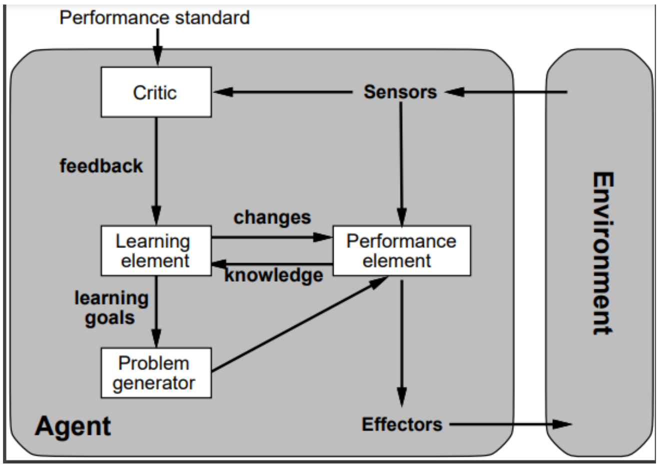
CS 3346 Assignment 3

Minsoo Park

18.1 Consider the problem faced by an infant learning to speak and understand a language. Explain how this process fits into the general learning model. Describe the percepts and actions of the infant, and the types of learning the infant must do. Describe the subfunctions the infant is trying to learn in terms of inputs and outputs, and available example data.



Environment: People (parents), physical objects (TV, radio, etc…), events.

Sensors: Eye and Ears; infant can see some stuff and hear the word from it.

Critic: A utility function which evaluate infant's performance.

Performance standard: Standard language, words, or sentences.

Performance element: Infant speaks learned language.

Learning element: Mimicry, positive and negative reinforcement learning (reward system or punishment).

Problem generator: Give better suggestion; but not really effectful to infant.

Effectors: Mouth, togue, and lips.

18.2 Repeat Exercise 18.1 for the case of learning to play tennis (or some other sport with which you are familiar). Is this supervised learning or reinforcement learning?

Environment: People (coach, friend, or anyone who play tennis), objects (TV, YouTube, etc.…) , events (tournament), Ball, court.

Sensors: Eye, arm, and leg.

Critic: A utility function which evaluate player such as how to swing, how to move.

Performance standard: Playing a good tennis game.

Performance element: A player plays a tennis.

Learning element: How to get a point from the game, positive and negative reinforcement learning (reward system or punishment).

Problem generator: Give better suggestion option; Can player do better with such action or practice?

Effectors: Eye, leg, and arm.

The reinforcement learning is best for this situation. We can have such environment for present state and observe the function to get best reward. Then we can decide best action from it.

18.3 Suppose we generate a training set from a decision tree and then apply decision-tree learning to that training set. Is it the case that the learning algorithm will eventually return the correct tree as the training-set size goes to infinity? Why or why not?

No, when we have infinity size of training set which means we have infinite input value with output data. But we can’t classify the conclusion of the function because there is possibility which output different value by same input data.

18.6 Consider the following data set comprised of three binary input attributes (A1, A2, and A3) and one binary output:

Example	A_1	A_2	A_3	Output y
$\mathbf{x}_1$	1	0	0	0
$\mathbf{x}_2$	1	0	1	0
$\mathbf{x}_3$	0	1	0	0
$\mathbf{x}_4$	1	1	1	1
$\mathbf{x}_5$	1	1	0	1

Use the algorithm in Figure 18.5 (page 702) to learn a decision tree for these data. Show the computations made to determine the attribute to split at each node.

Add: use the tree built to predict a new (test) example with values of A1-A3 to be 0, 1, 1, respectively. What class (y) the tree would predict?

$$E(Y) = -\tfrac{2}{5}\log_2\tfrac{2}{5} - \tfrac{3}{5}\log_2\tfrac{3}{5} = 0.971$$

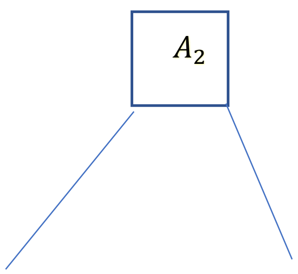
$$E(A_1) = \frac{4}{5}\left(-\frac{2}{4}\log_2\frac{2}{4} - \frac{2}{4}\log_2\frac{2}{4}\right) + \frac{1}{5}\ (-1\log_2 1)$$
$$= 0.8 + 0 = 0.8$$

$$E(A_2) = \frac{3}{5}\left(-\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3}\right) + \frac{2}{5}\ (-1\log_2 1)$$
$$= 0.551 + 0 = 0.551$$

$$E(A_3) = \frac{2}{5}\left(-\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2}\right) + \frac{3}{5}\left(-\frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3}\right)$$
$$= 0.4 + 0.551 = 0.951$$

Info gain; $A_1 = 0.971 - 0.8 = 0.171, A_2 = 0.971 - 0.551 = 0.420, A_3 = 0.971 - 0.951 = 0.020$

Therefore, A_2 *has maximum info gain, we should set up as root node.*



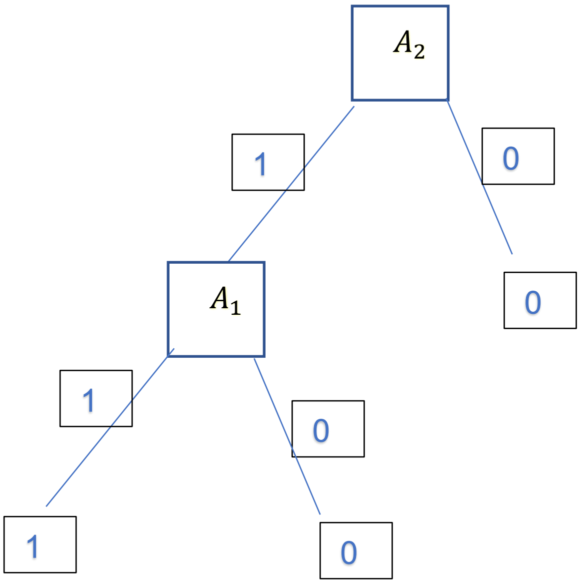
When we have a root node of A_2, let's calculate when it has 1.

$$E(A_1|A_2 = 1) = \frac{2}{3}(-1\log_2 1) + \frac{1}{3}(-1\log_2 1) = 0$$

$$E(A_3|A_2 = 1) = \frac{1}{3}(-1\log_2 1) + \frac{2}{3}\left(-\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2}\right) = 0 + 0.667 = 0.667$$

Therefore, A_1 *has smallest value so it should be come after* A_2

We don't care the
A_3 *, because it doesn't give any change of output* Y *by provided data set*



Try to put a new data (test) with $A_1 = 0, A_2 = 1, A_3 = 1$

It will return 0 for output y value.

18.6 continued: Apply k-NN algorithm with k=1 and k=3 to predict a new example with values of A1-A3 to be 0, 1, 1, respectively? If there is a tie, break the tie randomly.

$$when\ y = 1, \sqrt{\frac{2}{5}} = 0.632$$

$$when\ y = 0, \sqrt{\frac{2}{5}} = 0.775$$

When k = 1, with value 0,1,1 of A1-A3 it will give closet output y is 0.

When k = 3, It will be determined least two "0" of y, so it should be 0 as well.

18.6 continued: Apply the Naive Bayes learning algorithm using this small dataset on a new example with values of A1-A3 to be 0, 1, 1, respectively? When the probability is 0, use a small number (such as 0.05) to substitute 0. (If we don't have time to cover Naive Bayes, you can search the Web - it is a very simple algorithm).

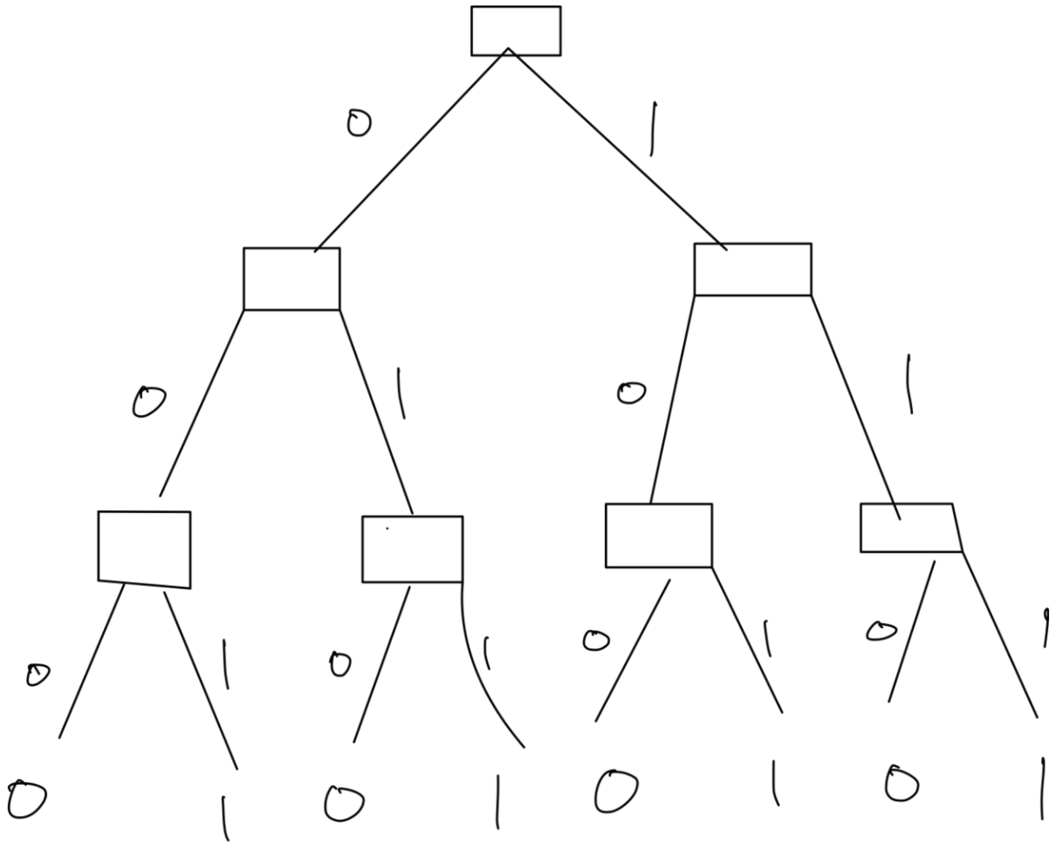
$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}$$

Example	A1	A2	A3	y
x1	1	0	0	0
x2	1	0	1	0
x3	0	1	0	0
x4	1	1	1	1
x5	1	1	0	1
test	0	1	1	0

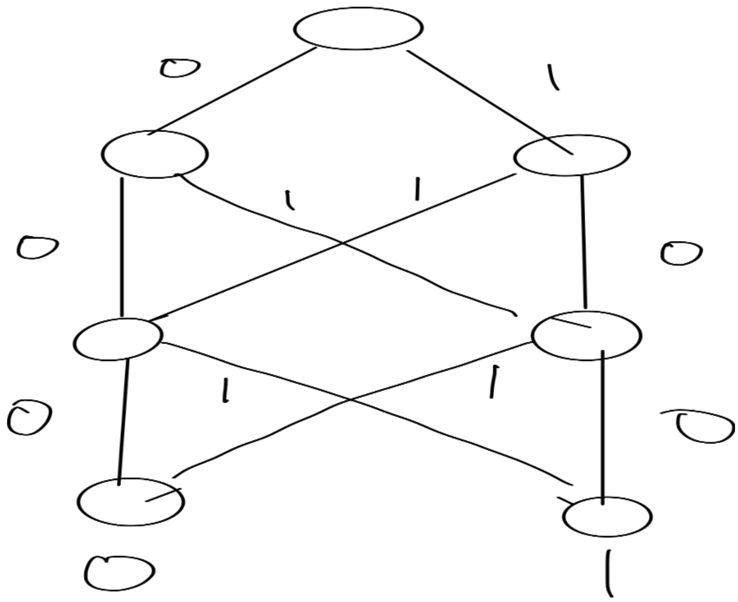
$$\frac{\left(\frac{2}{6} * \frac{1}{3}\right) + \left(\frac{4}{6} * \frac{1}{3}\right) + \left(\frac{3}{6} * \frac{1}{3}\right)}{\frac{4}{6}} = 0.75$$

18.7 A decision graph is a generalization of a decision tree that allows nodes (i.e., attributes used for splits) to have multiple parents, rather than just a single parent. The resulting graph must still be acyclic. Now, consider the XOR function of three binary input attributes, which produces the value 1 if and only if an odd number of the three input attributes has value 1.

a. Draw a minimal-sized decision tree for the three-input XOR function.



b. Draw a minimal-sized decision graph for the three-input XOR function.



18.12 Construct a decision list to classify the data below. Select tests to be as small as possible (in terms of attributes), breaking ties among tests with the same number of attributes by selecting the one that classifies the greatest number of examples correctly. If multiple tests have the same number of attributes and classify the same number of examples, then break the tie using attributes with lower index numbers (e.g., select A1 over A2).

Example	A_1	A_2	A_3	A_4	y
$\mathbf{x}_1$	1	0	0	0	1
$\mathbf{x}_2$	1	0	1	1	1
$\mathbf{x}_3$	0	1	0	0	1
$\mathbf{x}_4$	0	1	1	0	0
$\mathbf{x}_5$	1	1	0	1	1
$\mathbf{x}_6$	0	1	0	1	0
$\mathbf{x}_7$	0	0	1	1	1
$\mathbf{x}_8$	0	0	1	0	0

Y is 1 when x1, x2, x3, x5, x7

Y is 0 when x4, x6, x8

When A1 is 1, then y is 1. (x1, x2, x5)

If A1 is 0, then go A2, A3, A4 until find attribute output same y value

When A2 =1 and A3 = 1, y is 0 (x4)

If no, then go to next

when A2 = 1 and A4 = 0, y is 1 (x3)

If no, then go to next

When A2 = 1, y is 0 (x6)

If no, then go to next

When A4 = 1, y is 1 (x7).

If no y is 0 (x8).