

National University of Computer and Emerging Sciences, Lahore Campus



Course: Artificial Intelligence
Program: BS(Computer Science)
Duration: 20 min
Paper Date: 21-03-17
Section:
Exam: Quiz 3

Course Code: CS401
Semester: Fall 2016
Total Marks: 10
Weight: 2%
Page(s):
Reg. No

Instruction/Notes:

Given the training data what attribute will be chosen as root node? Using Gini as impurity measure.

Training Data

Instance #	Att1	Att2	Att3	Class
1	1	2	2	1
2	2	2	2	1
3	1	2	2	1
4	1	2	2	1
5	1	1	2	2
6	2	1	2	2
7	1	1	2	2
8	2	1	2	2

$$\text{Gini}(t) = 1 - \sum_{i=0}^{c-1} [p(i|t)]^2, \quad (4.4)$$

Where C is number of classes, and P(i|t) is fraction of records that belong to class i at node t.

Question 2 Classify the following test instance using kNN and training data given in above table. Use K=3

2	1	2	2
---	---	---	---

Note: All the attributes are binary and symmetric

Attr 1

x_1	$c_1: 3$ $c_2: 2$	x_2	$c_1: 1$ $c_2: 2$
-------	----------------------	-------	----------------------

Attr 2

y_1	$c_1: 4$ $c_2: 0$	y_2	$c_1: 4$ $c_2: 0$
-------	----------------------	-------	----------------------

Attr 3

z_1	$c_1: 0$ $c_2: 4$	z_2	$c_1: 4$ $c_2: 0$
-------	----------------------	-------	----------------------

$Gini(Attr 1) = 0.5$
 $Gini(x_1) = 1 - \left(\frac{3}{5}\right)^2 - \left(\frac{2}{5}\right)^2 = x$
 $Gini(x_2) = 1 - \left(\frac{1}{3}\right)^2 - \left(\frac{2}{3}\right)^2 = y$
 $Gain = 0.5 - \frac{5}{8}x - \frac{3}{8}y$

$Gini(y_1) = 1 - \left(\frac{4}{4}\right)^2 - \left(\frac{0}{4}\right)^2 = 0$
 $Gini(y_2) = 1 - \left(\frac{4}{4}\right)^2 - \left(\frac{0}{4}\right)^2 = 0$
 $Gain(Attr 2) = 0.5 - \frac{4}{8}(0) - \frac{0}{8}(0) = 0.5$

$Gini(z_1) = 1 - \left(\frac{0}{4}\right)^2 - \left(\frac{4}{4}\right)^2 = \text{undefined}$
 $Gini(z_2) = 1 - \left(\frac{4}{8}\right)^2 - \left(\frac{4}{8}\right)^2 = 0.5$
 $Gain(Attr 3) = \text{undefined}$

as $x, y > 0$ so $gain(Attr 2)$ is highest so it will be chosen as root node

Training Data

Instance #	Att1	Att2	Att3	Class
1	1	2	2	1
2	2	2	2	1
3	1	2	2	1
4	1	2	2	1
5	1	1	2	2
6	2	1	2	2
7	1	1	2	2
8	2	1	2	2

Q2

Similarity $\times 3$

1
2
1
1
2
3
2
3

SMC similarity of test instance with all training instances.
 2 is the most frequent class of k nearest neighbors
 so we will assign 2 to test instance

$Gini(t) = 1 - \sum_{i=1}^{c-1} [p(i|t)]^2$ (4.4)

Where C is number of classes, and $P(i|t)$ is fraction of records that belong to class i at node t .