Fall 2019

CISC 483-683 Assignment 2 Answers

1. a) Entropy overall $= -\frac{15}{22}\log\left(\frac{15}{22}\right) - \frac{5}{22}\log\left(\frac{5}{22}\right) - \frac{2}{22}\log\frac{2}{22}$
$= 1.177$

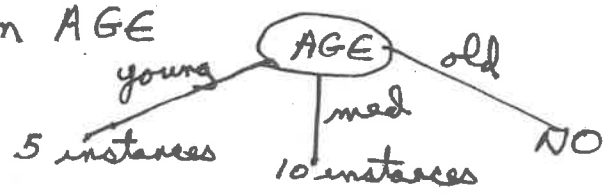
Info Gain(Split $_{AGE}$) $= 1.177 - \frac{7}{22}$(entropy old) $- \frac{10}{22}$(entropy med) $- \frac{5}{22}$(entropy you
$= 1.177 - 7/22(0) - \frac{10}{22}(1) - \frac{5}{22}(.971)$
$= .502$

Info Gain(Split $_{Hair}$) $= 1.177 - \frac{5}{22}$(entropy short) $- \frac{11}{22}$(entropy med)
$- \frac{4}{22}$(entropy long) $- \frac{2}{22}$(entropy horrid)
$= 1.177 - \frac{5}{22}(1.371) - \frac{11}{22}(.946) - \frac{4}{22}(.811) - \frac{2}{22}(0)$
$= .245$

Info Gain(Split $_{UPKEEP}$) $= 1.177 - \frac{4}{22}$(entropy low) $- \frac{6}{22}$(entropy med)
$- \frac{3}{22}$(entropy high) $- \frac{9}{22}$(ent tremendous)
$= 1.177 - \frac{4}{22}(.811) - \frac{6}{22}(1.46) - \frac{3}{22}(0) - \frac{9}{22}(.918)$
$= 1.177 - .921 = .256$

Info Gain(Split $_{SHOTS}$) $= 1.177 - \frac{5}{22}$(entropy all) $- \frac{11}{22}$(entropy some) $- \frac{6}{22}$(entropy none)
$= 1.177 - \frac{5}{22}(.971) - \frac{11}{22}(1.435) - \frac{5}{22}(0)$
$= 1.177 - .938 = .239$

So Split on AGE



Let's look at the young branch with 5 instances
Entropy = .971
Info Gain(Split Hair) $= .971 - \frac{3}{5}$(entropy short) $- 0$(entropy med)
$- \frac{2}{5}$(entropy long) $- 0$(entropy horrid)
$= .971 - \frac{3}{5}(.918) - \frac{2}{5}(1)$
$= .020$

Info Gain(Split UPKEEP) $= .971 - \frac{2}{5}$(entropy low) $- \frac{2}{5}$(entropy med)
$- \frac{1}{5}$(entropy high) $- 0$(entropy tremendous)
$= .971 - \frac{2}{5}(0) - \frac{2}{5}(0) - \frac{1}{5}(0) - 0$
$= .971$

Info Gain(Split SHOTS) $= .971 - \frac{3}{5}$(entropy none) $- \frac{2}{5}$(entropy some) $- 0$
$= .971 - \frac{3}{5}(0) - \frac{2}{5}(0) - 0$
$= .971$

So Split on either UPKEEP or SHOTS
I'll split on UPKEEP but you could split on SHOTS

Now look at the med branch with 10 instances

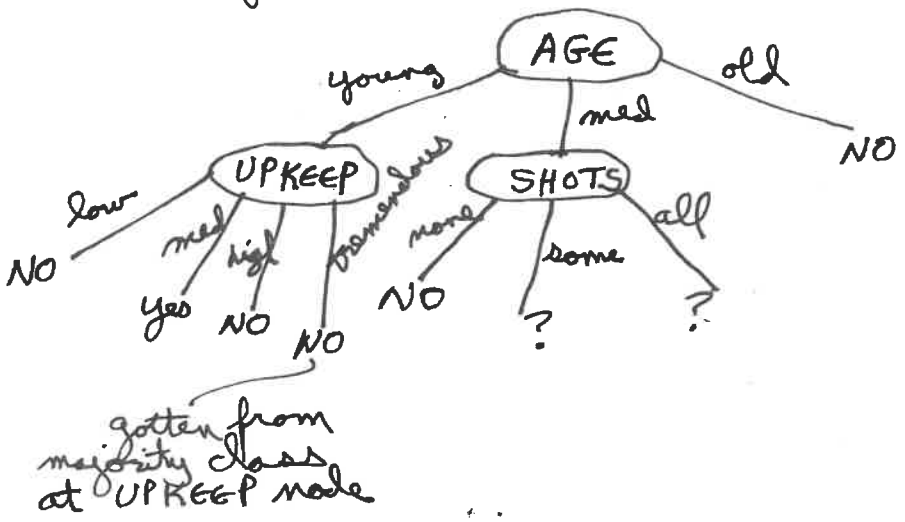
Entropy = .918

Info Gain(Split HAIR) = 1.0 − 2/10(entropy short) − 6/10(entropy med) − 0(entropy long) − 2/10(entropy horrid)
= 1.0 − 2/10(1) − 6/10(.918) − 0 − 2/10(0)
= 1 − .751
= .249

Info Gain(Split UPKEEP) = 1.0 − 1/10(entropy low) − 2/10(entropy med) − 0(entropy high) − 7/10(entropy tremendous)
= 1.0 − 1/10(0) − 2/10(1) − 7/10(.985)
= .110

Info Gain(Split SHOTS) = 1.0 − 3/10(entropy all) − 4/10(entropy some) − 3/10(entropy none)
= 1.0 − 3/10(.918) − 4/10(.811) − 3/10(0)
= .4

So split on SHOTS



Still need to work on the two ? above

Lets look at the branch labelled SHOTS = some

Entropy = .811

Info Gain(Split UPKEEP) = .811 − 1/4(entropy low) − 0(entropy med) − 0(entropy high) − 3/4(entropy trem)
= .811 − 0 − 0 − 0 − 3/4(.918)
= .122

Info Gain(Split HAIR) = .811 − 0(entropy short) − 3/4(entropy med) − 0(entropy high) − 1/4(entropy horrid)
= .811 − 0 − 3/4(0) − 0 − 1/4(0) = .811

So we split on HAIR

Now let's look at the branch labelled SHOTS = all

entropy = .918

Info Gain(Split UPKEEP) = .918 − 0(entropy low) − 1/3(entropy med) − 0(entropy high) − 2/3(entropy tremendous)

= .918 − 0 − 1/3(0) − 0 − 2/3(1)

= .251

Info Gain(Split HAIR) = .918 − 0 = .918

So split on HAIR

Tree is

1b) Accuracy on test data for the above tree is $9/11 \approx 82\%$

(the second and 10th instances are incorrectly classified)

TITLE OF PROJ OR STUDY _____ PROJ. OR STUDY NO. _____

SUBJECT _____ WORKS _____

COMPUTER _____ DATE _____ 20 _____

1c) Consider three possible binary splits on SHOTS

A. {none}, {some, all}

B. {some}, {none, all}

C. {all}, {none, some}

$$\text{Gini}(\{\text{none}\},\{\text{some, all}\}) = \frac{6}{22}(1-1) + \frac{16}{22}\left(1-\left(\frac{2}{16}\right)\left(\frac{2}{16}\right)-\left(\frac{5}{16}\right)\frac{5}{16}-\frac{9}{16}\cdot\frac{9}{16}\right)$$
$$= \frac{6}{22}(0) + \frac{16}{22}(1-.016-.098-.316)$$
$$= \frac{16}{22}(.570)$$
$$= .415$$

$$\text{Gini}(\{\text{some}\},\{\text{none, all}\} = \frac{11}{22}\left[1-\frac{2}{11}\cdot\frac{2}{11}-\frac{3}{11}\cdot\frac{3}{11}-\frac{6}{11}\cdot\frac{6}{11}\right]$$
$$+\frac{11}{22}\left[1-0-\frac{2}{11}\cdot\frac{2}{11}-\frac{9}{11}\cdot\frac{9}{11}\right]$$
$$= .297 + .149$$
$$= .446$$

$$\text{Gini}(\{\text{all}\},\{\text{some, none}\}) = \frac{5}{22}\left[1-0-\frac{2}{5}\cdot\frac{2}{5}-\frac{3}{5}\cdot\frac{3}{5}\right]$$
$$+\frac{17}{22}\left[1-\frac{2}{17}\cdot\frac{2}{17}-\frac{3}{17}\cdot\frac{3}{17}-\frac{12}{17}\cdot\frac{12}{17}\right]$$
$$= .109 + .353$$
$$= .462$$

So the best partition for SHOTS is {none}, {some, all}