0
0	1	1	1
1	1	1	0

Draw the decision tree that would be learned from this data using ID3.

9. **[6 Points]** Suppose you have another set of mystery data shown below. We get a new case, A=0, B=0, and C=1.

A	B	C	Class
0	0	0	1
1	0	0	1
1	1	0	1
0	1	1	1
1	0	1	0
1	1	1	0

(a) What is the most likely class of the case using naïve Bayes (assume no smoothing)?

(b) Now assume Laplace smoothing with strength k=1 on both prior and class conditional parameters. Now, how will the instance be classified?

10. **[10 Points]** Consider the following case:

Bill goes to the hospital because he is feeling ill. The physician thinks that there are three probable illnesses (I_1, I_2, I_3), which are marginally independent from one another. In order to discover the most probable cause of Bill's illness, the physician will run 4 different tests (T_1, T_2, T_3, and T_4). The tests are conditionally dependent on the three illnesses as follows: T1 depends only on I_1, T_2 depends on I_1 and I_2, T_3 depends on I_1 and I_3, and T_4 depends only on I_3. Assume that all variables are either 0 or 1.

(a) Draw the Bayesian network for this problem.

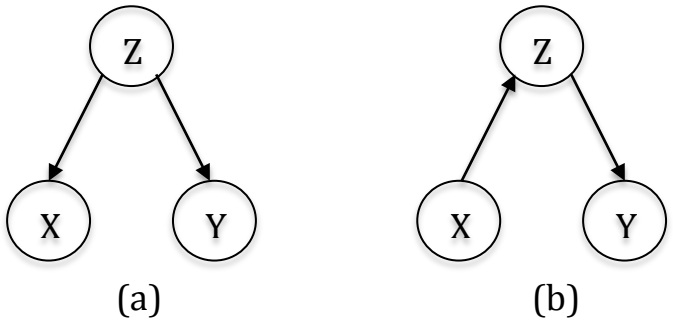
(b) What is the joint probability distribution as the product of conditional probabilities?

(c) How many independent parameters are needed to describe the joint distribution?

(d) What if there was no conditional independence between the variables, how many independent parameters would be needed to represent the joint distribution?

(e) What is the Markov Blanket of T_2?

11. **[6 Points]** Consider the following Bayesian networks. Both of the networks describe a joint probability distribution for P (X, Y, Z).



(a) P (X, Y, Z) = _____

(b) P (X, Y, Z) = _____

Part 2 (50 Points) Experiments

1. **[20 Points]** Answer the following questions comparing Naïve Bayes and k-NN algorithms.

(a) Compare Naïve Bayes and k-NN (use iB1) on the spambase.arff dataset using the train set option. Report your results. Which algorithm performs better and why?

Algorithm	% (training)
Nearest Neighbor	
Naïve Bayes	

(b) Repeat part (a) using 10-fold cross-validation as the test option. Report your results. What changes and why?

Algorithm	% (10-fold CV)
Nearest Neighbor	
Naïve Bayes	

2. **[30 Points]** Examine the effect of 2 parameters in the k-NN algorithm, the **optimal *k* value** and the **distance weighting function** using the spambase.arff dataset.

Type of attribute	Weighting Function	Model 3 No attr selection #attr=____		Model 4 attr selection #attr=____	
		k	%	k	%
Nominal attributes	None				
	1/d				
	1-d				
Numeric attributes	Weighting Function	Model 1 No attr selection ##attr=____		Model 2 attr selection #attr=____	
		k	%	k	%
	None				
	1/d				
	1-d				

(a) Select the k-NN (iBk) algorithm with KNN=30 and cross-validation=True. Run the algorithm 3 times with a different *distanceWeighting* function and report the best k and the classification accuracy. Perform 4 different experiments with the following options.

Model 1 Numeric attributes with default options in Weka (no pre-processing)

Model 2 Numeric attributes with supervised attribute selection

On preprocessing tab select filter→choose→filter→supervised→attribute→attributeSelection

Model 3 Nominal attributes with no attribute selection.
On preprocessing tab select filter→choose→filter→supervised→attribute→Discretize

Model 4 Nominal attributes with attribute selection and discretization
On preprocessing tab select filter→choose→filter→supervised→attribute→Discretize
On preprocessing tab select filter→choose→filter→supervised→attribute→attributeSelection

**Make sure you hit apply after choosing the filters. For model 4, it must be done twice.

Report your results in the table above including the number of attributes for

(b) Analyze your results. Describe the effect of the distance weighting function on the relative classification accuracy and the difference between numeric and nominal attributes. Also, analyze results with respect to *k*.