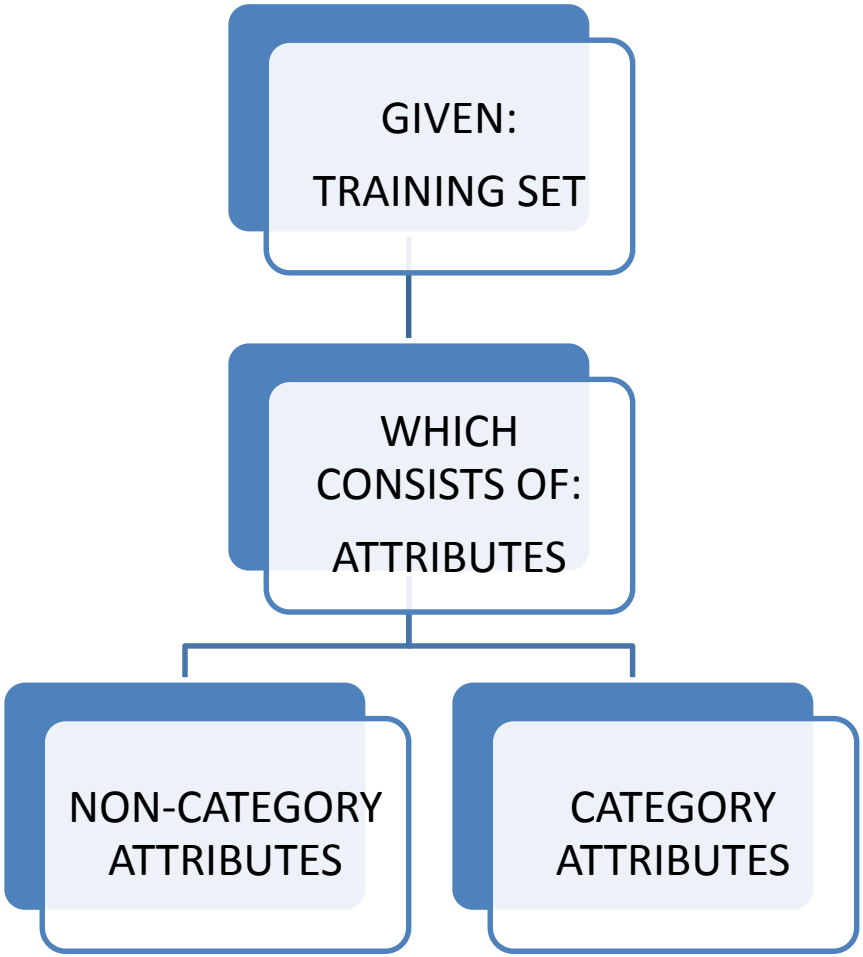
ID3 ALGORITHM

ID3 (Iterative Dichotomiser 3): Basic Idea

- Invented by J.Ross Quinlan in 1975.
- Used to generate a decision tree from a given data set by employing a top-down, greedy search, to test each attribute at every node of the tree.
- The resulting tree is used to classify future samples.

Introduction



Attribute				Classes
Gender	**Car Ownership**	**Travel Cost**	**Income Level**	**Transportation**
Male	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Female	1	Cheap	Medium	Train
Female	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Male	0	Standard	Medium	Train
Female	1	Standard	Medium	Train
Female	1	Expensive	High	Car
Male	2	Expensive	Medium	Car
Female	2	Expensive	High	Car

GIVEN TRAINING SET

We use non-category attributes to predict the values of category attributes.

ALGORITHM

- Calculate the entropy of every attribute using the data set
- Split the set into subsets using the attribute for which entropy is minimum (or, equivalently, information gain is maximum)
- Make a decision tree node containing that attribute
- Recurse on subsets using remaining attributes

Entropy

- In order to define information gain precisely, we need to discuss entropy first.
- A formula to calculate the homogeneity of a sample.
- A completely homogeneous sample has entropy of 0 (leaf node).
- An equally divided sample has entropy of 1.
- The formula for entropy is:

$$\text{Entropy}(S) = -p(I) \log_2 p(I)$$

- where p(I) is the proportion of S belonging to class I. ∑ is over total outcomes. Log2 is log base 2.

Example 1

- If S is a collection of 14 examples with 9 YES and 5 NO examples then
- Entropy(S) = - (9/14) Log2 (9/14) - (5/14) Log2 (5/14) = 0.940

Information Gain (IG)

- The information gain is based on the decrease in entropy after a dataset is split on an attribute.
- The formula for calculating information gain is:

$$Gain(S, A) = Entropy(S) - ((|S_v| / |S|) * Entropy(S_v))$$

Where:

- S_v = subset of S for which attribute A has value v
- $|S_v|$ = number of elements in S_v
- $|S|$ = number of elements in S

PROCEDURE

- First the entropy of the total dataset is calculated.
- The dataset is then split on the different attributes.
- The entropy for each branch is calculated. Then it is added proportionally, to get total entropy for the split.
- The resulting entropy is subtracted from the entropy before the split.
- The result is the Information Gain, or decrease in entropy.
- The attribute that yields the largest IG is chosen for the decision node.

EXAMPLE

Attribute				Classes
Gender	Car Ownership	Travel Cost	Income Level	Transportation
Male	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Female	1	Cheap	Medium	Train
Female	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Male	0	Standard	Medium	Train
Female	1	Standard	Medium	Train
Female	1	Expensive	High	Car
Male	2	Expensive	Medium	Car
Female	2	Expensive	High	Car

Probability :

-Bus → 4/10 = 0.4

-Train → 3/10 = 0.3

-Car → 3/10 = 0.3

Impurity using entropy :

$$E(S) = \sum - p(I) \log_2 p(I)$$

Entropy

– 0.4 log (0.4,2) – 0.3 log (0.3,2) – 0.3 log (0.3,2) = 1.571

Attribute				Classes
Gender	Car Ownership	Travel Cost	Income Level	Transportation
Male	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Female	1	Cheap	Medium	Train
Female	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Male	0	Standard	Medium	Train
Female	1	Standard	Medium	Train
Female	1	Expensive	High	Car
Male	2	Expensive	Medium	Car
Female	2	Expensive	High	Car

Attribute	Classes
Gender	Transportation
Male	Bus
Male	Bus
Female	Train
Female	Bus
Male	Bus
Male	Train
Female	Train
Female	Car
Male	Car
Female	Car

Attribute	Classes
Gender	Transportation
Male	Bus
Male	Bus
Male	Bus
Male	Train
Male	Car

Probability :
Bus : 3/5 = 0.6
Train : 1/5= 0.2
Car : 1/5 = 0.2
Entropy→ 1.522

Attribute	Classes
Gender	Transportation
Female	Train
Female	Bus
Female	Train
Female	Car
Female	Car

Probability :
Bus : 1/5 = 0.2
Train : 2/5= 0.4
Car : 2/5 = 0.4
Entropy → 1.371

Information Gain →
*1.571 – (((5/10)*1.522)+((5/10)*1.371)) = 0.12*

Attribute				Classes
Gender	Car Ownership	Travel Cost	Income Level	Transportation
Male	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Female	1	Cheap	Medium	Train
Female	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Male	0	Standard	Medium	Train
Female	1	Standard	Medium	Train
Female	1	Expensive	High	Car
Male	2	Expensive	Medium	Car
Female	2	Expensive	High	Car

Attribute	Classes
Car Ownership	Transportation
0	Bus
1	Bus
1	Train
0	Bus
1	Bus
0	Train
1	Train
1	Car
2	Car
2	Car

Attribute	Classes
Car Ownership	Transportation
0	Bus
0	Bus
0	Train

Attribute	Classes
Car Ownership	Transportation
1	Bus
1	Train
1	Bus
1	Train
1	Car

Attribute	Classes
Car Ownership	Transportation
2	Car
2	Car

Information Gain →
*1.571 – (((3/10)*0.918)+((5/10)*1.522)+((2/10)*0)) = 0.534*

Attributes				Classes
Gender	Car Ownership	Travel Cost	Income Level	Transportation
Male	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Female	1	Cheap	Medium	Train
Female	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Male	0	Standard	Medium	Train
Female	1	Standard	Medium	Train
Female	1	Expensive	High	Car
Male	2	Expensive	Medium	Car
Female	2	Expensive	High	Car

Attribute	Classes
Travel Cost	Transportation
Cheap	Bus
Cheap	Bus
Cheap	Train
Cheap	Bus
Cheap	Bus
Standard	Train
Standard	Train
Expensive	Car
Expensive	Car
Expensive	Car

Attribute	Classes
Travel Cost	Transportation
Cheap	Bus
Cheap	Bus
Cheap	Train
Cheap	Bus
Cheap	Bus

Attribute	Classes
Travel Cost	Transportation
Standard	Train
Standard	Train

Attribute	Classes
Travel Cost	Transportation
Expensive	Car
Expensive	Car
Expensive	Car

Information Gain →
*1.571 – (((5/10)*0.722)+((2/10)*0)+((3/10)*0)) = 1.21*

Attribute				Classes
Gender	Car Ownership	Travel Cost	Income Level	Transportation
Male	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Female	1	Cheap	Medium	Train
Female	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Male	0	Standard	Medium	Train
Female	1	Standard	Medium	Train
Female	1	Expensive	High	Car
Male	2	Expensive	Medium	Car
Female	2	Expensive	High	Car

Attribute	Classes
Income Level	Transportation
Low	Bus
Medium	Bus
Medium	Train
Low	Bus
Medium	Bus
Medium	Train
Medium	Train
High	Car
Medium	Car
High	Car

Attribute	Classes
Income Level	Transportation
Low	Bus
Low	Bus

Attribute	Classes
Income Level	Transportation
Medium	Bus
Medium	Train
Medium	Bus
Medium	Train
Medium	Train
Medium	Car

Attribute	Classes
Income Level	Transportation
High	Car
High	Car

Information Gain →

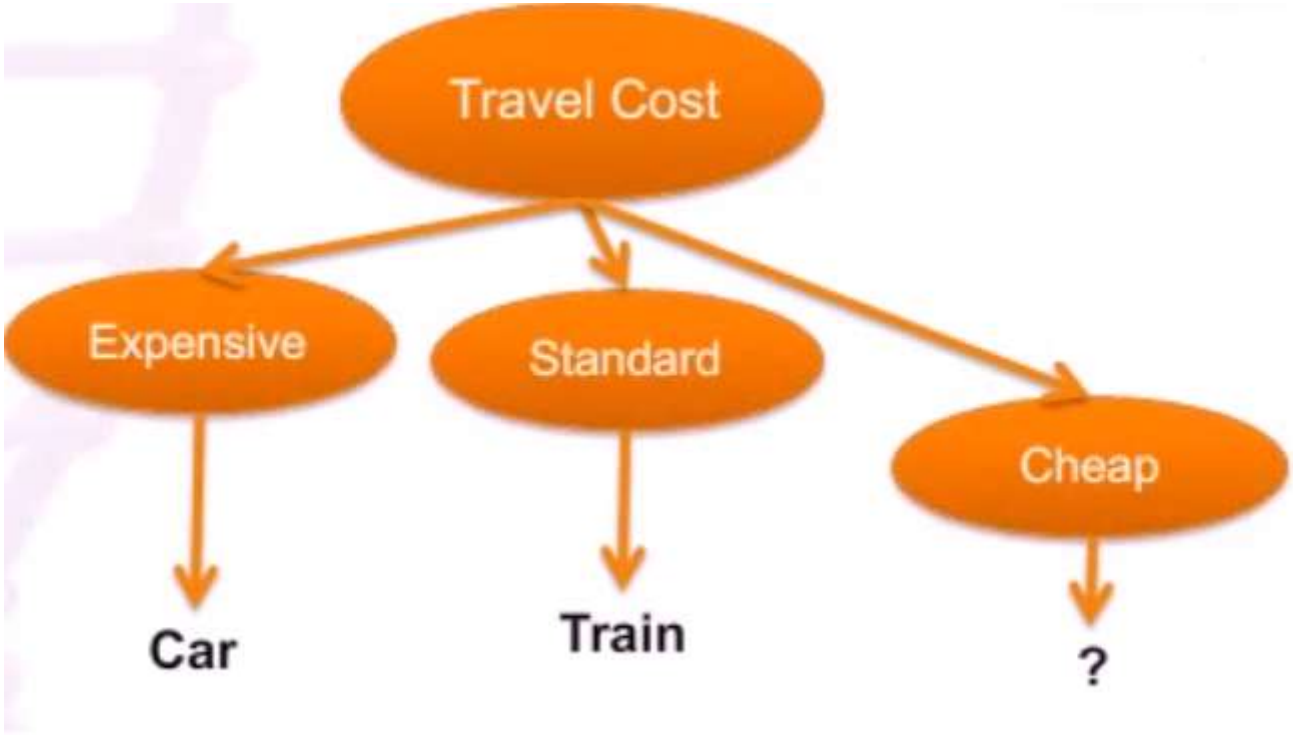
*1.571 – (((2/10)*0)+((6/10)*1.459)+((2/10)*0)) = 0.695*

Attribute	*Information Gain*
Gender	0.125
Car	0.534
Travel cost	1.21
Income Level	0.695

Attribute				Classes
Travel Cost	**Gender**	**Car Ownership**	**Income Level**	**Transportation**
Cheap	Male	0	Low	Bus
Cheap	Male	1	Medium	Bus
Cheap	Female	1	Medium	Train
Cheap	Female	0	Low	Bus
Cheap	Male	1	Medium	Bus

Attribute				Classes
Travel Cost	**Gender**	**Car Ownership**	**Income Level**	**Transportation**
Standard	Male	0	Medium	Train
Standard	Female	1	Medium	Train

Attribute				Classes
Travel Cost	**Gender**	**Car Ownership**	**Income Level**	**Transportation**
Expensive	Female	1	High	Car
Expensive	Male	2	Medium	Car
Expensive	Female	2	High	Car


Travel Cost
Expensive
Standard
Cheap
Car
Train
?

Attribute			Classes
Gender	**Car Ownership**	**Income Level**	**Transportation**
Male	0	Low	Bus
Male	1	Medium	Bus
Female	1	Medium	Train
Female	0	Low	Bus
Male	1	Medium	Bus

Probability :
Bus → 4/5 = 0.8
Train → 1/5 = 0.2
Entropy → 0.722

Attribute			Classes
Gender	Car Ownership	Income Level	Transportation
Male	0	Low	Bus
Male	1	Medium	Bus
Female	1	Medium	Train
Female	0	Low	Bus
Male	1	Medium	Bus

Attribute	Classes
Gender	Transportation
Male	Bus
Male	Bus
Female	Train
Female	Bus
Male	Bus

Attribute	Classes
Gender	Transportation
Male	Bus
Male	Bus
Male	Bus

Probability :
Bus : 3/3 = 1
Entropy → 0

Attribute	Classes
Gender	Transportation
Female	Train
Female	Bus

Probability :
Bus : 1/2 = 0.5
Train : 1/2 = 0.5
Entropy → 1

Information Gain →
0.722 – (((3/5)*0)+((2/5)*1)) = 0.322

Attribute			Classes
Gender	Car Ownership	Income Level	Transportation
Male	0	Low	Bus
Male	1	Medium	Bus
Female	1	Medium	Train
Female	0	Low	Bus
Male	1	Medium	Bus

Attribute	Classes
Car Ownership	Transportation
0	Bus
1	Bus
1	Train
0	Bus
1	Bus

Attribute	Classes
Car Ownership	Transportation
0	Bus
0	Bus

Probability :
Bus : 2/2 = 1
Entropy → 0

Attribute	Classes
Car Ownership	Transportation
1	Bus
1	Train
1	Bus

Probability :
Bus : 2/3 = 0.67
Train : 1/3 = 0.33
Entropy → 0.918

Information Gain →
*0.722 – (((2/5)*0)+((3/5)*0.9[illegible])*0.[illegible]1)*

Attribute			Classes
Gender	Car Ownership	Income Level	Transportation
Male	0	Low	Bus
Male	1	Medium	Bus
Female	1	Medium	Train
Female	0	Low	Bus
Male	1	Medium	Bus

Attribute	Classes
Income Level	Transportation
Low	Bus
Medium	Bus
Medium	Train
Low	Bus
Medium	Bus

Attribute	Classes
Income Level	Transportation
Low	Bus
Low	Bus

Probability :
Bus : 2/2 = 1
Entropy → 0

Attribute	Classes
Income Level	Transportation
Medium	Bus
Medium	Train
Medium	Bus

Probability :
Bus : 2/3 = 0.67
Train : 1/3 = 0.33
Entropy → 0.918

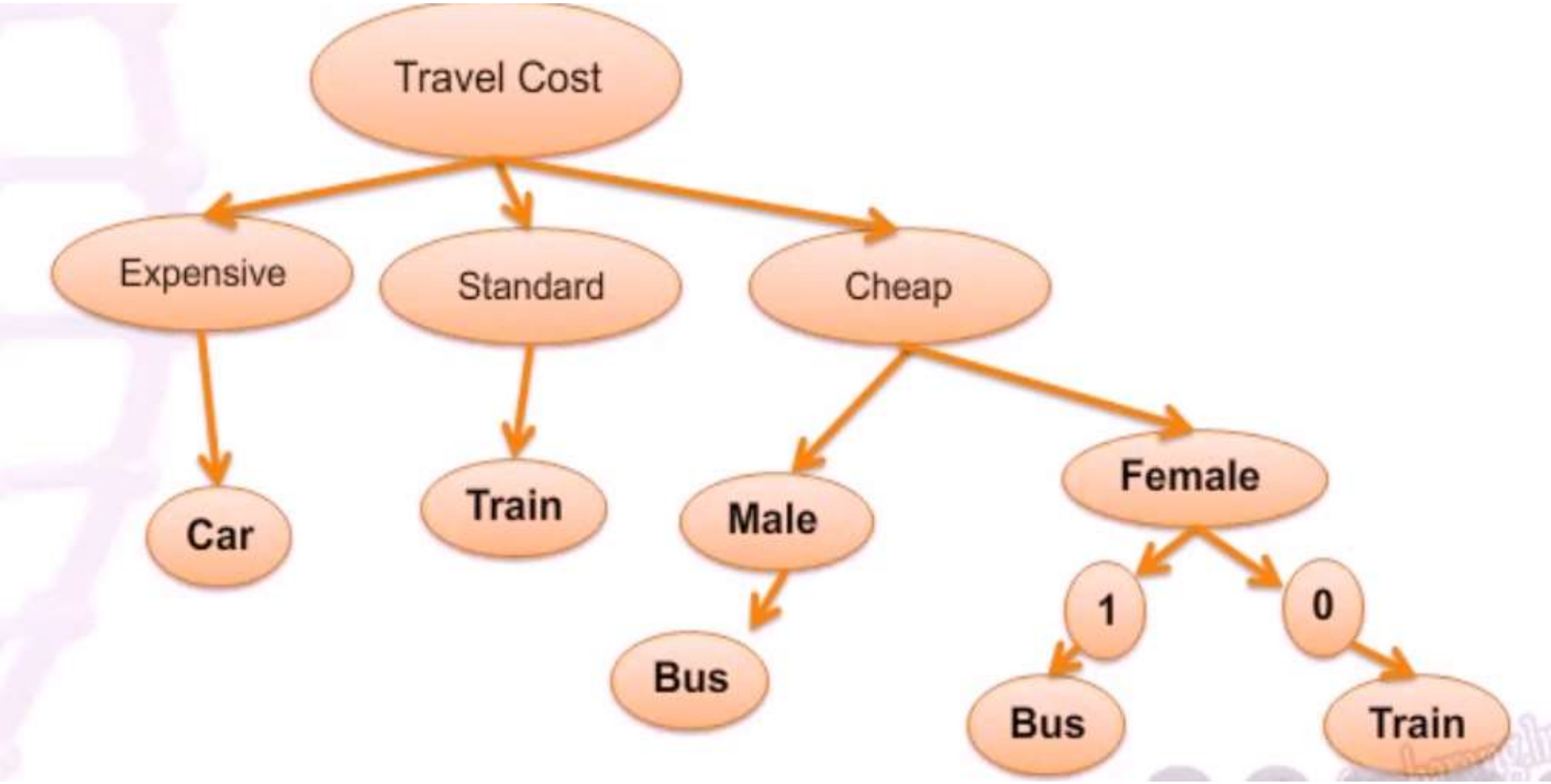
Information Gain →
0.722 – (((2/5)*0)+((3/5)*0.918)) = 0.171

Attribute	Information Gain
Gender	0.322
Car	0.171
Income Level	0.171

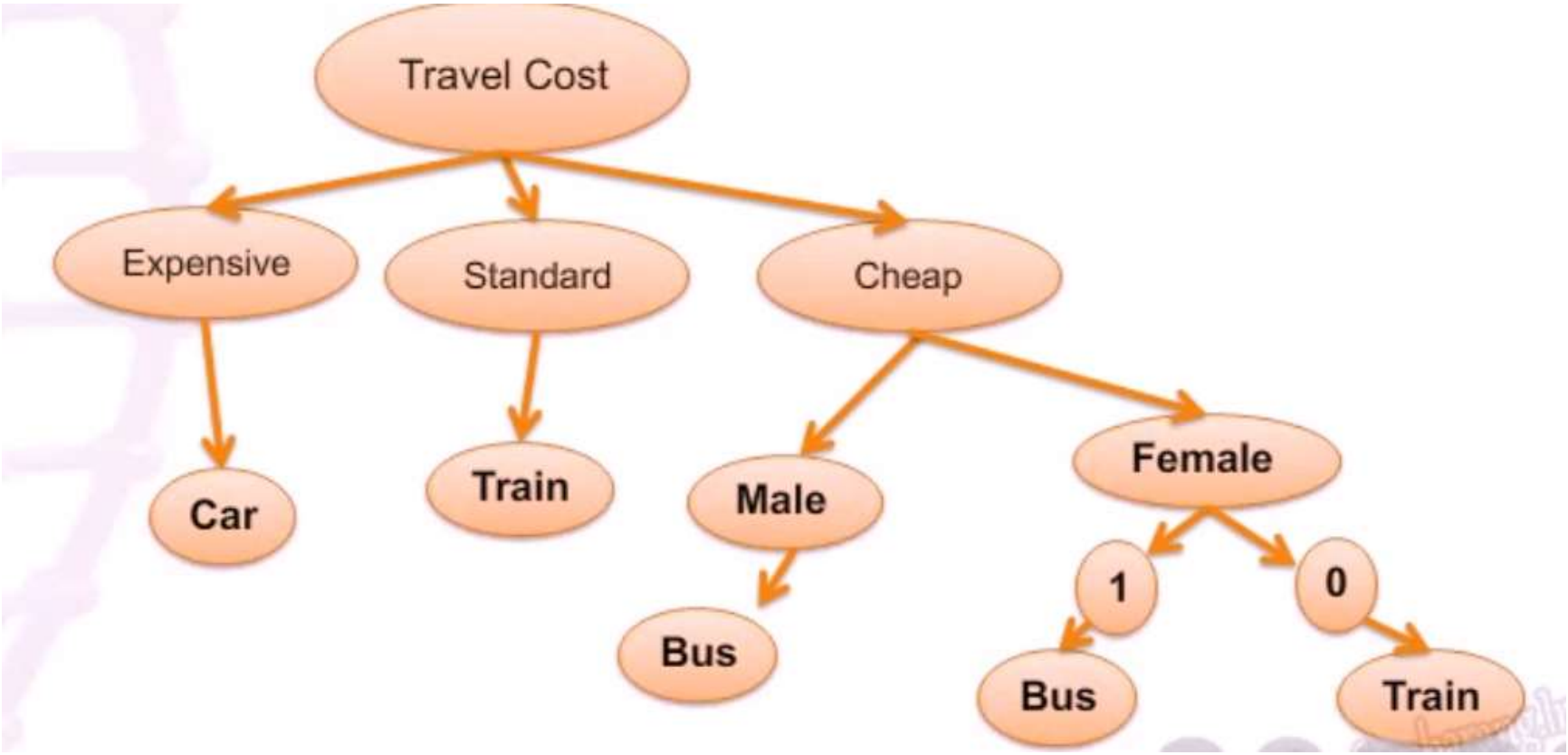
Attribute			Classes
Gender	Car Ownership	Income Level	Transportation
Male	0	Low	Bus
Male	1	Medium	Bus
Female	1	Medium	Train
Female	0	Low	Bus
Male	1	Medium	Bus

Attribute			Classes
Gender	Car Ownership	Income Level	Transportation
Male	0	Low	Bus
Male	1	Medium	Bus
Male	1	Medium	Bus

Attribute			Classes
Gender	Car Ownership	Income Level	Transportation
Female	1	Medium	Train
Female	0	Low	Bus


Travel Cost
Expensive
Standard
Cheap
Car
Train
Male
Female
Bus
1
0
Bus
Train

Name	Gender	Car ownership	Travel Cost	Income Level	Transportation
Alex	Male	1	Standard	High	?
Buddy	Male	0	Cheap	Medium	?
Cherry	Female	1	Cheap	High	?



Name	Gender	Car ownership	Travel Cost	Income Level	Transportation
Alex	Male	1	*Standard*	High	*Train*
Buddy	**Male**	0	*Cheap*	Medium	*Bus*
Cherry	**Female**	**1**	*Cheap*	High	*Bus*

Advantages of using ID3

- Understandable prediction rules are created from the training data.
- Builds the fastest tree.
- Builds a short tree.
- Only need to test enough attributes until all data is classified.
- Finding leaf nodes enables test data to be pruned, reducing number of tests.
- Whole dataset is searched to create tree.

Disadvantages of using ID3

- Data may be over-fitted or over-classified, if a small sample is tested.
- Only one attribute at a time is tested for making a decision.
- Classifying continuous data may be computationally expensive, as many trees must be generated to see where to break the continuum.