Pattern and Speech Recognition WS2015-16
Exercise 7

Atanas Poibrenski(2554135), Marimuthu Kalimuthu(2557695), Furkat Kochkarov(2557017)

January 4, 2016

Decision Trees

Exercise 1

1 Done.

2 See “decision_tree.m”

3 Histogram plot:

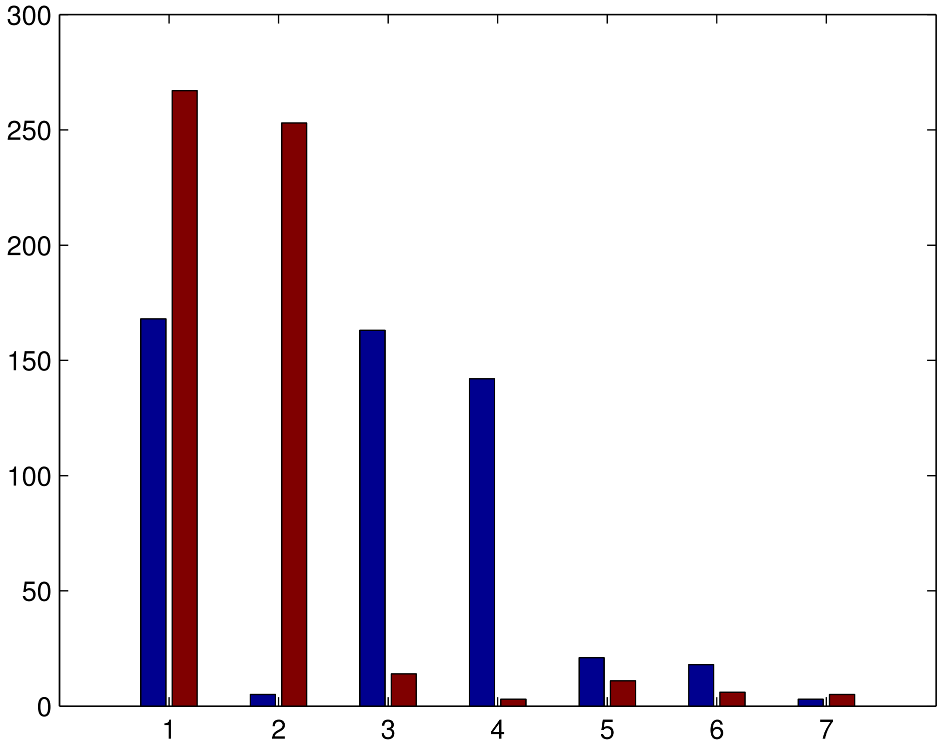

Figure 1: Histogram for two class labels (blue=republican, red=democrat)

(From the histogram), the highest probability class at every node are as follows:

$$\begin{aligned}
node1 &= democrat\\
node2 &= democrat\\
node3 &= republican\\
node4 &= republican\\
node5 &= republican\\
node6 &= republican\\
node7 &= democrat
\end{aligned}$$

The histogram tells us that the quality of the questions is good since the splits result in low misclassification rates and are almost pure.

4 Misclassification Rates:

$$\begin{aligned}
node1 &= 0.38\\
node2 &= 0.019\\
node3 &= 0.07\\
node4 &= 0.02\\
node5 &= 0.34\\
node6 &= 0.25\\
node7 &= 0.375
\end{aligned}$$

5 Entropy of the nodes:

$$\begin{aligned}
node1 &= 0.96\\
node2 &= 0.137\\
node3 &= 0.398\\
node4 &= 0.145\\
node5 &= 0.92\\
node6 &= 0.811\\
node7 &= 0.95
\end{aligned}$$

6 Information Gain:

$$\begin{aligned}
node1 &= 0.96 - (258/435 * 0.137 + 167/435 * 0.398) = 0.72\\
node2 &= 0.398 - (145/177 * 0.145 + 32/177 * 0.92) = 0.11\\
node3 &= 0.92 - (24/32 * 0.811 + 8/32 * 0.95) = 0.07
\end{aligned}$$

7 Why Information Gain?
The information gain is a good measure because for a given node, maximizing the information gain is equivalent to minimizing the entropy of its children. Minimizing the entropy of

the children will result in more "purity", therefore less classification error.

The question about *anti-satellite-test-ban* will maximize the information gain. This question will result in *pure splits.*

8 Algorithm to construct complete decision tree:
* Start with the *root node.*
* For all the possible questions to ask, choose the one which results into the highest possible information gain.
* Repeat this for all the nodes.
* If a splitting of a node results into a *pure split*, stop splitting on that node.

Exercise-2

1 Prior probabilities of the classes:

republican = 0.38
democrat = 0.62

2 Conditional Probabilities:

P(x ["education-spending"] = no | republican) = 0.11
P(x ["education-spending"] = no | democrat) = 0.79

3 Highest probability class:
The *democrat* class has the highest probability given that *x["education-spending] = no*