Data Mining and Web Algorithms
15B22CI621

Tutorial 7

Q1: Consider the following data set:

Color	Size	Shape	Edible?
Yellow	Small	Round	+
Yellow	Small	Round	-
Green	Small	Irregular	+
Green	Large	Irregular	-
Yellow	Large	Round	+
Yellow	Small	Round	+
Yellow	Small	Round	+
Yellow	Small	Round	+
Green	Small	Round	-
Yellow	Large	Round	-
Yellow	Large	Round	+
Yellow	Large	Round	-
Yellow	Large	Round	-
Yellow	Large	Round	-
Yellow	Small	Irregular	+
Yellow	Large	Irregular	+

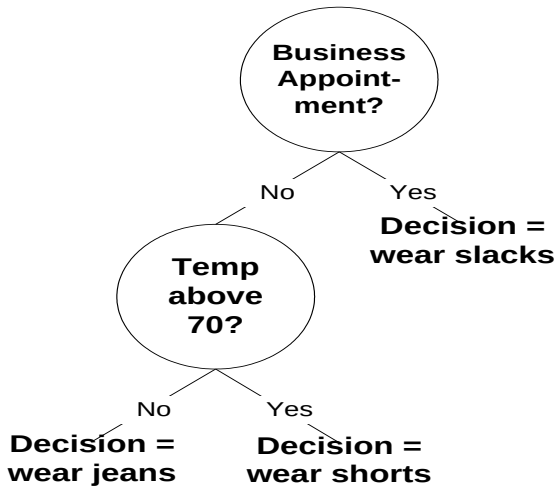
(a) What is the entropy of this collection of training data?

(b) What is the best split? (Among color, size and shape attributes) according to the information gain using entropy)

(c) What is the best split (between color and size) according to the classification error rate?

(d) Built a decision tree using ID3 Algorithm upto 2 level.

Q2: Which of the following is a valid production rule for the decision tree below?



a. IF Business Appointment = No & Temp above 70 = No

THEN Decision = wear slacks

b. IF Business Appointment = Yes & Temp above 70 = Yes
THEN Decision = wear shorts

c. IF Temp above 70 = No
THEN Decision = wear shorts

d. IF Business Appointment= No & Temp above 70 = No
THEN Decision = wear jeans

Q3. Consider the following data set and find the Gini of Past Trend and Open Interest.

Past Trend	Open Interest	Trading Volume	Return
Positive	Low	High	Up
Negative	High	Low	Down
Positive	Low	High	Up
Positive	High	High	Up
Negative	Low	High	Down
Positive	Low	Low	Down
Negative	High	High	Down
Negative	Low	High	Down
Positive	Low	Low	Down
Positive	High	High	Up