

CSCI 4380/6380 Data Mining

Assignment Number 2: Due 2/19/2026

The use of Generative AI is not Allowed

1. [25 points] Consider the following training set of samples for data mining:

Example	A1	A2	A3	A4	label
1	1	2	2	2	0
2	1	1	1	1	0
3	2	3	2	1	1
4	1	3	3	3	0
5	3	1	2	1	1
6	1	1	1	2	0

The attributes A1 through A4 are integers with values in the range [1,2,3] each.

- (a) Give a minimal size (measured by the total number of nodes) decision tree that can correctly classify all the training examples.
 - (b) How would the tree given in Part (a) above classify the following examples: (1,2,2,3) and (3,2,1,1)?
 - (c) Give three association rules consistent with this training set and specify the support and confidence for each rule.
2. [25 points] Short answers please
- (a) How can a decision tree be converted to a set of rules?
 - (b) How does Naive Bayes handle the missing value problem in training and in testing?
 - (c) How does the 1R method attempt to avoid over-fitting?
 - (d) Give one advantage to using Winnow learning over Perceptron learning of linear models.
 - (e) Give one advantage to using Perceptron learning over Winnow learning of linear models.
3. [25 points] Short answers please
- (a) Give a major difference between divide and conquer learning of decision trees and separate and conquer covering methods in data mining?
 - (b) Why does a problem occur in Naive Bayes if a particular attribute value does not occur in the training set in conjunction with every class value? Briefly describe a way to fix this problem.
 - (c) What is the difference between decision lists and general rule sets?
4. [25 points] Short answers please

- (a) Why are there usually many more association rules that can be inferred from a data-set than there are classification rules?
- (b) Give one advantage to using Logistic Regression over linear regression for classification.
- (c) Why is there a potential problem with highly branching attributes in decision tree learning? Briefly describe a method to overcome this problem.