Machine Learning

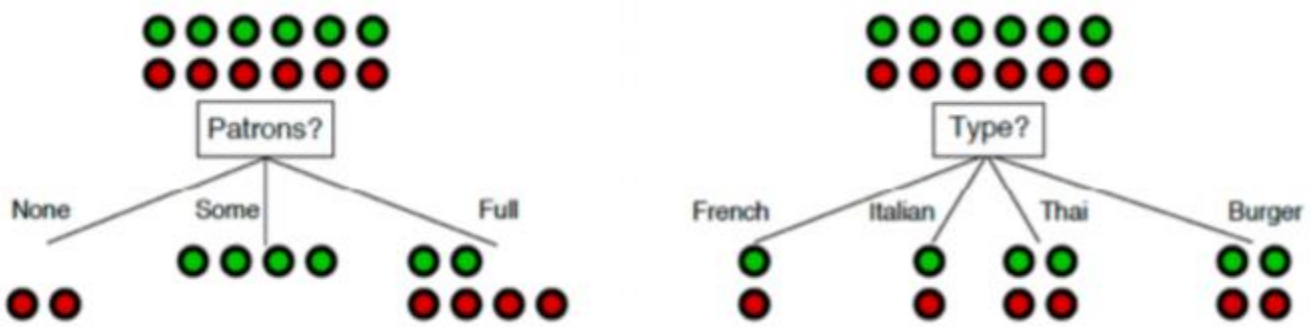
V. Adamchik | CSCI 567 | Fall 2019

Discussion Set 2 | University of Southern California

Decision Trees
Naïve Bayes

Problem 1

Compute Gini Index for both cases



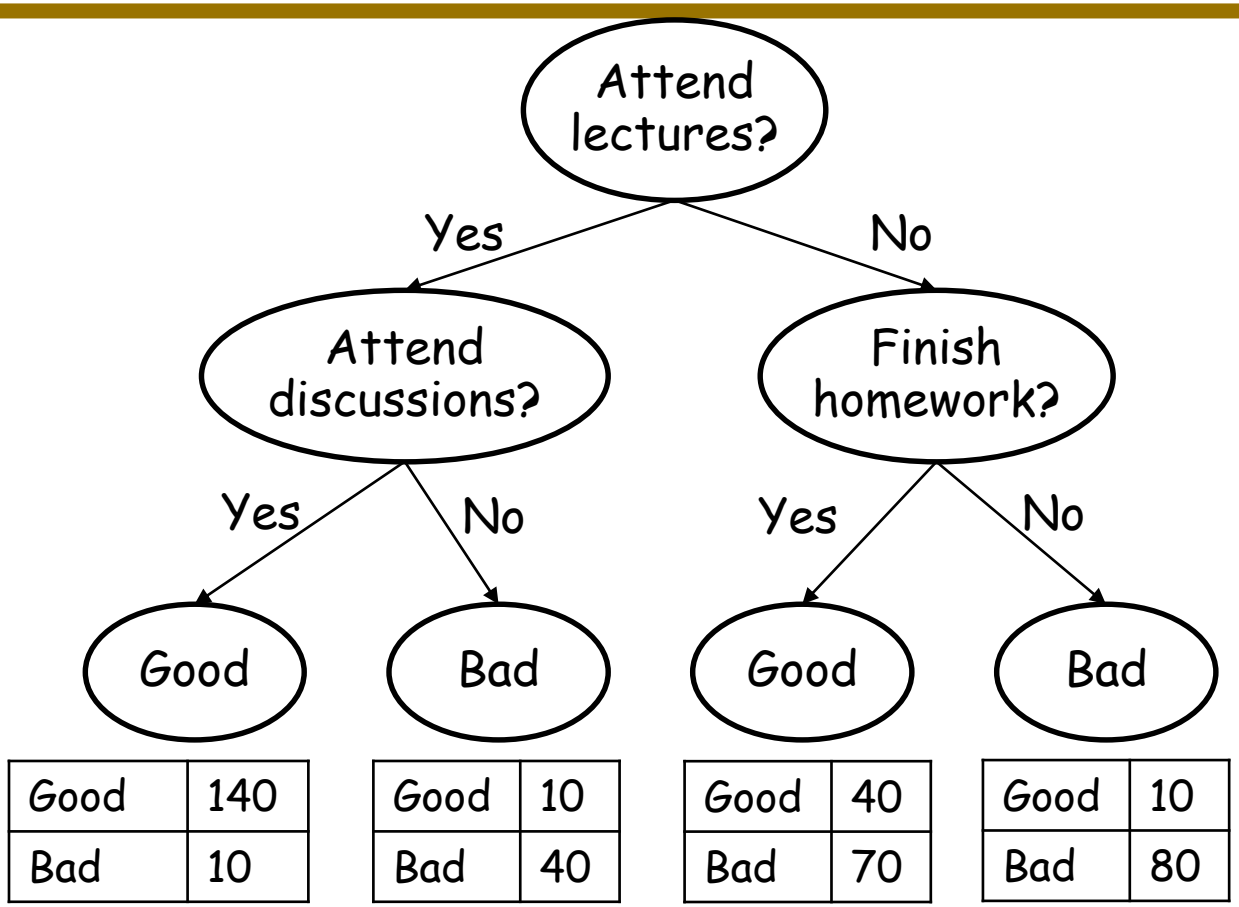
Problem 2

Name	Hair	Height	Weight	Lotion	Result
Sarah	blonde	average	light	no	sunburned
Dana	blonde	tall	average	yes	none
Alex	brown	short	average	yes	none
Annie	blonde	short	average	no	sunburned
Emily	red	average	heavy	no	sunburned
Pete	brown	tall	heavy	no	none
John	brown	average	heavy	no	none
Katie	blonde	short	light	yes	none

1) Which attribute will be best for the root of the tree?
Assume any attribute other than name.
Don't do any computation. Instead, try to make
an estimate.

2) Using the best test at the root, finish the tree.

Problem 3



Using REP, which subtree will you prune and why?

Problem 4

The Tough ML 567 is attended by students majoring in ML and some students that don't major in ML. In it only 50% of the ML students and 30% of the non-ML students pass the midterm exam. Unfortunately 60% of the entire class are non-ML students. What is the percentage of ML students among those that actually pass the exam.

Problem 5

Suppose you are given the following set of data with three Boolean input variables a, b, and c, and a single Boolean output variable K.

a	b	c	K
1	0	1	1
1	1	1	1
0	1	1	0
1	1	0	0
1	0	1	0
0	0	0	1
0	0	0	1
0	0	1	0

According to the naive Bayes classifier, what is a value of $P(K = 1|a = 1,b = 1,c = 0)$?

Problem 6

Example No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

Given training examples above, classify a Red Domestic SUV.