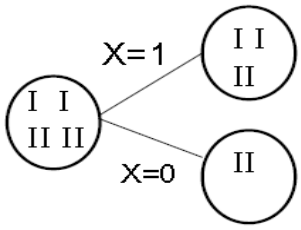
Question 2

If you have the following training set (3 features and 2 classes)

X	Y	Z	Class
1	1	1	I
1	1	0	I
0	0	1	II
1	0	0	II

How would you distinguish class I from class II?

1) Split on attribute X



If X is the best attribute, this node would be further split.

$E_{child1} = -(1/3)\log_2(1/3)-(2/3)\log_2(2/3)$
$= .5284 + .39$
$= .9184$

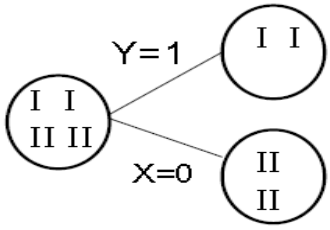
$E_{child2} = 0$

2 Grades

$E_{parent} = 1$

GAIN = 1 – (3/4)(.9184) – (1/4)(0) = .3112

2) Split on attribute Y



$E_{child1} = 0$

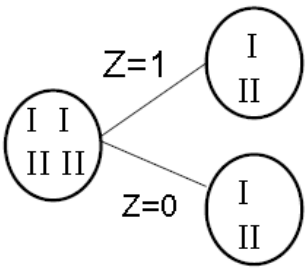
$E_{child2} = 0$

2 Grades

$E_{parent} = 1$

GAIN = 1 –(1/2) 0 – (1/2)0 = 1; BEST ONE

3) Split on attribute Z



$E_{child1} = 1$

$E_{child2} = 1$

2 Grades

$E_{parent} = 1$

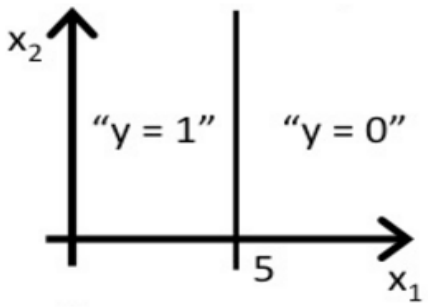
GAIN = 1 – (1/2)(1) – (1/2)(1) = 0 ie. NO GAIN; WORST

Question 3

Consider a logistic regression with two features ***x1*** and ***x2*** suppose $\theta_0 = 5$, $\theta_1 = -1$, $\theta_2 = 0$, write the hypothises function and draw the decision boundary of it ?

$$h_\theta(x) = g(5 - x_1)$$

1.5 Grades



1.5 Grades

Question 4

Compare between Supervised and unsupervised Learning techmiques

- Supervised learning
  - Learning from labelled data
  - Classification, Regression, Prediction, Function Approximation

1 Grade

- Unsupervised learning
  - Learning from unlabelled data
  - Clustering, Visualization, Dimensionality Reduction

1 Grade