Machine Learning INF2008

Lecture 05: Tree Based Models: Decision Trees

Donny Soh

Singapore Institute of Technology

Preamble: types of variables

- Ordinal versus Nominal variables.

Transmission	Car Type	Size	Price
Manual	Honda	1.5	150,000
Automatic	Toyota	2.4	200,000
Automatic	Nissan	1.6	155,000

- Ordinal numbers
  - natural numbers to describe a way to arrange different elements.

- Nominal numbers
  - numbers that are used for identifications. (eg NRIC number)

Transmission	Car Type	Size	Price
			150,000
			200,000
			155,000

Preamble: types of variables

- Ordinal versus Nominal variables.

Transmission	Car Type	Size	Price
Manual	Honda	1.5	150,000
Automatic	Toyota	2.4	200,000
Automatic	Nissan	1.6	155,000

- The transmission can be seen as a Boolean variable. True if it is automatic transmission and False if it is manual transmission

Transmission	Car Type	Size	Price
False			150,000
True			200,000
True			155,000

Preamble: types of variables

- Ordinal versus Nominal variables.

Transmission	Car Type	Size	Price
Manual	Honda	1.5	150,000
Automatic	Toyota	2.4	200,000
Automatic	Nissan	1.6	155,000

- The size of the car can be considered as an ordinal variable.

Transmission	Car Type	Size	Price
False		1.5	150,000
True		2.4	200,000
True		1.6	155,000

Preamble: types of variables

- Ordinal versus Nominal variables.

Transmission	Car Type	Size	Price
Manual	Honda	1.5	150,000
Automatic	Toyota	2.4	200,000
Automatic	Nissan	1.6	155,000

- How then do we represent the car type?
- What is the issue if we encode it as the following:
  - Honda: 0
  - Toyota: 1
  - Nissan: 2

Transmission	Car Type	Size	Price
False	0	1.5	150,000
True	1	2.4	200,000
True	2	1.6	155,000

Preamble: types of variables

- Ordinal versus Nominal variables.

Transmission	Car Type	Size	Price
Manual	Honda	1.5	150,000
Automatic	Toyota	2.4	200,000
Automatic	Nissan	1.6	155,000

- How then do we represent the car type?
- What is the issue if we encode it as the following:
  - Honda: 0
  - Toyota: 1
  - Nissan: 2

Transmission	Car Type	Size	Price
False	0	1.5	150,000
True	1	2.4	200,000
True	2	1.6	155,000

The inherent number may mislead most algorithms.

Preamble: types of variables

- How then do we represent the car type?
- We hence create three new variables:
  - car_type_honda
  - car_type_toyota
  - car_type_nissan

Transmission	car_type_honda	car_type_toyota	car_type_nissan	Size	Price
False	1	0	0	1.5	150,000
True	0	1	0	2.4	200,000
True	0	0	1	1.6	155,000

This is known as One-Hot Encoding

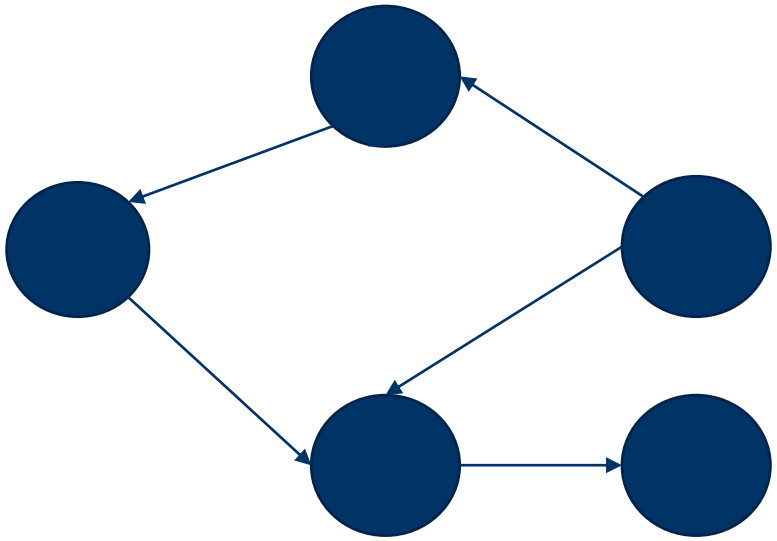
What is a Tree?

In graph theory a tree is an undirected, connected graph containing no cycles. The following conditions hold:

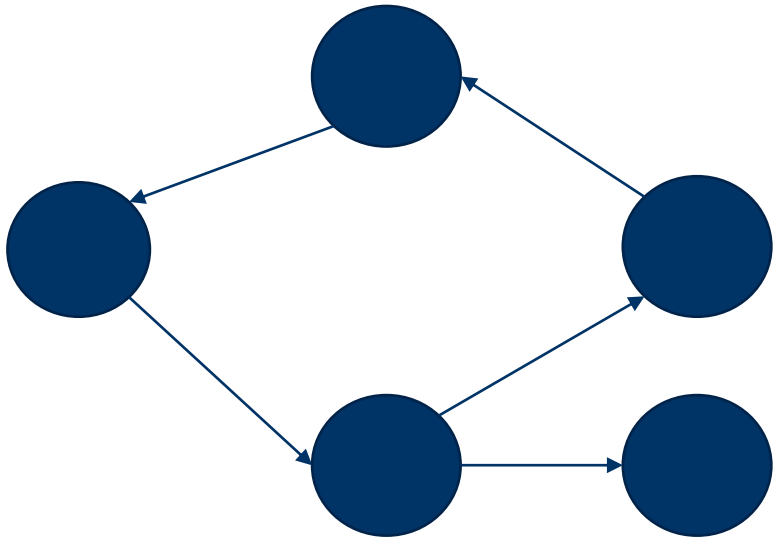
- The tree contains a single node called the root of the tree. Therefore, we say that node p is the parent of node u if we reach u from p after starting to traverse the tree from the selected root. Similarly, we say that u is a child of p. It's worth noting that we can choose multiple nodes as the root of the tree.

- Each node, except the root, must have a single parent. In other words, each node must be reached only from its parent when traversing the tree starting from the root.

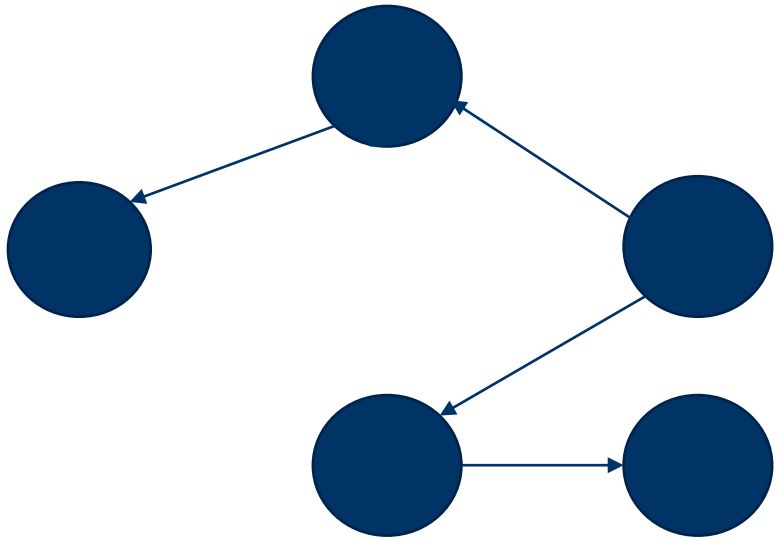
- Starting from the root, we must be able to visit all the nodes of the tree. Therefore, the tree should always be connected.



Is this graph a tree?



Is this graph a tree?

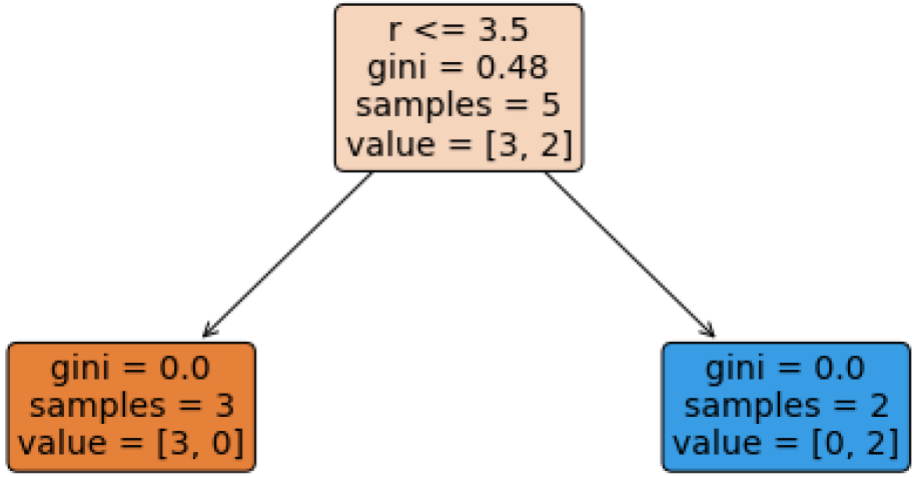


Is this graph a tree?

What is a Decision Stump?

There are a few terms one must takeaway from this simple decision tree stump. They are the following:

- The stump has a depth of one.
- There are two binary outcomes.
- There are two leaves.
- There are three nodes.



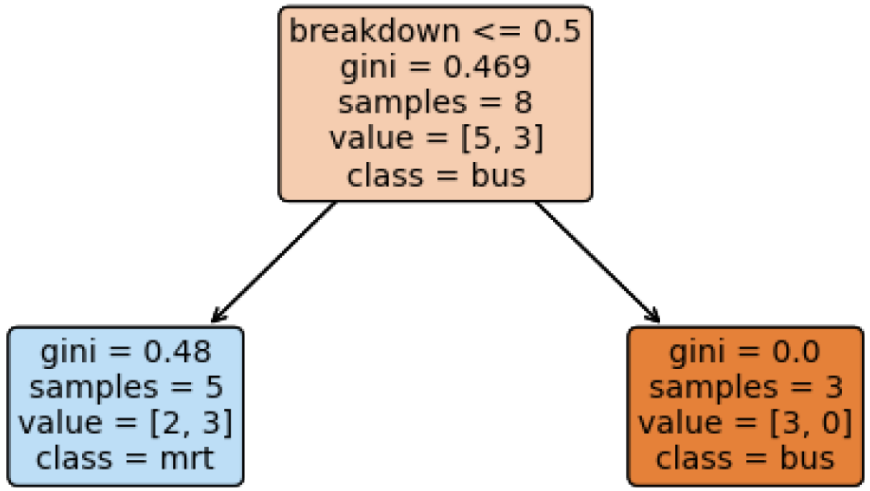
Running Example: How should I go to school today?

To decide on this, perhaps I will need to check if the MRT has broken down or not. If the MRT has broken down, I will take the bus. Else I will take the MRT.

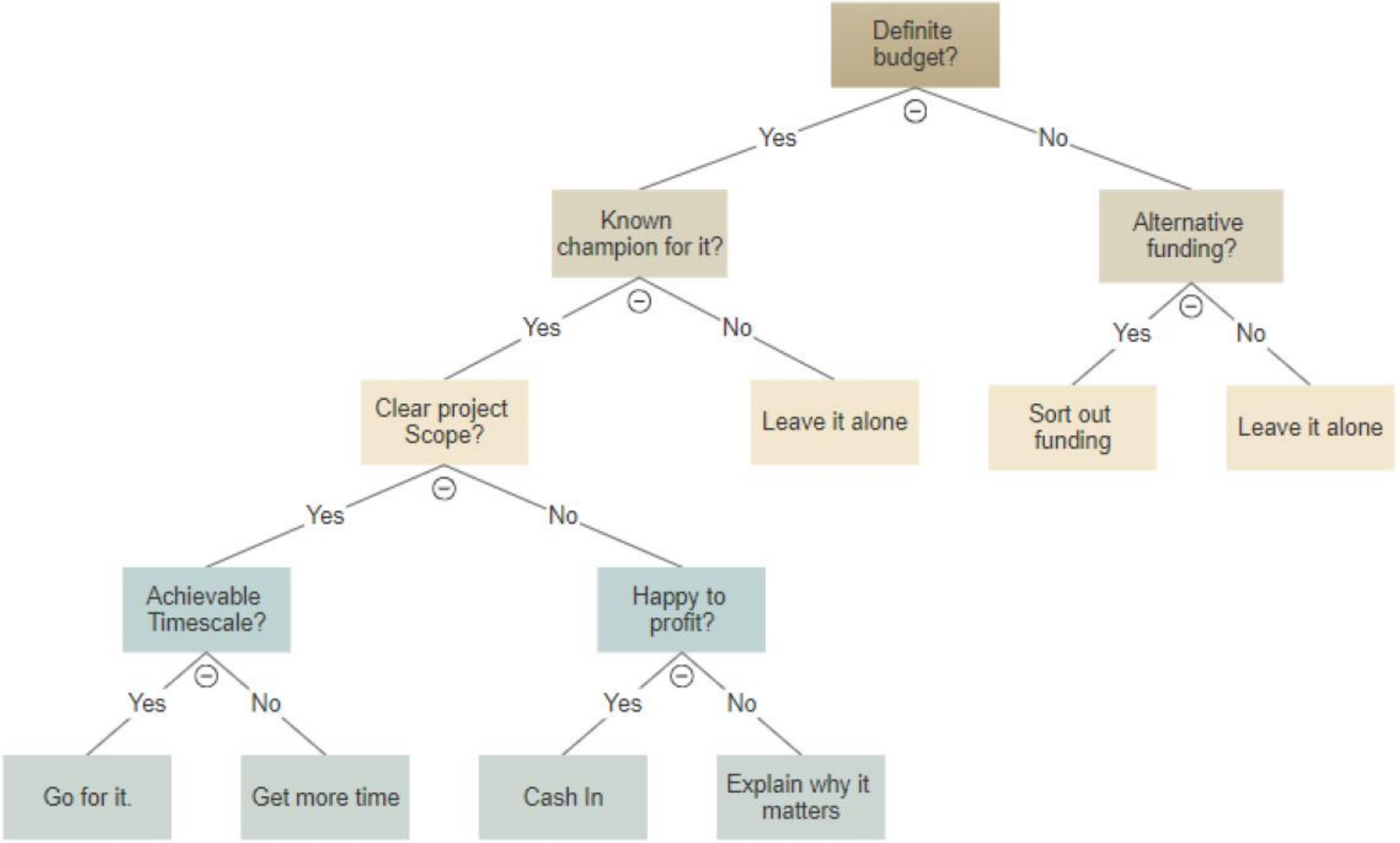
Discussion:

- Is this a regression or classification problem?
- What is the target variable?
- What are the possible values the target variable is able to take?
- What is the x-variable?
- What are the possible values the x-variable is able to take?

Num	Breakdown	Transport
0	1	Bus
1	1	Bus
2	1	Bus
3	0	MRT
4	0	MRT
5	0	Bus
6	0	Bus
7	0	MRT



What is a Decision Tree?

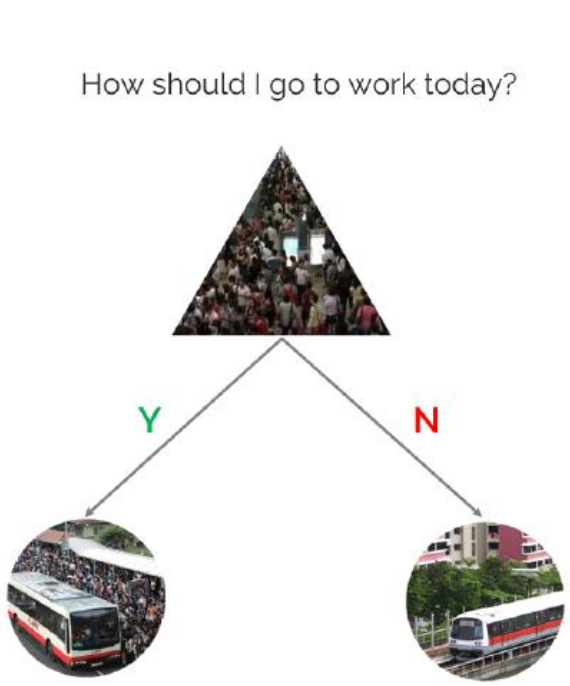


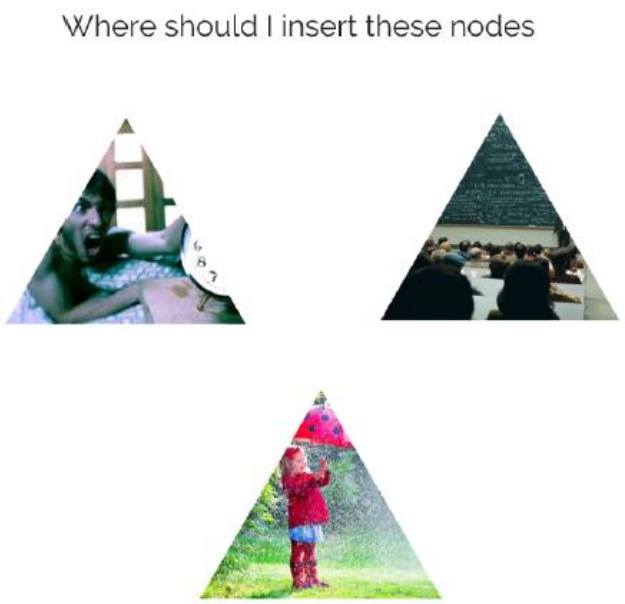
Running Example: How should I go to school today?

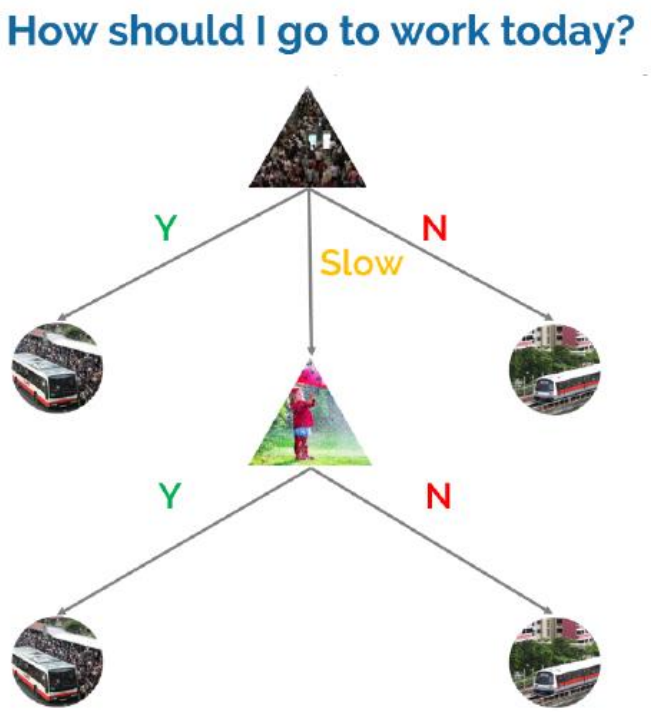
Suppose other than finding if the MRT has broken or not, we have the following other factors:

- Did I oversleep?
- Is there an important lecture?
- Is it raining?

Hence the challenge we face is taking the decision stump and all the new conditions (on the left), combining them together to form a decision tree on the right.



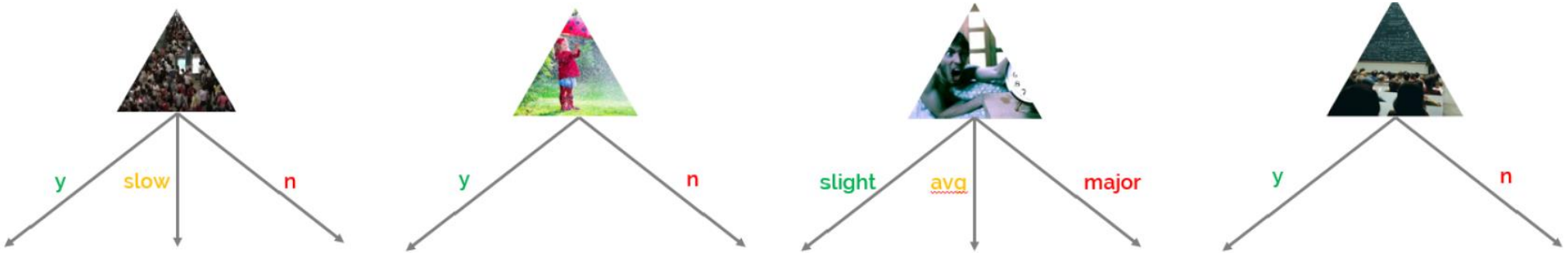




Get training data

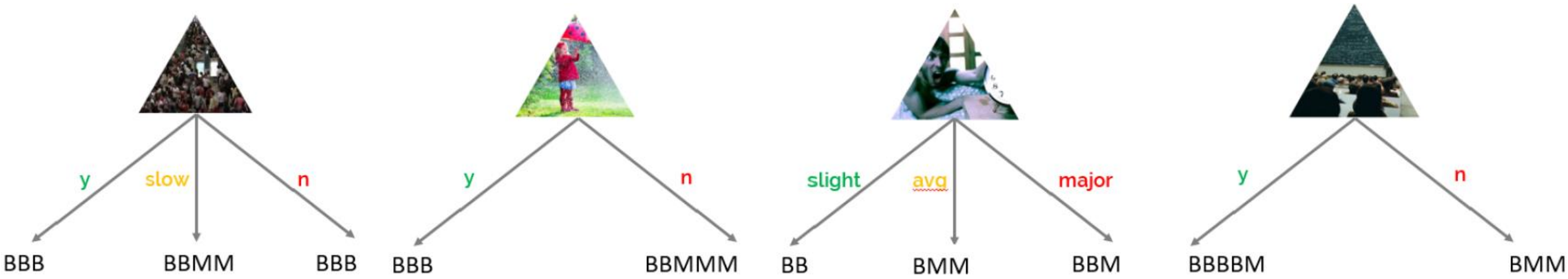
Being a great student, you then realize that in order to build this supervised learning model, you would first h
to collect sufficient data to train the model. Therefore for the next eight days, you start to collect the data of your travel habits. You tabulate them and end up with the following table:

Breakdown	Rain	Oversleep	ML Lecture	Bus / MRT
Slow	Y	Slight	Y	B
Yes	Y	Major	Y	B
Slow	N	Major	Y	M
No	N	Average	N	M
Slow	N	Average	N	B
Yes	N	Major	Y	B
Yes	N	Average	Y	B
Slow	Y	Slight	N	B



Building the tree

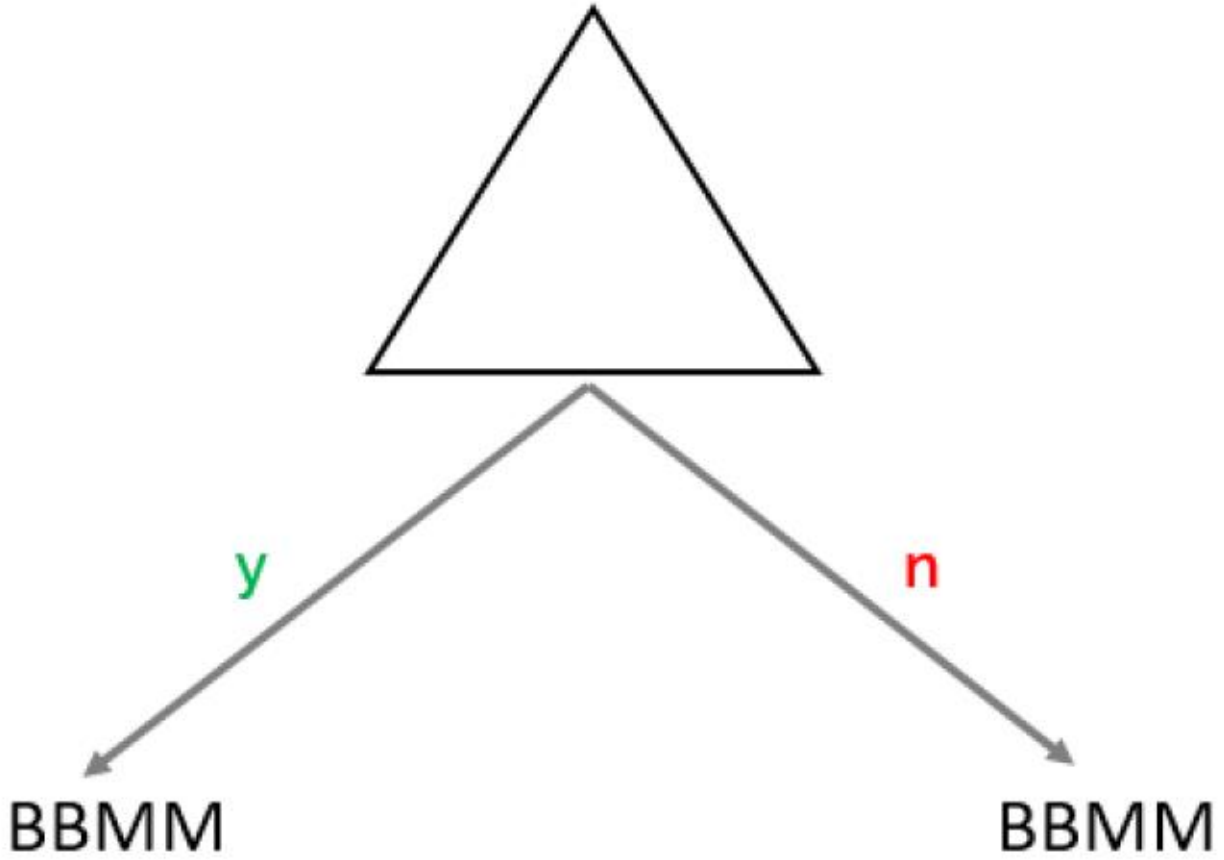
Breakdown	Rain	Oversleep	ML Lecture	Bus / MRT
Slow	Y	Slight	Y	B
Yes	Y	Major	Y	B
Slow	N	Major	Y	M
No	N	Average	N	M
Slow	N	Average	N	B
Yes	N	Major	Y	B
Yes	N	Average	Y	B
Slow	Y	Slight	N	B



Building the tree

Faced with these four decision stumps, the key issue will then be to pick the decision stump which best splits the data.

This would be a decision stump which gives the most unambigious splitting. For instance, this would be an example of a bad node to choose.

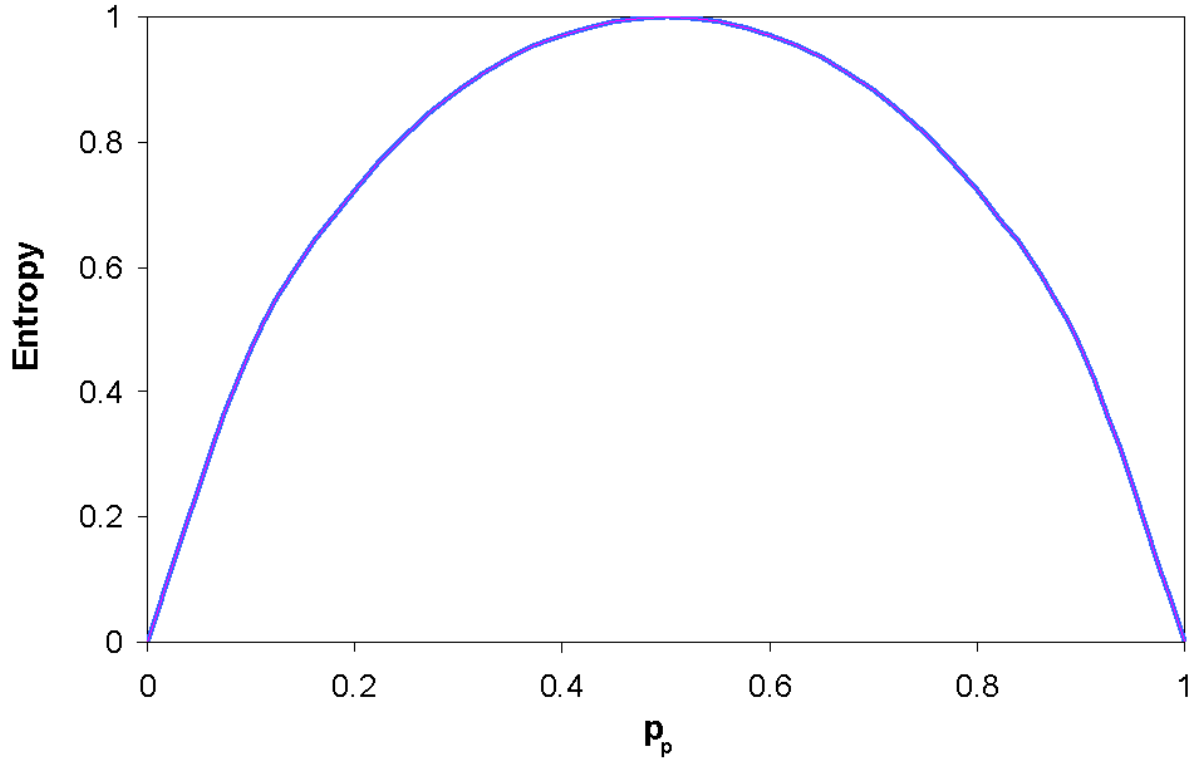


What is entropy?

The English definition of entropy is the lack of order.

By this definition, one can intuitive infer that the more random an event is, the ??????? the value of the entropy.

Hence the experiment of a coin would have the ?????? entropic value of ???????. How about the coin of DC villain / character Two-Face? What would the entropic value of his coin be?



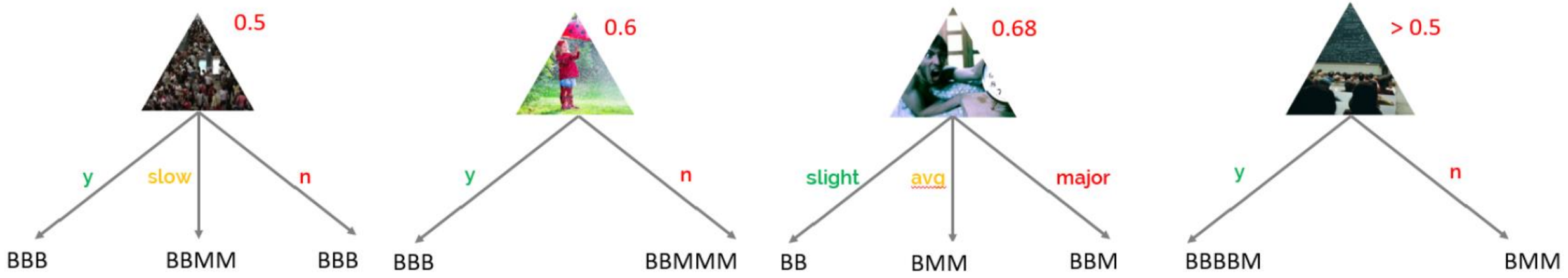
$$H = -\sum_i p_i(log_2 p_i)$$

This entropy follows certain interesting characteristics.
The slope of the graph goes up relatively fast.

- At p = 0.3, the entropy already exceeds 0.88
- At p = 0.2, the entropy already exceeds 0.72

Building the tree

Breakdown	Rain	Oversleep	ML Lecture	Bus / MRT
Slow	Y	Slight	Y	B
Yes	Y	Major	Y	B
Slow	N	Major	Y	M
No	N	Average	N	M
Slow	N	Average	N	B
Yes	N	Major	Y	B
Yes	N	Average	Y	B
Slow	Y	Slight	N	B



Building the tree: second level

We then select the decision stump with the lowest entropy value as the root node.

Hence in this case, x-variable if the MRT broke down or not is chosen as the root node. We are then left with three other decision stumps and the next step is to select the next best node.

The same actions are taken.

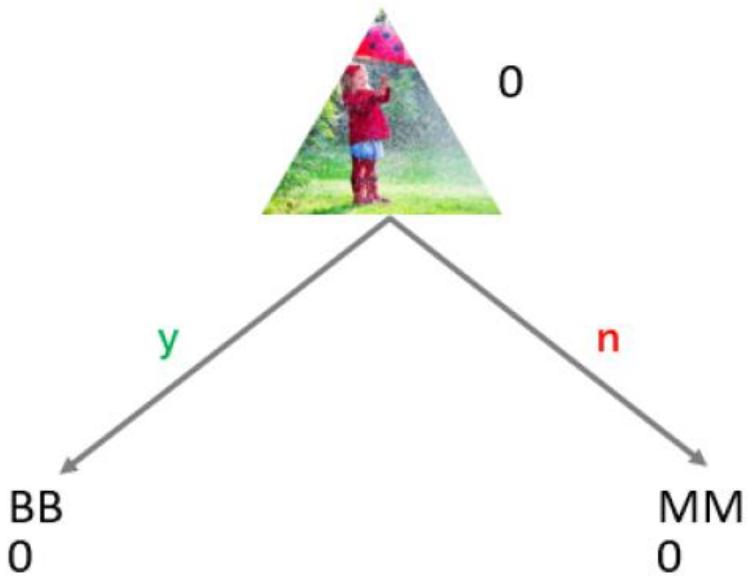
Firstly we consider only the four observations that are still incorrectly classified from the root node.

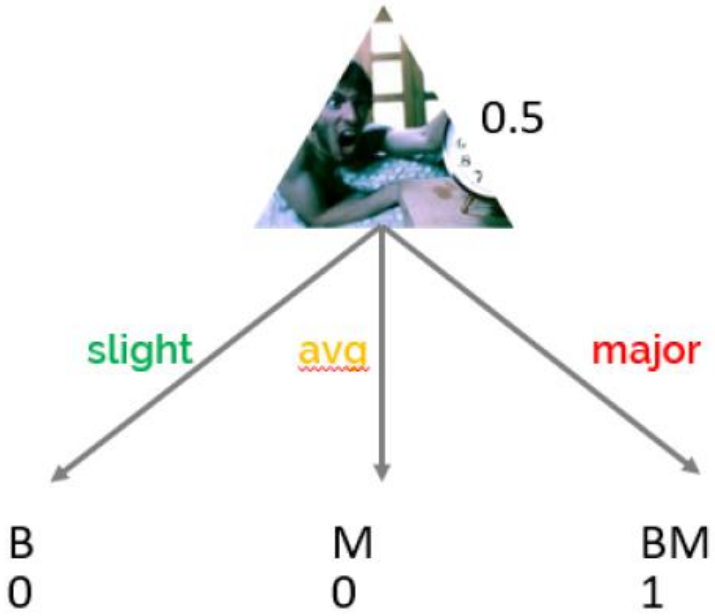
The four observations are:

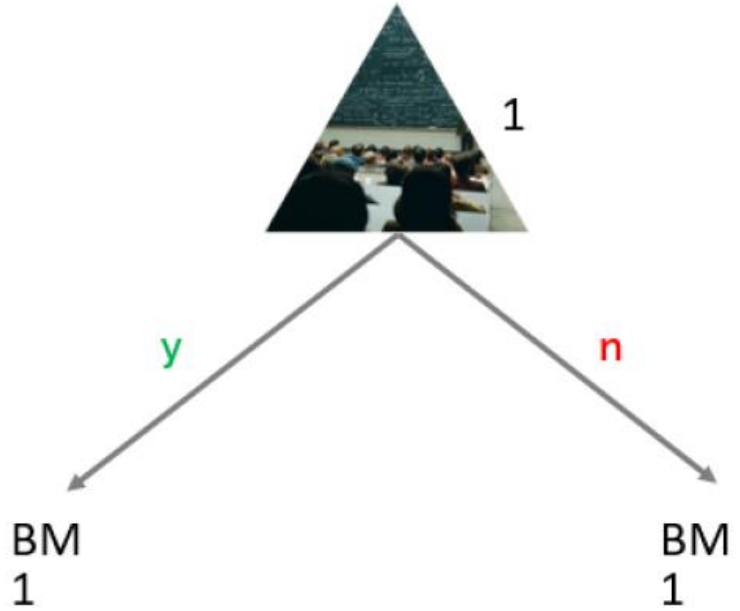
Breakdown	Rain	Oversleep	ML Lecture	Bus / MRT
Slow	Y	Slight	Y	B
Slow	N	Major	Y	M
Slow	N	Average	N	B
Slow	Y	Slight	N	B

Building the tree: second level

Breakdown	Rain	Oversleep	ML Lecture	Bus / MRT
Slow	Y	Slight	Y	B
Slow	N	Major	Y	M
Slow	N	Average	N	B
Slow	Y	Slight	N	B







Completing the tree