Nick Palacio

Raj Dasgupta

October 1, 2018

Introduction to AI

Homework 2

Mathematics Part

Question 1.

You have a decision tree algorithm and you are trying to figure out which attribute is the best to test on first. You are using the “information gain” metric.

- *You are given a set of 128 examples, with 64 positively labeled and 64 negatively labeled.*
- *There are three attributes: Homeowner (H), In Debt (ID), and Rich (R).*
- *For 64 examples, Home Owner is true. The Homeowner=true examples are 1/4 negative and 3/4 positive.*
- *For 96 examples, In Debt is true. Of the In Debt=true examples, 1/2 are positive and half are negative.*
- *For 32 examples, Rich is true. 3/4 of the Rich=true examples are positive and 1/4 are negative*

You must show all mathematical calculations/steps to get full points for each subpart (a) – (d) below. Just writing the final answer in each subpart (correct or not) will get zero points.

a) *What is the entropy of the initial set of examples?*
b) *What is the information gain of splitting on the Home Owner attribute as the root node?*
c) *What is the information gain of splitting on the In Debt attribute as the root node?*
d) *What is the information gain of splitting on the Rich attribute as the root node?*
e) *Which attribute do you split on?*

<u>1a</u>

Since in the initial set half the examples are positive and half are negative our initial entropy is given as:

$$H_{initial}\left(<\frac{1}{2},\frac{1}{2}>\right) = -\frac{1}{2} * log_2\left(\frac{1}{2}\right) - \frac{1}{2} * log_2\left(\frac{1}{2}\right)$$

$$H_{initial}\left(<\frac{1}{2},\frac{1}{2}>\right) = \frac{1}{2} + \frac{1}{2}$$

$$\boldsymbol{H_{initial}\left(<\frac{1}{2},\frac{1}{2}>\right) = 1}$$

<u>1b</u>

For the 64 examples where Home Owner is true, 16 (1/4) are negative, and 48 (3/4) are positive. This means that for the other 64 examples where Home Owner is false, 48 (3/4) are negative and 16 (1/4) are positive. The information gain of splitting on the Home Owner as the root node is given as:

$$InfoGain_{HomeOwner} = H_{initial}\left(<\frac{1}{2},\frac{1}{2}>\right) - \left(\frac{64}{128} * EntropyOfPositiveHomeOwnerExamples + \frac{64}{128} * EntropyOfNegativeHomeOwnerExamples\right)$$

$$InfoGain_{HomeOwner} = H_{initial}\left(<\frac{1}{2},\frac{1}{2}>\right) - (\frac{64}{128} * H\left(<\frac{3}{4},\frac{1}{4}>\right) + \frac{64}{128} * H\left(<\frac{1}{4},\frac{3}{4}>\right))$$

$$InfoGain_{HomeOwner} = 1 - (\frac{64}{128} * \left(-\frac{3}{4} * log_2\left(\frac{3}{4}\right) - \frac{1}{4} * log_2\left(\frac{1}{4}\right)\right) + \frac{64}{128} * \left(-\frac{1}{4} * log_2\left(\frac{1}{4}\right) - \frac{3}{4} * log_2\left(\frac{3}{4}\right)\right))$$

$$InfoGain_{HomeOwner} = 1 - (\frac{64}{128} * (.811) + \frac{64}{128} * (.811))$$

$$InfoGain_{HomeOwner} = 1 - (.811)$$

$$\boldsymbol{InfoGain_{HomeOwner} = .189}$$

1c

For the 96 examples where In Debt is true, 48 (1/2) are negative and 48 (1/2) are positive. This means that for the other 32 examples where In Debt is false, 16 (1/2) are negative and 16 (1/2) are positive. The information gain of splitting on In Debt as the root node can be given as:

$$InfoGain_{InDebt} = H_{initial}\left(<\frac{1}{2},\frac{1}{2}>\right) - \left(\frac{96}{128} * EntropyOfPositiveInDebtExamples + \frac{32}{128} * EntropyOfNegativeInDebtExamples\right)$$

$$InfoGain_{InDebt} = H_{initial}\left(<\frac{1}{2},\frac{1}{2}>\right) - \left(\frac{96}{128} * H\left(<\frac{1}{2},\frac{1}{2}>\right) + \frac{32}{128} * H\left(<\frac{1}{2},\frac{1}{2}>\right)\right)$$

$$InfoGain_{InDebt} = 1 - \left(\frac{96}{128} * 1 + \frac{32}{128} * 1\right)$$

$$InfoGain_{InDebt} = 1 - (1)$$

$$\boldsymbol{InfoGain_{InDebt} = 0}$$

1d

For the 32 examples where Rich is true, 8 (1/4) are negative and 24 (3/4) are positive. This means that for the 96 examples where Rich is false, 56 are negative and 40 are positive. The information gain of splitting on Rich as the root node can be given as:

$$InfoGain_{Rich} = H_{initial}\left(<\frac{1}{2},\frac{1}{2}>\right) - \left(\frac{32}{128} * EntropyOfPositiveRichExamples + \frac{96}{128} * EntropyOfNegativeRichExamples\right)$$

$$InfoGain_{Rich} = 1 - \left(\frac{32}{128} * H(<\frac{3}{4},\frac{1}{4}>) + \frac{96}{128} * H(<\frac{40}{96},\frac{56}{96}>)\right)$$

$$InfoGain_{Rich} = 1 - \left(\frac{32}{128} * \left(-\frac{3}{4} * log_2\left(\frac{3}{4}\right) - \frac{1}{4} * log_2\left(\frac{1}{4}\right)\right) + \frac{96}{128} * \left(-\frac{40}{96} * log_2\left(\frac{40}{96}\right) - \frac{56}{96} * log_2\left(\frac{56}{96}\right)\right)\right)$$

$$InfoGain_{Rich} = 1 - \left(\frac{32}{128} * (.811) + \frac{96}{128} * (.526 + .454)\right)$$

$$InfoGain_{Rich} = 1 - (.203 + .735)$$

$$\boldsymbol{InfoGain_{Rich} = .062}$$

1e

Since:

$$InfoGain_{HomeOwner} > InfoGain_{Rich} > InfoGain_{InDebt}$$

I would split on Home Owner so I can have the largest information gain.

Question 2.

You are given the following training set with three attributes A1, A2 and A3 and one binary output y. Construct the decision tree for the hypothesis in the training set. You have to show the steps for the calculations of the information gain of each attribute and the complete decision tree to get full points.

Example	*A1*	*A2*	*A3*	*Output 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*

In order to construct the decision tree I first need to find out what attribute to split on for the root node. To do this I will calculate the information gain for splitting on A1, A2 and A3 as the root node and select the one with the highest information gain.

The first thing to do is calculate the initial entropy of our dataset. Since we have 2 examples where Y is positive and 3 where Y is negative our initial entropy can be given as:

$$H_{Initial}\left(<\frac{2}{5},\frac{3}{5}>\right) = -\frac{2}{5} * log_2\left(\frac{2}{5}\right) - \frac{3}{5} * log_2\left(\frac{3}{5}\right)$$

$$H_{Initial}\left(<\frac{2}{5},\frac{3}{5}>\right) = .529 + .442$$

$$\boldsymbol{H_{Initial}\left(<\frac{2}{5},\frac{3}{5}>\right) = .971}$$

<u>A1</u>

For the A1 attribute we have 4 rows where A1 is positive. In these 4 rows, Y is positive in 2 rows and negative in 2 rows. We also have 1 row where A1 is negative. In this row Y is negative. Our information gain for splitting on this attribute as the root node can be given by:

$$InfoGain_{A1} = H_{Initial}\left(<\frac{2}{5},\frac{3}{5}>\right) - \left(\frac{4}{5} * EntropyOfPositiveA1Examples + \frac{1}{5} * EntropyOfNegativeA1Examples\right)$$

$$InfoGain_{A1} = H_{Initial}\left(<\frac{2}{5},\frac{3}{5}>\right) - \left(\frac{4}{5} * H(<\frac{2}{4},\frac{2}{4}>) + \frac{1}{5} * H(<\frac{0}{1},\frac{1}{1}>)\right)$$

$$InfoGain_{A1} = .971 - \left(\frac{4}{5} * \left(-\frac{2}{4} * log_2\left(\frac{2}{4}\right) - \frac{2}{4} * log_2\left(\frac{2}{4}\right)\right) + \frac{1}{5} * 0\right)$$

$$InfoGain_{A1} = .971 - \left(\frac{4}{5} * \left(\frac{1}{2} + \frac{1}{2}\right) + 0\right)$$

$$InfoGain_{A1} = .971 - \left(\frac{4}{5}\right)$$

$$\boldsymbol{InfoGain_{A1} = .171}$$

<u>A2</u>

For the A2 attribute we have 3 rows that are positive. Of those 3 rows where A2 is positive, Y is positive in 2 rows and negative in 1 row. For the 2 other rows where A2 is negative, Y is positive in 0 rows and negative in 2 rows. Our information gain for splitting on this attribute as the root node can be given by:

$$InfoGain_{A2} = H_{Initial}\left(<\frac{2}{5},\frac{3}{5}>\right) - \left(\frac{3}{5} * EntropyOfPositiveA2Examples + \frac{2}{5} * EntropyOfNegativeA2Examples\right)$$

$$InfoGain_{A2} = H_{Initial}\left(<\frac{2}{5},\frac{3}{5}>\right) - \left(\frac{3}{5} * H(<\frac{2}{3},\frac{1}{3}>) + \frac{2}{5} * H(<\frac{0}{2},\frac{2}{2}>)\right)$$

$$InfoGain_{A2} = .971 - \left(\frac{3}{5} * \left(-\frac{2}{3} * log_2\left(\frac{2}{3}\right) - \frac{1}{3} * log_2\left(\frac{1}{3}\right)\right) + \frac{2}{5} * 0\right)$$

$$InfoGain_{A2} = .971 - \left(\frac{3}{5} * (.39 + .528) + 0\right)$$

$$InfoGain_{A2} = .971 - \left(\frac{3}{5} * (.918)\right)$$

$$InfoGain_{A2} = .971 - (.551)$$

$$\boldsymbol{InfoGain_{A2} = .42}$$

A3

For the A3 attribute we have 2 rows where A3 is positive. Of those two rows where A3 is positive, Y is positive in 1 and negative in 1. For the other 3 rows where A3 is negative, Y is positive in 1 and negative in 2. Our information gain for splitting on this attribute as the root node can be given by:

$$InfoGain_{A3} = H_{Initial}\left(<\frac{2}{5},\frac{3}{5}>\right) - \left(\frac{2}{5} * EntropyOfPositiveA3Examples + \frac{3}{5} * EntropyOfNegativeA3Examples\right)$$

$$InfoGain_{A3} = .971 - \left(\frac{2}{5} * H(<\frac{1}{2},\frac{1}{2}>) + \frac{3}{5} * H(<\frac{1}{3},\frac{2}{3}>)\right)$$

$$InfoGain_{A3} = .971 - \left(\frac{2}{5} * 1 + \frac{3}{5} * \left(-\frac{1}{3} * log_2\left(\frac{1}{3}\right) - \frac{2}{3} * log_2\left(\frac{2}{3}\right)\right)\right)$$

$$InfoGain_{A3} = .971 - \left(\frac{2}{5} + \frac{3}{5} * (.918)\right)$$

$$InfoGain_{A3} = .971 - (.951)$$

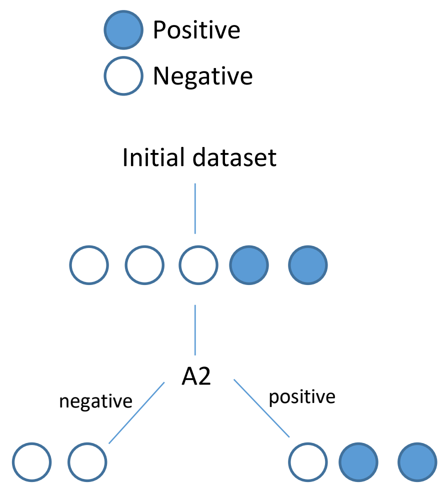
$$\boldsymbol{InfoGain_{A3} = .02}$$

First Split

Since:

$$InfoGain_{A2} > InfoGain_{A1} > InfoGain_{A3}$$

We will split on A2 first, resulting in the start of our tree looking like this:



Next we need to figure out how what attribute to split the leftover data from the positive A2 results.

New Initial Entropy

In order to know which attribute to split on we will need to recalculate the information gain for A1 and A3 with a new initial dataset.

Our new initial dataset has Y as positive in 2 examples and negative in 1 example. This gives our *new* initial entropy as:

$$H_{Initial}\left(<\frac{2}{3},\frac{1}{3}>\right) = -\frac{2}{3} * log_2\left(\frac{2}{3}\right) - \frac{1}{3} * log_2\left(\frac{1}{3}\right)$$

$$\boldsymbol{H_{Initial}\left(<\frac{2}{3},\frac{1}{3}>\right) = .918}$$

A1

Now we can calculate the information gain from splitting on A1 given a new initial dataset and entropy. In the new dataset, A1 is positive in 2 rows and negative in 1 row. In the 2 rows where A1 is positive, Y is also positive. In the 1 row where A1 is negative, Y is also negative. This gives our information gain as:

$$InfoGain_{A1} = H_{Initial}\left(<\frac{2}{3},\frac{1}{3}>\right) - \left(\frac{2}{3} * EntropyOfPositiveA1Examples + \frac{1}{3} * EntropyOfNegativeA1Examples\right)$$

$$InfoGain_{A1} = H_{Initial}\left(<\frac{2}{3},\frac{1}{3}>\right) - \left(\frac{2}{3} * H(<\frac{2}{2},\frac{0}{2}>) + \frac{1}{3} * H(<\frac{0}{1},\frac{1}{1}>)\right)$$

$$InfoGain_{A1} = H_{Initial}\left(< \frac{2}{3}, \frac{1}{3} >\right) - \left(\frac{2}{3} * 0 + \frac{1}{3} * 0\right)$$

$$InfoGain_{A1} = .918 - (0)$$

$$\boldsymbol{InfoGain_{A1} = .918}$$

<u>A3</u>

We will also calculate the information gain from splitting on A3 given a new initial dataset and entropy. In the new dataset, A3 is positive in 1 row where Y is also positive. A3 is negative in 2 rows where Y is positive in 1 of them and negative in the other. This gives our information gain as:

$$InfoGain_{A3} = H_{Initial}\left(< \frac{2}{3}, \frac{1}{3} >\right) - \left(\frac{1}{3} * EntropyOfPositiveA3Examples + \frac{2}{3} * EntropyOfNegativeA3Examples\right)$$

$$InfoGain_{A3} = H_{Initial}\left(< \frac{2}{3}, \frac{1}{3} >\right) - \left(\frac{1}{3} * H\left(< \frac{1}{1}, \frac{0}{1} >\right) + \frac{2}{3} * H\left(< \frac{1}{2}, \frac{1}{2} >\right)\right)$$

$$InfoGain_{A3} = .918 - \left(\frac{1}{3} * 0 + \frac{2}{3} * \left(-\frac{1}{2} * log_2\left(\frac{1}{2}\right) - \frac{1}{2} * log_2\left(\frac{1}{2}\right)\right)\right)$$

$$InfoGain_{A3} = .918 - \left(0 + \frac{2}{3} * (1)\right)$$

$$InfoGain_{A3} = .918 - \left(\frac{2}{3}\right)$$

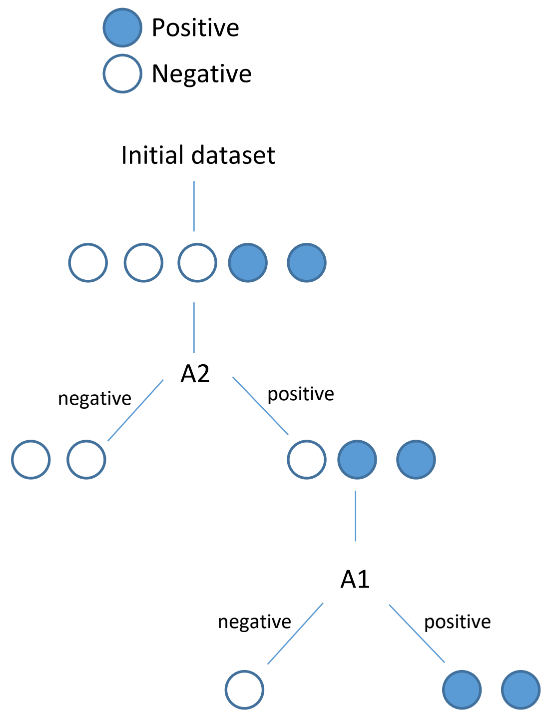
$$\boldsymbol{InfoGain_{A3} = .251}$$

Final Split

Since:

$$InfoGain_{A1} > InfoGain_{A3}$$

We will split on A1, resulting in our final decision tree looking like this:



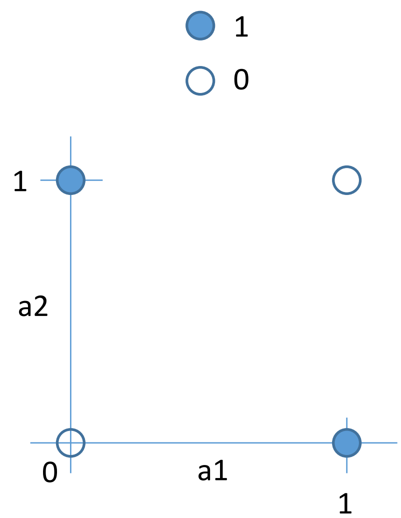
Question 2.

Construct by hand a neural network for the Boolean XOR function with two inputs.

The truth table for the XOR function can be given as:

a_1	a_2	XOR
0	0	0
0	1	1
1	0	1
1	1	0

The plot of this function can be given as:



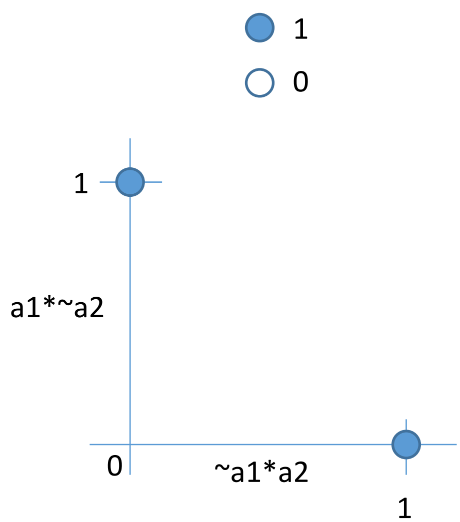
The equation for the XOR function can be given as:

$$XOR = (a_1 + a_2) * \sim(a_1 * a_2)$$

Where ~ is the NOT operator. This can be simplified as:

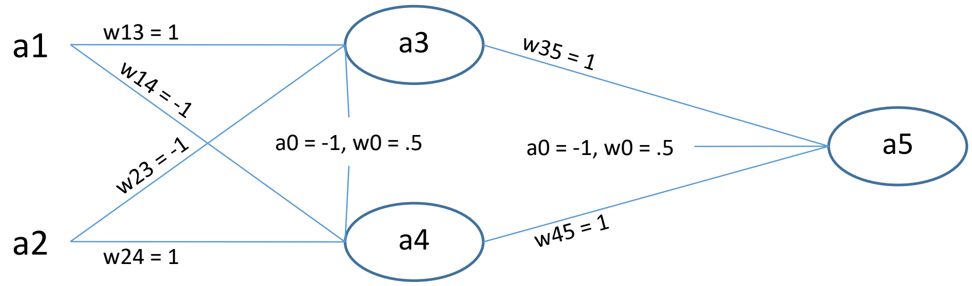
$$XOR = (a_1 * \sim a_2) + (\sim a_1 * a_2)$$

This simplified equation can be plotted as:



Notice how the x and y axis can never both be the same value, 0 or 1, hence there are no circles on those points.

I can then write the neural network for the XOR function as:



Where a1 and a2 are our inputs and the rest can be defined as:

$$a3 = a1 * w13 + a2 * w23 + a0 * w0$$

$$a4 = a1 * w14 + a2 * w24 + a0 * w0$$

$$a5 = a3 * w35 + a4 * w45 + a0 * w0$$

Each output cell (a3, a4, a5) produces either a 1 or zero based on if its values that it sums come out to a positive (1) or negative (0) value.

We can validate this neural network by running through the 4 combinations of a1 and a2 and comparing to the XOR function:

a1	**a2**	**a3**	**a4**	**a5**	**XOR**
0	0	0+0-.5=-.5 → 0	0+0-.5=-.5 → 0	0+0-.5=-.5 → 0	0
0	1	0-1-.5=-1.5 → 0	0+1-.5=.5 → 1	0+1-.5=.5 → 1	1
1	0	1+0-.5=.5 → 1	-1+0-.5=-1.5 → 0	1+0-.5=.5 → 1	1
1	1	1-1-.5=-.5 → 0	-1+1-.5=-.5 → 0	0+0-.5=-.5 → 0	0