uOttawa

L'Université canadienne
Canada's university

Ottawa-Carleton Institute for Computer Science
School of Information Technology and Engineering
CSI 5387: Concept Learning and Data Mining
Final Examination Fall 2010

Instructor: Dr. Stan Matwin

Closed Text Exam; Time: 3 hrs. Total points: 100

Write all your answers in the exam booklet. Use one booklet for rough work, and the other for proper answers. Calculators are allowed (but NOT laptops).

Good luck, and have a nice Christmas!

Name:

Student #

Indicate your home university: OTTAWA / CARLETON

Q.	MAX	OBTAINED	Q.	MAX	OBTAINED
1a	4		6a	5	
1b	8		6b	3	
2	6		6c	4	
3	4		6d	4	
4a	5		7a	3	
4b	5		7b	5	
5a	6		7c	4	
5b	3		8a	3	
5c	5		8b	4	
5d	4		9a	3	
			9b	5	
			10a	3	
			10b	4	
TOTAL	100				

1. [Decision trees] You are given a dataset S describing performance of people taking a driving test. The class is PA (pass) or FA (fail).

Attributes are:

Three-point turn: C: correct, I: incorrect
Parking: S: successful U: unsuccessful, P: partly successful
Lane change: M: mirror consulted, N: mirror not consulted

Age: Young, Old
Sex: Male, Female
Violence in the movie : Yes/No

Three-point turn	Parking	Lane change	Pass/Fail
I	U	N	FA
C	S	N	FA
I	P	M	PA
C	S	M	PA
C	P	M	PA
I	S	N	FA
C	U	N	FA
C	P	N	FA
C	U	M	PA
I	S	M	PA

a) `(4pts)` give the value of info(S)
b) `(8pts)` which attribute will be chosen as the root? Why? (you can use the calculator here)

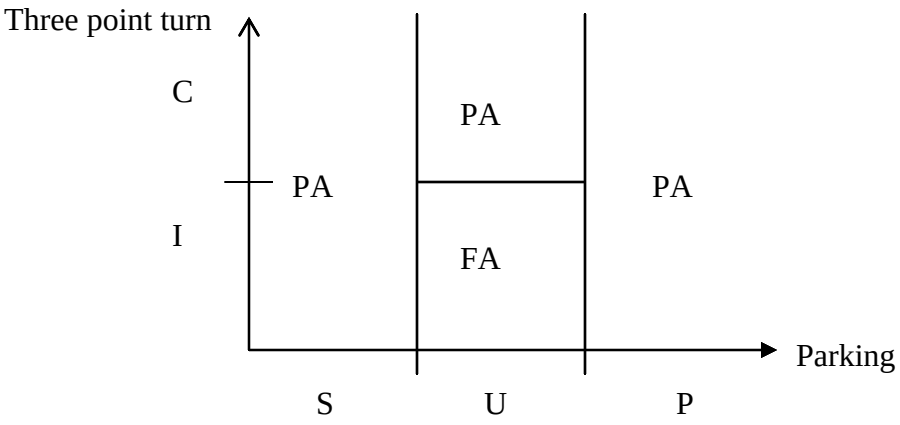
2. [Naïve Bayes] `(6pts)` Suppose the following new instance of the data from Question 1 is to be classified by an Naïve Bayes classifier:

Three-point turn	Parking	Lane change	Pass/Fail
I	P	N	?

The NB formula is $v_{NB} = \text{argmax}_{v_j \in V} P(v_j) \prod_i P(a_i | v_j)$

Show the calculations and the predicted class. If one of the probabilities is computed from the data as =0, use a simple form of Laplace smoothing: add 1 to the numerator and to the denominator of the fraction representing this conditional probability.

3. [Decision trees]. (4pts) Suppose the data in question (1) is reduced to two attributes, Three point turn and Parking (the same two classes). Given the following partition of the instance space, what tree does it result from?



Draw the tree. For tests, assume that if the test is true the right branch is followed.

4. [Naïve Bayes]

a) (5 pts) Characterizing NB geometrically we know it is a linear classifier. Show why. Hint: start with the formula

$$class(x) = \frac{P(class = 1 \mid x)}{P(class = 0 \mid x)} > \theta$$

b) (5 pts)show how an NB classifier can compute the confidence in its decision

5.[ROC]. Given a labeled data set and a scoring function (that could be, e.g., be obtained from a Bayesian classifier):

instance	score	label
A	0.8	+
B	0.7	+
C	0.6	-
D	0.4	+
E	0.3	-
F	0.2	-

a) (6 pts)show in the ROC space given below SEVEN classifiers (numbered 1-7) obtained as follows: (i) two trivial classifiers – always positive, always negative (ii) five other classifiers obtained by generating the prediction with a threshold between each two scores – e.g. a threshold = 0.5 will classify A, B, and C as positive, D, E, and F as negative.

b) (3 pts) draw a convex ROC curve

c)(5 pts) assume the classifier will be deployed in the environment in which pos:neg ratio is expected to be 2:1. Which classifier will you use? Why?

d)(4pts) For the classifier identified in c) above, what will be its expected accuracy? Justify.

6. [SVM] Φ: $R^2 \rightarrow R^3$ is the mapping from the input space (D = R^2) to the feature space (FS = R^3) defined as follows

$$(x_1, x_2) \rightarrow (z_1, z_2, z_3) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)$$

a) (5 pts) show that Φ above defines a polynomial kernel of the second degree 2.

b) (3 pts)Draw Pi, Ni, i=1…4 in the input space, (indicating their coordinates).

c) (4 pts) Draw Φ(Pi), Φ(Ni), i=1…4 in the feature space (indicating their coordinates).

d) (4 pts) Draw the largest margin hyperplane in the feature space (just a line will do so as not to complicate the drawing).

7. [Association rules – Apriori]. The following table database of 10 transactions is given:

Assume the following database of transactions is given (I symbol is omitted in from of the item symbols for simplicity):

T00 1 2
T10 2 3 4 5
T20 1 3 5
T30 2 5
T40 1 3 4 5
T50 2 4

T60 1 4
T70 4 5

a) (3 pts) give all frequent 2-item sets with support ≥ 0.3

b) (5 pts) give all frequent 3-item sets with support ≥ 0.3

b) (4 pts) from the item sets identified in c), give an association rule with confidence > 0.7.

In c) show how you obtained the rule and the confidence. In (a) and (b), show which candidate itemsets are rejected due to the monotonicity, and which due to insufficient support.

8 . Suppose you have a dataset with 3000 attributes. You are building a classifier, and you need to work with at most 100 attributes (e.g. you are building a decision tree on a laptop with limited memory).

a) (5 pts) name one specific preprocessing method (algorithm) that you will use

b) (4 pts) explain what will be the effect of that method on

(i) bias
(ii) variance

of this learning task.

9. a) (3 pts) Define the VC dimension

b)(5 pts) Consider a hypothesis language *H*. Show that the VC dimension of learning a hypothesis from *H* grows with the size of *H*.

10. Recall the Multinomial Naïve Bayes (MNB) formula: $$P(c \mid d) = \frac{P(c)\prod_{i=1}^{n} P(w_i \mid c)^{f_i}}{P(d)}$$

where c is a class, d is a document, and f_i is the frequency of occurrence of the word w_i in d .

a) (3 pts) what is the source of efficiency of the MNB algorithm?

b) (4 pts) how is the value of $P(w_i|c)$ estimated? (remember that MNB is a generative method)