Comp7103B: Second Assignment

Discussing with other students, the professor and the teaching assistant is allowed and encouraged, however, everybody should submit his/her own solutions. Students submitting the same solutions (or very similar) for one or more exercises will be penalized or might fail the assignment. Students should write their solutions in a short report (5 pages maximum) and submit it through the moodle website (pdf or scanned copy) together with the Python code. There is no limit on the length of the Python code.

1 Question 1 (Decision tree), 25%

Consider the data given in Table 1. a) Construct a decision tree using the following rules: (i) at each step compute the gini index for every possible split considering all attributes and select the split with the best gini value; (ii) Stopping rule: when the gini value of a node is zero or no further split is possible; and (iii) the class of a leaf node is determined by a majority rule (breaking ties arbitrarily). Include all the steps and calculations. b) Prune the decision tree you built so as to improve the generalization error (if possible). c) Predict the class value of (a='1',b='1') using the pruned decision tree (if it has been pruned).

a	b	Class
0	0	1
1	0	0
0	0	1
1	1	0
0	1	1
1	1	0
1	1	1
1	0	1

Table 1: Binary records with attributes a, b. Last attribute is the class (binary).

Assignment Project Quiz Exam Essay Help
WeChat: cestbon6888
Email: accoder_0verseas@163.com

2 Question 2 (NaiveBayes), 25%

Consider the data given in Table 2 from an employee database, where "Status" is the class attribute. Using the Naive-Bayes classifier, predict the class value of the record: ('marketing', 31-35,46k-50k). Show all the steps.

Department	Age	Salary	**Status**
marketing	26-30	46K-50K	junior
marketing	31-35	41K-45K	junior
secretary	46-50	36K-40K	senior
sales	26-30	25K-30K	junior
systems	41-45	66K-70K	senior
sales	31-35	46K-50K	junior
marketing	36-40	46K-50K	senior
systems	21-25	46K-50K	junior
systems	31-35	66K-70K	senior
secretary	26-30	25K-30K	junior
sales	31-35	31K-35K	junior

Table 2: Data specifying for each employee his/her department, age, salary, and whether he/she is senior or junior. Status is the class attribute.

3 Question 3 (hands-on on decision tree), 25%

Check the material provided to you, including in particular a text describing your tasks, a tutorial and a dataset.

4 Question 4 (hands-on on naive Bayes), 25%

Check the material provided to you, including in particular a text describing your tasks, a tutorial and a dataset.

Assignment Project Quiz Exam Essay Help
WeChat: cestbon688
Email: accoder_overseas@163.com