National University of Singapore
School of Computing
CS3243: Foundations of Artificial Intelligence
Solutions for Tutorial 10

Readings: AIMA Chapter 18

1. (Question 18.4 from AIMA) We never test the same attribute twice along one path in a decision tree. Why not?

   Because for the attribute node that is lower down the tree, all the sample points will map to the same value.

2. Given the following sets of class values. For each set, calculate the Information Content (Entropy) given by: $I(\frac{p}{p+n}, \frac{n}{p+n}) = -\frac{p}{p+n} \log_2 \frac{p}{p+n} - \frac{n}{p+n} \log_2 \frac{n}{p+n}$. Explain your findings.

   (a) {p, p, p, p}

   $$I = -\frac{4}{4} log_2 \frac{4}{4} - \frac{0}{4} log_2 \frac{0}{4} = 0$$

   (b) {p, p, p, n}

   $$I = -\frac{3}{4} log_2 \frac{3}{4} - \frac{1}{4} log_2 \frac{1}{4} = 0.81$$

   (c) {p, p, n, n}

   $$I = -\frac{2}{4} log_2 \frac{2}{4} - \frac{2}{4} log_2 \frac{2}{4} = 1$$

   (d) {p, n, n, n}

   $$I = -\frac{1}{4} log_2 \frac{1}{4} - \frac{3}{4} log_2 \frac{3}{4} = 0.81$$

   (e) {n, n, n, n}

   $$I = -\frac{0}{4} log_2 \frac{0}{4} - \frac{4}{4} log_2 \frac{4}{4} = 0$$

   In the first and last data set, all elements of each data set belong to one class - In this data sets, knowing either data set will allow you to correctly predict a new item in the data set. This means no further information is needed in either case, and the value I of zero captures this.

   If the middle data set is all the only information you have about the distribution, then you cannot do better than random chance in predicting a new item. The value I of one captures this - You need a maximum amount of information to be able to better predict what each new item added to the data set would be.

   Values for all other data sets fall in between.

3. Given the following training data set about exotic dishes -We want to predict whether or not a dish is appealing given attributes 'Temperature', 'Taste', and 'Size'.

ID	Temperature	Taste	Size	Appealing
1	Hot	Salty	Small	No
2	Cold	Sweet	Large	No
3	Cold	Sweet	Large	No
4	Cold	Sour	Small	Yes
5	Hot	Sour	Small	Yes
6	Hot	Salty	Large	No
7	Hot	Sour	Large	Yes
8	Cold	Sweet	Small	Yes
9	Cold	Sweet	Small	Yes
10	Hot	Salty	Large	No

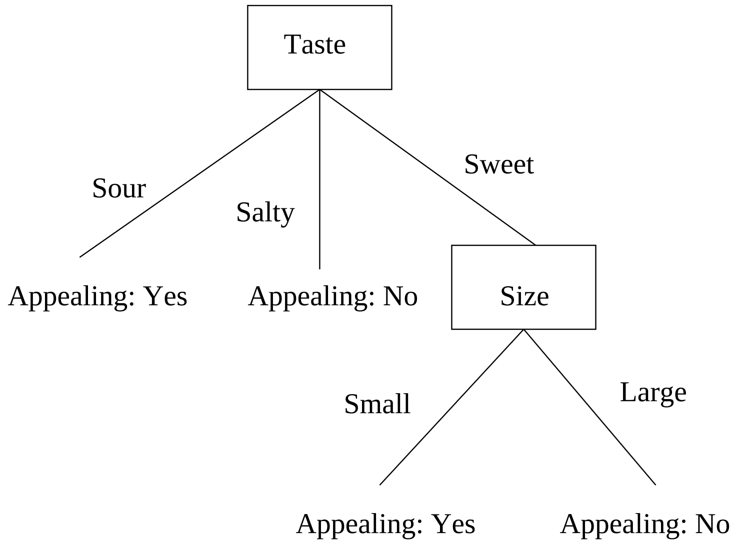
(a) What is the information gain $IG(A)$ associated with choosing the attribute 'Taste' as the root of the decision tree?

$$\begin{aligned} IG(A) &= I\left(\frac{p}{p+n}, \frac{n}{p+n}\right) - remainder(A) \\ remainder(A) &= \sum_{i=1}^{v} \frac{p_i + n_i}{p+n} I\left(\frac{p_i}{p_i+n_i}, \frac{n_i}{p_i+n_i}\right) \\ I\left(\frac{5}{10}, \frac{5}{10}\right) &= -\frac{5}{10}\log_2\frac{5}{10} - \frac{5}{10}\log_2\frac{5}{10} = 1 \end{aligned}$$

There are three branches 'Salty', 'Sweet' and 'Sour':
Salty - #Yes = 0, #No = 3, Total = 3
Sweet - #Yes = 2, #No = 2, Total = 4
Sour - #Yes = 3, #No = 0, Total = 3

$$\begin{aligned} remainder(Taste) &= \frac{3}{10}I\left(\frac{0}{3}, \frac{3}{3}\right) + \frac{4}{10}I\left(\frac{2}{4}, \frac{2}{4}\right) + \frac{3}{10}I\left(\frac{3}{3}, \frac{0}{3}\right) \\ &= \frac{4}{10} \times 1 = \frac{2}{5} \\ IG(Taste) &= I\left(\frac{5}{10}, \frac{5}{10}\right) - remainder(Taste) \\ &= 1 - \frac{2}{5} = \frac{3}{5} \end{aligned}$$

(b) Draw a decision tree with 'Taste' as the root.



(c) Use the decision tree to predict the class value for the record given by

ID	Temperature	Taste	Size
11	Hot	Salty	Small
12	Cold	Sweet	Large

ID 11 - Predicted Appealing = No
ID 12 - Predicted Appealing = No.

4. The loans department of a bank has the following past loan processing records each containing an applicant's income, credit history, debt, and the final approval decision. These records can serve as training examples to build a decision tree for a loan advisory system.

Income	Credit History	Debt	Decision
$0-5K$	Bad	Low	Reject
$0-5K$	Good	Low	Approve
$0-5K$	Unknown	High	Reject
$0-5K$	Unknown	Low	Approve
$0-5K$	Unknown	Low	Approve
$0-5K$	Unknown	Low	Reject
$5-10K$	Bad	High	Reject
$5-10K$	Good	High	Approve
$5-10K$	Unknown	High	Approve
$5-10K$	Unknown	Low	Approve
Over $10K$	Bad	Low	Reject
Over $10K$	Good	Low	Approve

(a) Construct a decision tree based on the above training examples. (Note: $\log_2 \frac{x}{y} = \log_2 x - \log_2 y$, $\log_2 1 = 0$, $\log_2 2 = 1$, $\log_2 3 = 1.585$, $\log_2 4 = 2$, $\log_2 5 = 2.322$, $\log_2 6 = 2.585$, $\log_2 7 = 2.807$, $\log_2 8 = 3$, $\log_2 9 = 3.170$, $\log_2 10 = 3.322$, $\log_2 11 = 3.459$, and $\log_2 12 = 3.585$)

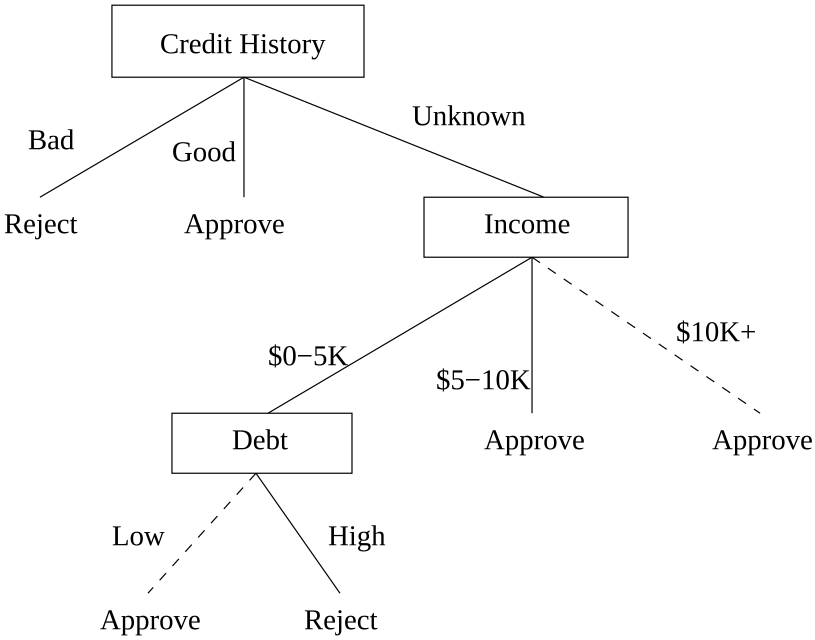
$$\begin{aligned}
IG(Income) &= I(\frac{5}{12}, \frac{7}{12}) - remainder(Income) \\
remainder(Income) &= -\frac{6}{12}(-\frac{3}{6}\log_2\frac{3}{6} - \frac{3}{6}\log_2\frac{3}{6}) - \frac{4}{12}(-\frac{1}{4}\log_2\frac{1}{4} \\
&\quad -\frac{3}{4}\log_2\frac{3}{4}) - \frac{2}{12}(-\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2}) \\
&= 0.937 \\
IG(Income) &= 0.98 - 0.937 = 0.043 \\
remainder(CreditHistory) &= -\frac{3}{12}(-\frac{3}{3}\log_2\frac{3}{3} - \frac{0}{3}\log_2\frac{0}{3}) - \frac{3}{12}(-\frac{3}{3}\log_2\frac{3}{3} \\
&\quad -\frac{0}{3}\log_2\frac{0}{3}) - \frac{6}{12}(-\frac{2}{6}\log_2\frac{2}{6} - \frac{4}{6}\log_2\frac{4}{6}) \\
&= 0.459 \\
IG(CreditHistory) &= 0.98 - 0.459 = 0.521 \\
remainder(Debt) &= -\frac{8}{12}(-\frac{3}{8}\log_2\frac{3}{8} - \frac{5}{8}\log_2\frac{5}{8}) - \frac{4}{12}(-\frac{2}{4}\log_2\frac{2}{4} - \frac{2}{4}\log_2\frac{2}{4}) \\
&= 0.970 \\
IG(Debt) &= 0.98 - 0.970 = 0.01
\end{aligned}$$

Since Credit History has the highest gain, choose it as the root, which has three values, i.e., "Bad", "Good", and "Unknown". Since all examples for "Bad" have the same classification (i.e., "Reject") and all examples for "Good" have the same classification (i.e., "Approve"), both nodes have no further subtree. For "Unknown", a subtree for the following subset of examples is to be constructed:

Income	Debt	Decision
$0-5K$	High	Reject
$0-5K$	Low	Approve
$0-5K$	Low	Approve
$0-5K$	Low	Reject
$5-10K$	High	Approve
$5-10K$	Low	Approve

$$\begin{aligned}
I(\frac{2}{6}, \frac{4}{6}) &= -\frac{2}{6}\log_2\frac{2}{6} - \frac{4}{6}\log_2\frac{4}{6} = 0.918 \\
remainder(Income) &= \frac{4}{6}(-\frac{2}{4}\log_2\frac{2}{4} - \frac{2}{4}\log_2\frac{2}{4}) + \frac{2}{6}(-\frac{2}{2}\log_2\frac{2}{2} - \frac{0}{2}\log_2\frac{0}{2}) = 0.667 \\
gain(Income) &= 0.918 - 0.667 = 0.251 \\
remainder(Debt) &= \frac{2}{6}(-\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2}) + \frac{4}{6}(-\frac{1}{4}\log_2\frac{1}{4} - \frac{3}{4}\log_2\frac{3}{4}) = 0.874 \\
gain(Debt) &= 0.918 - 0.874 = 0.044
\end{aligned}$$

Since Income has a higher gain than Debt, Income is chosen as the root of the subtree under Credit History=Unknown.



The two decisions on dotted lines are those that are not quite so straight forward. In one case, there is insufficient data for us to decide what to do, so we do something reasonable. In the other case, the resulting sample is not decisive. There are two approves and one reject. We can use a simple majority to decide.

Question is: why do we have such a node where there seems to be ambiguity? There is no clear answer (or at least we don't have the information to determine). One possibility is that we are missing some attribute from the data that will allow us to differentiate at this last node. Another possibility is that there is some non-determinism in the underlying function that we are trying to approximate with this decision tree.

(b) What is decision tree classifier's decision for a person who has 4K yearly income, a good credit history and a high amount of debt?

For the individual, the decision tree approves of the application from the root node of the tree.

(c) Use Naïve Bayes to calculate the decision. Does it differ from the decision tree classifier?

To calculate the decision for the given case, we have to find the unconditional probabilities of the two classes $P(Reject)$ *and* $P(Approve)$*, and the conditional probabilities of the features given the classes.*

$P(Approve) = \frac{7}{12}$
$P(Reject) = \frac{5}{15}$

$P(0 - 5K|Approve) = \frac{3}{7}$
$P(0 - 5K|Reject) = \frac{3}{5}$
$P(Good|Approve) = \frac{3}{7}$
$P(Good|Reject) = \frac{0}{5}$
$P(High|Approve) = \frac{2}{7}$
$P(High|Reject) = \frac{2}{5}$

$\hat{P}(Approve|0 - 5K, Good, High) = \frac{7}{12} \times \frac{3}{7} \times \frac{3}{7} \times \frac{2}{7} = \frac{1}{98}.$

$\hat{P}(Reject|0-5K, Good, High) = \frac{5}{12} \times \frac{3}{5} \times \frac{0}{5} \times \frac{2}{5} = 0.$

Note that as Naïve Bayes' final step does not really compute probabilities (as it assumes independence between features) we use $\hat{P}()$ instead of $P()$ to denote that NB doesn't calculate a formal probability. We take the argmax over the possible classes to find the solution. Here, as $\hat{P}(Approve) > \hat{P}(Reject)$, we choose Approve. This is the same decision as was rendered by the decision tree classifier.

5. We want to use the data to predict whether Ah Bui, whose weight is 78 kg and whose height is 179 cm, will be recruited into the Fit And Trim club.

   (a) By using the Nearest Neighbour learning algorithm on the raw data, predict whether Ah Bui will be recruited into the Fit And Trim club. Is there a problem with this prediction?

   The Euclidean distance between Ah Bui's weight and height from others are as follows:

Person	Weight (kg)	Height (cm)	Accepted?	Distance
A	79	189	Yes	10.050
B	76	170	Yes	9.220
C	77	155	Yes	24.021
D	72	163	No	17.088
E	73	195	No	16.763
F	70	182	No	8.544

   Since the distance between F and Ah Bui is the shortest, the Nearest Neighbour algorithm predicts Ah Bui not to be accepted into the Fat And Trim club.

   There is a problem in this prediction because it is obvious that the recruitment criterion is mainly by weight. The mis-prediction caused by the large difference between the scales of the weights and heights, which is about 1:5, but the Euclidean distance treats both weights and heights as having the same scale.

   (b) Reapply the Nearest Neighbour algorithm on the normalized data to predict whether Ah Bui will be recruited into the Fit And Trim club. Did normalizing the data improve the prediction?

   Ah Bui's weight and height, when normalized, becomes 0.80 and 0.58 respectively. The normalized weights and heights, as well as the re-computed Euclidean distances, are shown below:

Person	Weight	Height	Accepted?	Distance
A	0.90	0.78	Yes	0.224
B	0.60	0.40	Yes	0.269
C	0.70	0.10	Yes	0.490
D	0.20	0.26	No	0.680
E	0.30	0.90	No	0.594
F	0.00	0.64	No	0.802

   Now the shortest distance from Ah Bui is A. As such, Ah Bui is predicted to be recruited into the Fat And Trim club.

   The normalization causes both weights and heights to have similar scales, and therefore made a better prediction.