

Homework Problem Set 4: Rule-sets, Bayes, Neural Networks

Due Wednesday, March 1 at 11:59 PM

Upload a pdf to Canvas

Each question is worth the same number of points.

Question 1:

Consider the problem of predicting if a given person is a defaulted borrower (DB) based on the following attribute values:

1. Home Owner = Yes, No
2. Marital Status = Single, Married, Divorced
3. Annual Income = Low, Medium, High
4. Currently Employed = Yes, No

Suppose a rule-based classifier produces the following rules:

1. Home Owner = Yes \rightarrow DB = Yes
2. Marital Status = Single \rightarrow DB = Yes
3. Annual Income = Low \rightarrow DB = Yes
4. Annual Income = High, Currently Employed = No \rightarrow DB = Yes
5. Annual Income = Medium, Currently Employed = Yes \rightarrow DB = No
6. Home Owner = No, Marital Status = Married \rightarrow DB = No
7. Home Owner = No, Marital Status = Single \rightarrow DB = Yes

Answer the following questions. Make sure to provide a brief explanation or an example to illustrate the answer.

- (a) Are the rules mutually exclusive?
- (b) Is the rule set exhaustive?
- (c) Is ordering needed for this set of rules?
- (d) Do you need a default class for the rule set?

Question 3:

Consider the following set of rules based on the vertebrate data set given in Table 3.2 in the textbook (reproduced below):

Table 3.2. A sample data for the vertebrate classification problem.

Vertebrate Name	Body Temperature	Skin Cover	Gives Birth	Aquatic Creature	Aerial Creature	Has Legs	Hibernates	Class Label
human	warm-blooded	hair	yes	no	no	yes	no	mammal
python	cold-blooded	scales	no	no	no	no	yes	reptile
salmon	cold-blooded	scales	no	yes	no	no	no	fish
whale	warm-blooded	hair	yes	yes	no	no	no	mammal
frog	cold-blooded	none	no	semi	no	yes	yes	amphibian
komodo dragon	cold-blooded	scales	no	no	no	yes	no	reptile
bat	warm-blooded	hair	yes	no	yes	yes	yes	mammal
pigeon	warm-blooded	feathers	no	no	yes	yes	no	bird
cat	warm-blooded	fur	yes	no	no	yes	no	mammal
leopard	cold-blooded	scales	yes	yes	no	no	no	fish
shark								
turtle	cold-blooded	scales	no	semi	no	yes	no	reptile
penguin	warm-blooded	feathers	no	semi	no	yes	no	bird
porcupine	warm-blooded	quills	yes	no	no	yes	yes	mammal
eel	cold-blooded	scales	no	yes	no	no	no	fish
salamander	cold-blooded	none	no	semi	no	yes	yes	amphibian

R1: (Gives Birth = No, Aerial Creature = Yes) \implies Birds

R2: (Gives Birth = No, Aquatic Creature = Yes) \implies Fishes

R3: (Body Temperature = warm-blooded) \implies Mammals

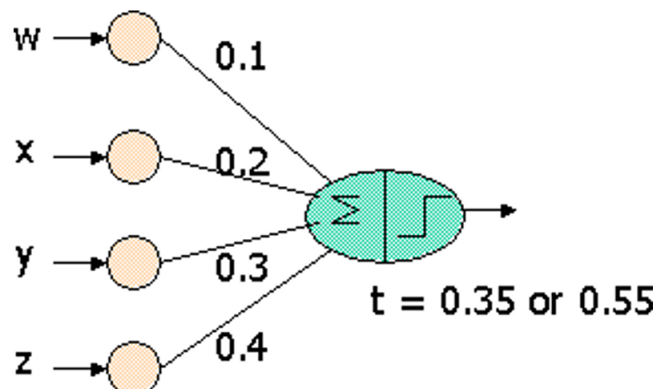
R4: (Gives Birth = No, Aerial Creature = No, Aquatic Creature = No) \implies Reptiles

For each of the following questions, provide a yes/no answer with a short explanation or a suitable example to justify your answer.

- Are the rules mutually exclusive?
- Are the rules exhaustive?
- Is ordering needed for this set of rules?

Question 4:

For the perceptron below:



What inputs will make the output 1 if the threshold t is 0.35, but NOT if the threshold is 0.55? After writing a Boolean expression in terms of w, x, y , and z for the set of possible inputs, identify one of those inputs in the lists below.

- a) $w = 0; x = 0; y = 1; z = 0$
- b) $w = 0; x = 1; y = 0; z = 0$
- c) $w = 0; x = 1; y = 1; z = 0$
- d) $w = 0; x = 1; y = 0; z = 1$
- e) $w = 1; x = 1; y = 0; z = 1$
- f) $w = 1; x = 1; y = 1; z = 0$

Question 5:

You are given a classification dataset with 100 instances, which has been partitioned into two subsets, dataset A with 50 instances and dataset B with 50 instances. Dataset A is used for training and dataset B is used for testing. You are supposed to compare two classification models: Model 1, which is an unpruned decision tree, and Model 2, which is a pruned version of the decision tree. The accuracy of the two classification models on datasets A and B are shown in the table below.

Classification Accuracy	Dataset A	Dataset B
Model 1	0.98	0.72
Model 2	0.82	0.8

- (a) Based on the accuracies shown in the table above, which classification model would you expect to have better performance on unseen instances? Support your answer with a brief explanation.
- (b) Now, you tested Model 1 and Model 2 on the entire dataset (A + B) and found that the classification accuracy of Model 1 on dataset (A + B) is 0.85, whereas the classification accuracy of Model 2 on the dataset (A + B) is 0.81. Based on this new information and your observations from the table above, which classification model would you finally choose for classification? Provide a brief explanation.

Question 6:

Both Minimum Description Length (MDL) and the pessimistic error estimate are techniques used for incorporating model complexity into the loss function. State one similarity and one difference between them in the context of decision trees. If you do not know the answer, use the web to do some research. You can also ask ChatGPT, but if you do, you must state so in your report. You will lose points if chatGPT provides the wrong answer. You will get a score of zero if you use chatGPT and do not state so.

Question 7:

State whether the following statements are True or False, giving a one sentence justification. Read the book chapter on ANN to retrieve the answer if you do not know it. You will be graded on the justification. Review how to compute the weight update based on Stochastic Gradient Descent to help answer the question. Do NOT use chatGPT or similar tool.

- a) In the back-propagation algorithm for training ANN models, the gradients of weights at the $k+1$ th layer can be computed using the gradients of weights at the k th layer.
- b) While applying an ANN model on a test instance, the activations at nodes at the $k+1$ th layer can be computed using the activations at nodes at the k th layer.
- c) An ANN training procedure is said to suffer from the vanishing gradient problem if the training errors vanish to zero while the test errors are still large.
- d) If at a given iteration of the back-propagation algorithm, the ANN model perfectly classifies all training instances, then the gradients of loss with respect to weights at all layers will be 0.

Question 8: Bayesian Classifiers

Consider a training set with 3 features, X_1 , X_2 , and X_3 , for a binary classification problem. The class distribution is shown in the table below.

X_1	X_2	X_3	Number of positive examples	Number of negative examples
1	1	1	20	8
1	0	0	20	17
0	1	0	5	8
0	0	0	5	17

- (a) Based on the information above, determine whether X_1 and X_2 are independent of each other.
- (b) Determine whether X_1 and X_2 are conditionally independent of each other given the class.
- (c) Compute the class conditional probabilities $P(X_1 = 1|+)$, $P(X_1 = 1|-)$, $P(X_2 = 1|+)$, $P(X_2 = 1|-)$, $P(X_3 = 1|+)$, and $P(X_3 = 1|-)$.
- (d) Use the class conditional probabilities given in the previous question to predict the class label of each example with the feature set given in the training set above. Use your results to compute the training error rate of the naïve Bayes classifier.

Question 9: Nearest Neighbor Classifier

For each of the two given scenarios, make a right choice of K for the KNN classifier in order to obtain better performance with a brief explanation.

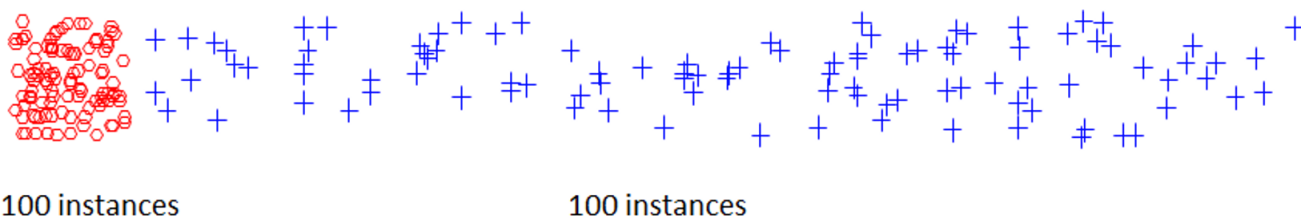


Figure KNN (a)

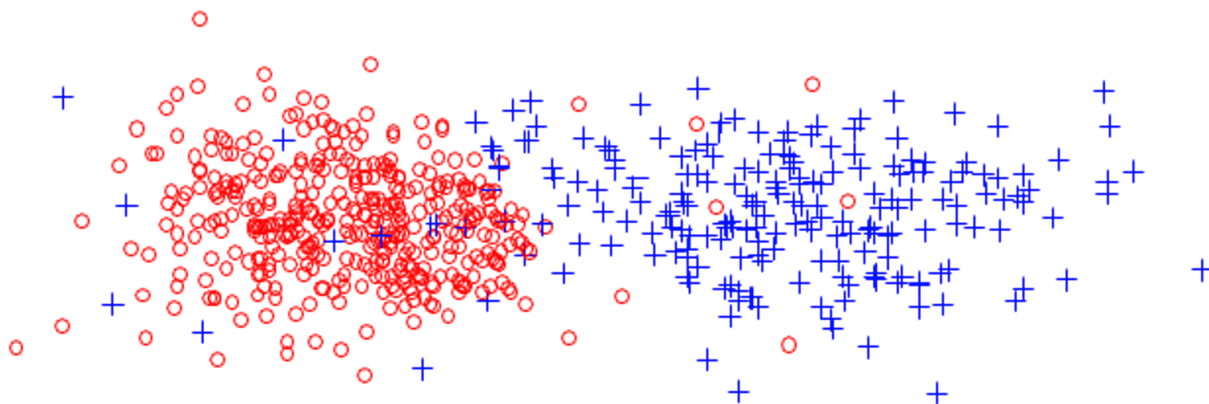


Figure KNN (b)

(a) $K = 1$ or $K = 5$ or $K = 50$?

(b) $K = 1$ or $K = 5$ or $K = 50$?

Question 10:

Consider the data set shown in the Table below.

Instance	A	B	C	Class
1	0	0	1	-
2	1	0	1	+
3	0	1	0	-
4	1	0	0	-
5	1	0	1	+
6	0	0	1	+
7	1	1	0	-
8	0	0	0	-
9	0	1	0	+
10	1	1	1	+

- Estimate the conditional probabilities for $P(A = 1|+)$, $P(B = 1|+)$, $P(C = 1|+)$, $P(A = 1|-)$. Explain how you arrive at the answers for the first case.
- Use the conditional probabilities in part (a) to predict the class label for a test sample $(A = 1, B = 1, C = 1)$ using the naïve Bayes approach.
- Compare $P(A = 1)$, $P(B = 1)$, and $P(A = 1, B = 1)$. State the relationships between A and B .
- Repeat the analysis in part (c) using $P(A = 1)$, $P(B = 0)$, and $P(A = 1, B = 0)$.
- Compare $P(A = 1, B = 1|Class = +)$ against $P(A = 1|Class = +)$ and $P(B = 1|Class = +)$. Are the variables conditionally independent given the class? Break down your derivations.