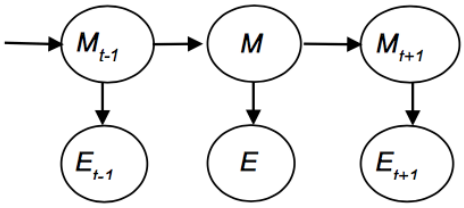
[CMSC-671] Home-Work 4

Mohit Khatwani, Rajat Patel, Priyank Agarwal, Sarthak Mehta

Part I: Filtering

The dependence between whether movie is popular (M) or not and the hopper is empty (E) or not is represented by following model



M_{t-1}	$P(M_t \mid M_{t-1})$
T	0.75
F	0.1

E_t	$P(E_t \mid M_t)$
T	0.9
F	0.1

Given: 1.) $M_0 = 0.75$

2.) Hopper is empty on day1 and day2

3.) Hopper is full on day3

$P(M_1) = \sum_{M0} P(M_1|M_0)P(M_0)$

= <0.75,0.25> 0.75 + <0.1,0.9> 0.25

= <0.562,0.187> + <0.025,0.225>

≅ <0.587, 0.412>

$P(M_2) = \sum_{M1} P(M_2|M_1)P(M_1)$

= <0.75, 0.25> 0.587 + <0.1, 0.9> 0.412

= <0.440, 0.148> + <0.041, 0.370>

≅ <0.482, 0.518>

$P(M_2 \mid E_2) = \alpha\ P(E_2, M_2)\ P(M_2)$

= α <0.9, 0.1> <0.482, 0.518>

= α <0.434, 0.052>

≅ <0.893, 0.107>

$P(M_3|E_2) = \sum_{M2} P(M_2|M_2)P(M_2|E_2)$

= <0.75, 0.25> 0.893 + <0.1, 0.9> 0.107

= <0.670, 0.223> + <0.0107, 0.096>

≅ <0.680, 0.320>

$P(M_3|E_2, -E_3) = \alpha\ P(-E_3, M_3)\ P(M_3, E_2)$

= α <0.1, 0.9> <0.680, 0.320>

= α <0.068, 0.288>

≅ <0.191, 0.809>

Probability of movie being popular on day 3 is 0.191 or 19.1%

Part II: Learning in the Wild

The problem here depicts the child as a learning agent present in an unknown environment predicting the likes and dislikes. Thus, this problem formulates into general learning model, where the child is given large data set of food choices as inputs. Further, the output of the prediction is usually influenced by the child's domain knowledge. For example, child might have prior knowledge of particular food, also it can be influenced by nature which acts as an environment, where the child predicts the food as inedible based on the knowledge imparted by nature. The performance of the child's ability to like or dislike a particular eatable purely depends on its domain knowledge, supervised by a critic which evaluates child's strategy based on the ability of prediction by testing it on unknown or novel data items. Thus, according to the environment and taking into consideration the learning agent, this learning model would be a kind of semi-supervised learning where the child is initially supervised based on training data and would later become independent.

The percepts of the child would be sense organs trained in accordance with nutritional knowledge of food. For example, if the training data imparts the child with the knowledge of ‘food A’ being nutritious though not tasting as good as ‘food B’, the child would be supervised to choose ‘food A’ based on the nutrition content. Thus, the types of learning involved here would be semi-supervised and reinforcement learning, where it would develop a sense in choosing a particular kind and variety of food and develop its liking towards it.

The sub-function that the child learns is through categorization based on nutritional parameter depending on the food items and getting trained over the same. For example, the child learns to develops a liking towards particular type of food depending on a particular parameter of nutrition value.

Part III: Decision Tree Learning

1. Information Gain

a. At the root node for your decision tree in this domain, what is the information gain associated with a split on the attribute B.

$E_{overall}$ (4,5,4) = $\sum -plog(p)$

$E_{overall}$ (4,5,4) = $-[4/13 \log\left(\frac{4}{13}\right) + 5/13 \log\left(\frac{5}{13}\right) + 4/13 \log\left(\frac{4}{13}\right)]$

$E_{overall}$ (4,5,4) = 1.574 ……(1)

Splitting on attribute B

$E_b = P_{very\ far}\ E_{b=very\ far} + P_{far}\ E_{b=far} + P_{near}\ E_{b=near}$

$E_{b=very\ far}$ (3,1) = $-[3/4 \log\left(\frac{3}{4}\right) + 1/4 \log\left(\frac{1}{4}\right)]$ = 0.811

$E_{b=far}$ (2,1,2) = $-[2/5 \log\left(\frac{2}{5}\right) + 1/5 \log\left(\frac{1}{5}\right) + 2/5 \log\left(\frac{2}{5}\right)]$ = 1.522

$E_{b=near}$ (2,2) = $-\ [2/4 \log\left(\frac{2}{4}\right) + 2/4 \log\ (\frac{2}{4})]$ = 1

For calculating total Entropy for attribute B
$P_{very\ far}$ = 4/13; P_{far} = 5/13; P_{near} = 4/13

E_b = 4/13*0.811 + 5/13*1.522 + 4/13*1 = 1.143

$Information\_Gain_B = E_{overall} - E_b = 0.431$ …..(2)

b. What would it be for a split at the root on the attribute L? (Use a threshold of 50 for L (i.e., assume a binary split: L ≤ 50, L > 50)

Using Equation (1) for $E_{overall}$

Splitting of attribute L

$E_L = P_{L<=50}\ E_{L<=50} + P_{L>50}\ E_{L>50}$

$E_{L<=50}$ (3,1,3) = $-[3/7 \log\left(\frac{3}{7}\right) + 1/7 \log\left(\frac{1}{7}\right) + 3/7 \log(\frac{3}{7})]$ = 1.45

$E_{L>50}$ (2,3,1) = $-[2/6 \log\left(\frac{2}{6}\right) + 3/6 \log\left(\frac{3}{6}\right) + 1/6 \log(\frac{1}{6})]$ = 1.055

For calculating total Entropy of attribute L
$P_{L<=50}$ = 7/13; $P_{L>50}$ = 6/13

E_L = 7/13*1.45 + 6/13*1.055 = 1.268

Information_Gain$_L$ = $E_{overall}$ – E_L = 0.306 .....(3)

2. Gain Ratios

a) Again, at the root node, what is the *gain ratio* associated with the attribute B? What is the gain ratio for the Y attribute at the root (using the same threshold as in 1b)? Gain Ratio is ratio of information gain and split information at that attribute

Gain Ratio associated with attribute B
Gain_Ratio$_B$ = $\frac{Information\_Gain_B}{Split\_Information_B}$ = $\frac{0.431}{1.143}$ = 0.377

Gain Ratio associated with attribute L
Gain_Ratio$_L$ = $\frac{Information\_Gain_L}{Split\_Information_L}$ = $\frac{0.306}{1.268}$ = 0.241

3. Decision Tree

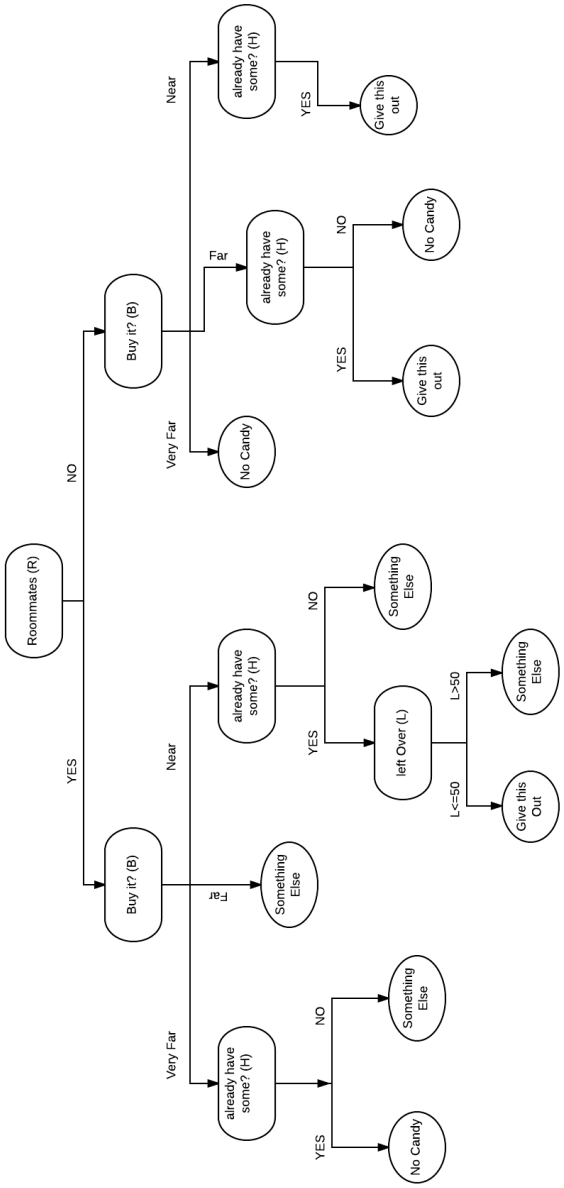
a) Suppose you build a decision tree that splits on the H attribute at the root node. How many child nodes are there are at the *first level* of the decision tree?

If H attribute is selected in decision tree as root node, then there will be **two** child nodes of that tree at first level. As there are only two possible values that H attribute can take: No and Yes.

b) After H, which branches require a further split?

After H both branches require a further split because both children nodes have non-zero entropy i.e. they are not leaf nodes.

c) Draw the smallest (fewest nodes) decision tree you can you construct for this dataset. The tree should show which attribute you split on for each branch, and show the decisions (class predictions) at the leaves.



d) What method(s) did you use to find that tree? Show all calculations

For selecting Root node of decision tree

Information_Gain$_H$:

$E_H = P_{H=Yes}\, E_{H=Yes} + P_{H=No}\, E_{H=No}$

$E_{H=Yes}(2,2,4) = -[2/8 \log(\frac{2}{8}) + 2/8 \log(\frac{2}{8}) + 4/8 \log(\frac{4}{8})] = 1.5$

$E_{H=No}(3,2) = -[3/5 \log(\frac{3}{5}) + 2/5 \log(\frac{2}{5})] = 0.971$

For final attribute entropy calculations

$P_{H=Yes} = 8/13$; $P_{H=No} = 5/13$

$E_H = 8/13*1.5 + 5/13*0.971 = 1.296$

Information_Gain$_H$ = $E_{overall} - E_H$ = 0.278

Information_Gain$_R$:

$E_R = P_{R=Yes}\, E_{R=Yes} + P_{R=No}\, E_{R=No}$

$E_{R=Yes}(5,1,1) = -[5/7 \log(\frac{5}{7}) + 1/7 \log(\frac{1}{7}) + 1/7 \log(\frac{1}{7})] = 1.149$

$E_{R=No}(3,3) = -[3/6 \log(\frac{3}{6}) + 3/6 \log(\frac{3}{6})] = 1$

For final attribute entropy calculations

$P_{R=Yes} = 7/13$; $P_{R=No} = 6/13$

$E_R = 7/13*1.149 + 6/13*1 = 1.080$

Information_Gain$_R$ = $E_{overall} - E_R$ = 0.494

Splitting attribute at level 1 of decision tree

Information_Gain$_H$:

$E_H = P_{H=Yes}\, E_{H=Yes} + P_{H=No}\, E_{H=No}$

$E_{H=Yes}(2,1,1) = -[2/4 \log(\frac{2}{4}) + 1/4 \log(\frac{1}{4}) + 1/4 \log(\frac{1}{4})] = 1.5$

$E_{H=No}$ (3,0,0) = 0

For final entropy calculations
$P_{H=Yes}$ = 4/7; $P_{H=No}$ = 3/7

E_H = 4/7*1.5 + 3/7*0 = 0.857

Information_Gain$_H$ = E_R – E_H = 0.223

Information_Gain$_B$:

$E_b = P_{very\ far}\ E_{b=very\ far} + P_{far}\ E_{b=far} + P_{near}\ E_{b=near}$

$E_{B=very\ far}$ (1,1) = $-\ [1/2\log\ (\frac{1}{2}) + 1/2\log\ (\frac{1}{2})]$ = 1

$E_{B=far}$ (2,0,0) = 0

$E_{B=near}$ (2,0,1) = $-\ [2/3\ \log\left(\frac{2}{3}\right) + 1/3\log\ (\frac{1}{3})]$ = 0.917

For final Entropy calculations
$P_{B=very\ far}$ = 2/7; $P_{B=far}$ = 2/7; $P_{B=near}$ = 3/7

E_B = 2/7*1 + 0 + 0.917*3/7 = 0.679
Information_Gain$_B$ = E_R – E_B = 0.401

Information_Gain$_{L:}$

$E_L = P_{L<=50}\ E_{L<=50} + P_{L>50}\ E_{L>50}$

$E_{L<=50}$ (3,1,1) = $-\ [3/5\log\ (\frac{3}{5}) + 1/5\log\ (\frac{1}{5}) + 1/5\log\ (\frac{1}{5})]$ = 1.371

$E_{L>50}$ (2,0,0) = 0 ; $P_{L<=50}$ = 5/7; $P_{L>50}$ = 2/7

E_L = 5/7*1.371 + 0*2/7 = 0.979
Information_Gain$_L$ = E_R – E_L = 0.101

Splitting attribute at level 2 of decision tree

Information_Gain$_H$:

$E_H = P_{H=Yes}\ E_{H=Yes} + P_{H=No}\ E_{H=No}$

$E_{H=Yes}$ (1,3) = $-[1/4\ \log\left(\frac{1}{4}\right) + 3/4\ \log\left(\frac{3}{4}\right)]$ = 0.811

$E_{H=No}$ (2,0,0) = 0

For final attribute calculations
$P_{H=Yes}$ = 4/6; $P_{H=No}$ = 2/6

E_H = 4/6*0.811 + 2/6*0 = 0.541

Information_Gain$_H$ = E_B – E_H = 0.138

Information_Gain$_L$:

$E_L = P_{L<=50} E_{L<=50} + P_{L>50} E_{L>50}$

$E_{L<=50}$ (2,0,0) = 0

$E_{L>50}$ (3,1) = $-[3/4 \log\left(\frac{3}{4}\right) + 1/4 \log\left(\frac{1}{4}\right)]$ = 0.811

For final attribute calculations
$P_{L<=50}$ = 2/6; $P_{L>50}$ = 4/6

E_L = 2/6*0 + 4/6*0.811 = 0.541

Information_Gain$_L$ = E_B– E_L = 0.138