

Machine Learning for Finance

Prepared by Dr. Diaa Salama

FNT633
Decision Tree

Lecture 5



Machine Learning for Finance

Lecture

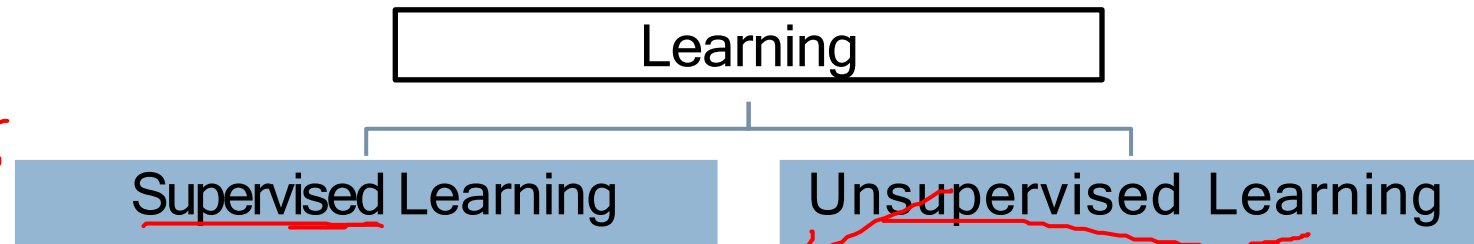
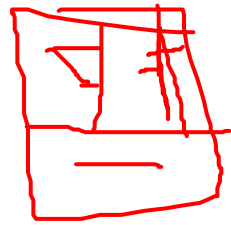
FNT633
Decesion Tree

Prepared by
Dr. Diaa Salama

www.midocean.ae

- Classification
- Decision Tree
- Classification: Basic Concepts
- Decision Tree Induction
- Prediction using classification model
- Model Evaluation and Selection
- Summary

Supervised vs. Unsupervised Learning



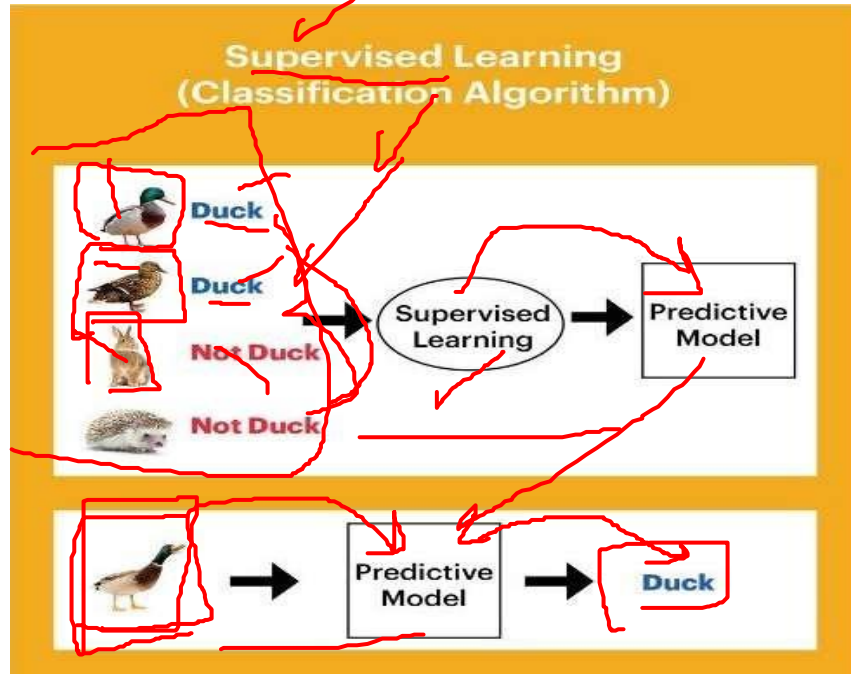
□ Supervised learning (classification)

- ▮ The class labels and the number of classes of training data is known.
- ▮ New data is classified based on the training set.

□ Unsupervised learning (clustering)

- ▮ The class labels of training data is unknown.
- ▮ Given a set of measurements, observations, etc.... with the aim of establishing the existence of classes or clusters in the data.

Example: Supervised vs. Unsupervised



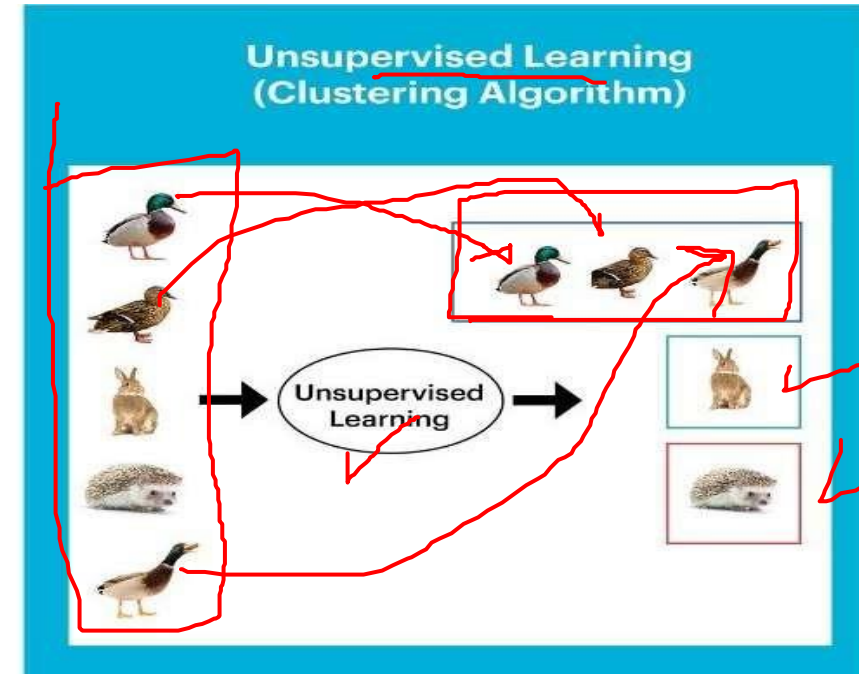
Class labels

Number of classes

✓✓

X : Unknown in advance

✓ : Known in advance



Class labels X

Number of classes ✓ or

X

Prediction Problems: Classification vs. Regression

- Classification and Regression are two major types of prediction problems.

- Classification:

Predicts categorical class labels (discrete or nominal).

Classifies data (constructs a model) based on the training set and the values (class labels) in a classifying attribute and uses it in classifying new data.

- Regression:

Models continuous-valued functions, i.e., predicts unknown or missing values.



Activity: (classification or regression?)

For some loan application data → predict 'safe' or 'risky'.

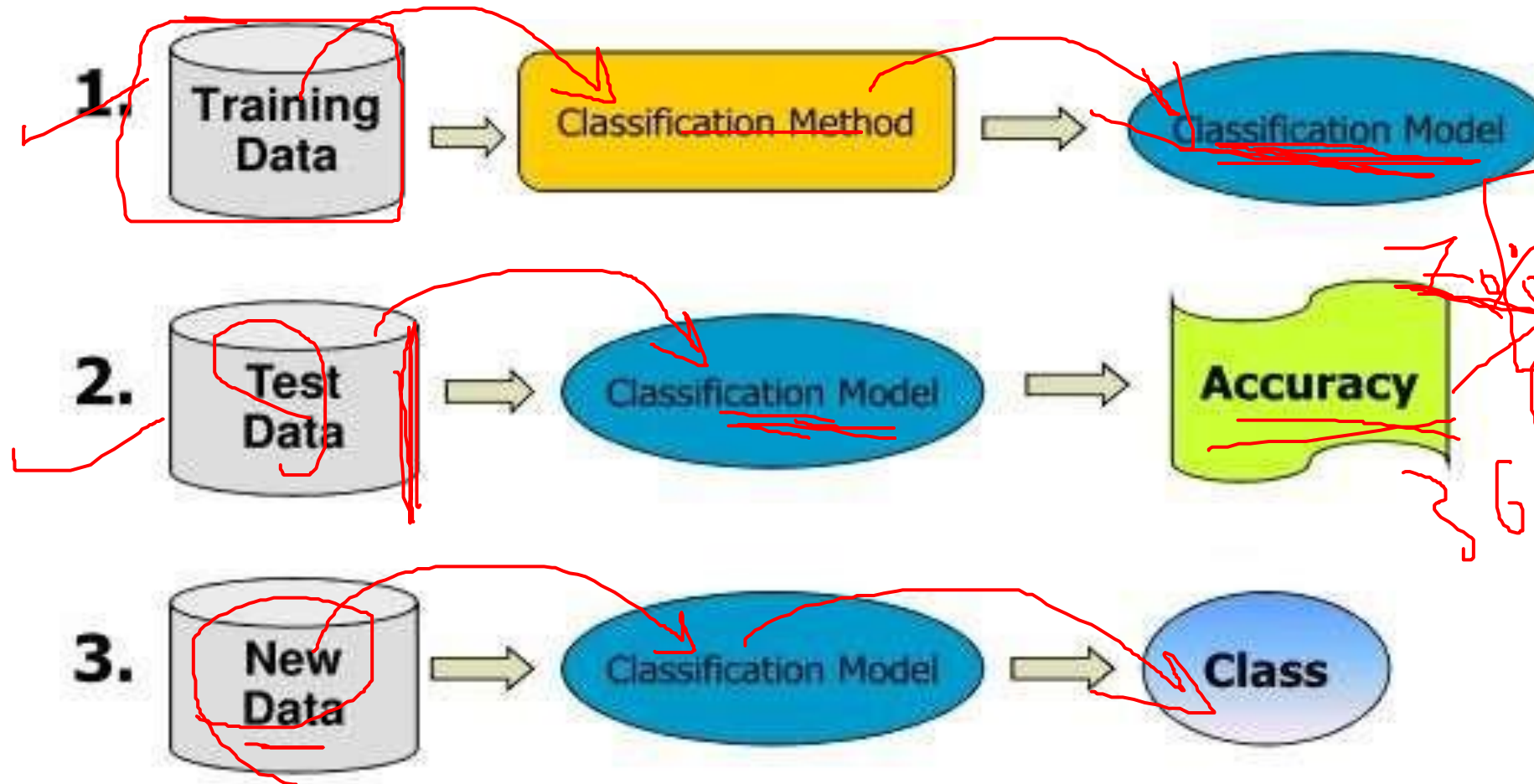
For a marketing manager → predict how much a given customer will spend during a sale.



Typical applications of classification

- Credit/loan approval: if a customer should be approved or not.
- Medical diagnosis: if a tumor is malignant or benign.
- Fraud detection: if a transaction is fraudulent.

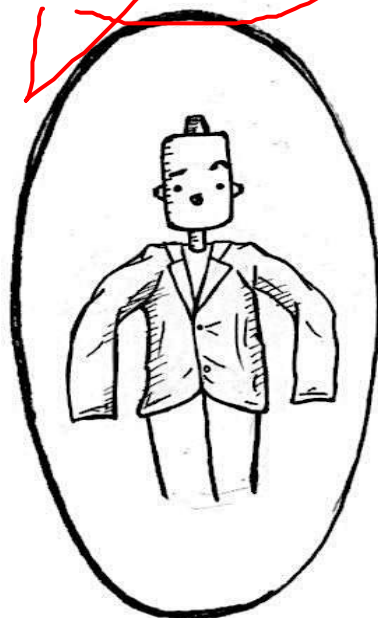
Classification Process: the 3 steps



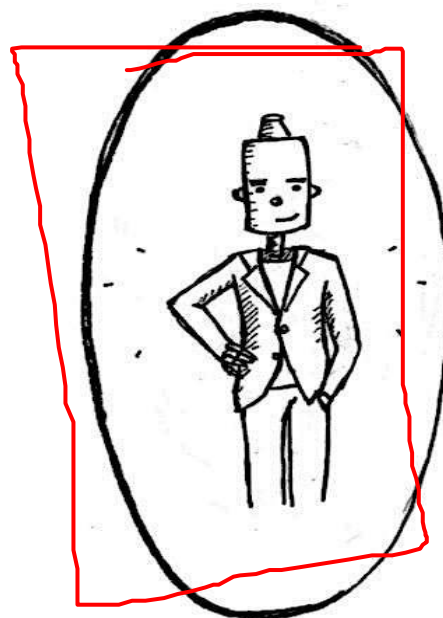
MACHINE LEARNING GENERALIZATION

FINDING THE PERFECT FIT

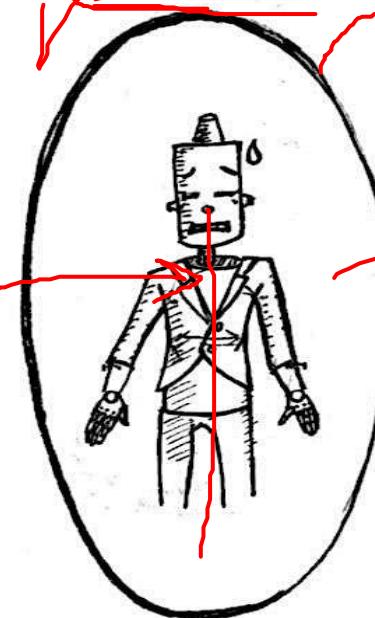
UNDERFIT



GOLDILOCKS ZONE



OVERFIT

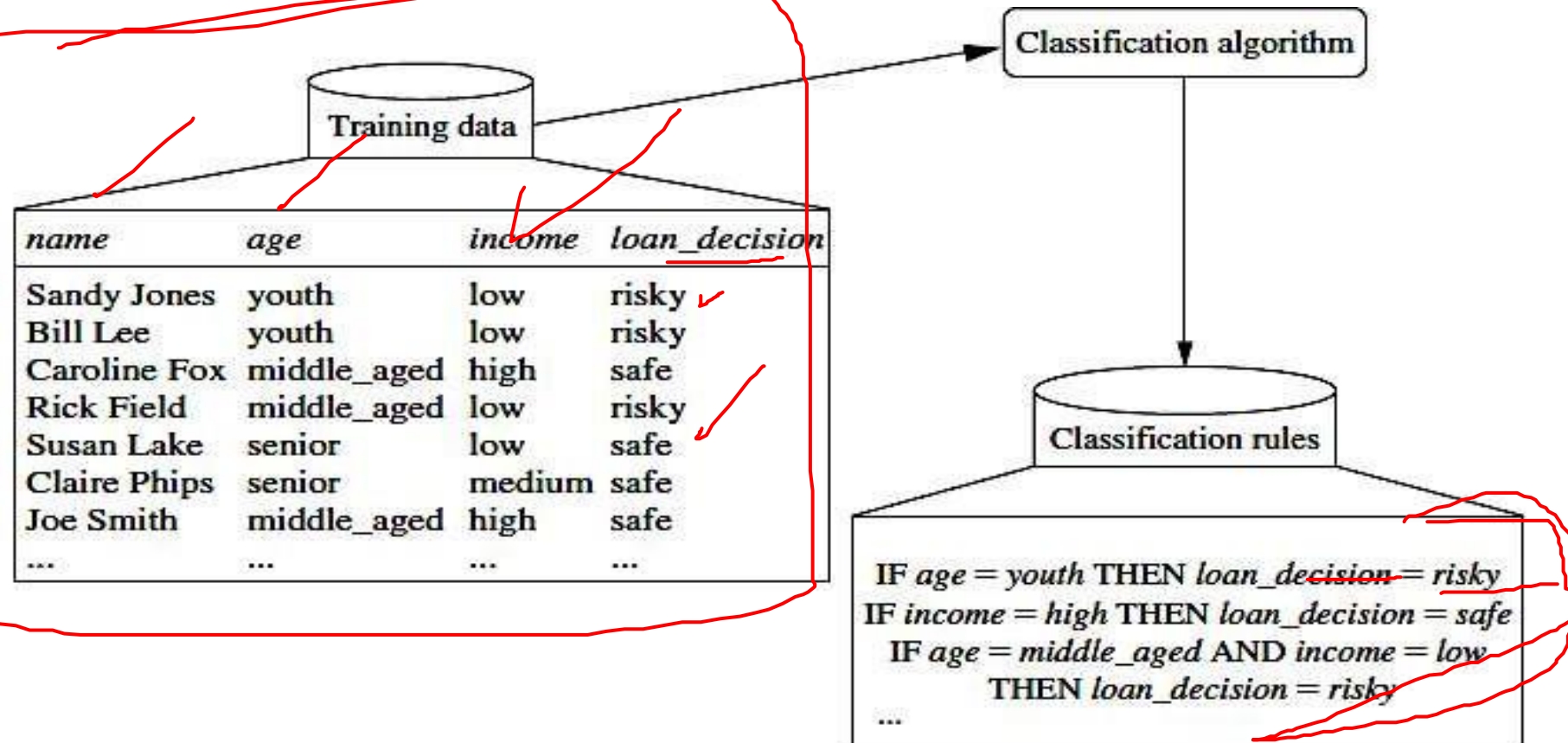


26 3.9

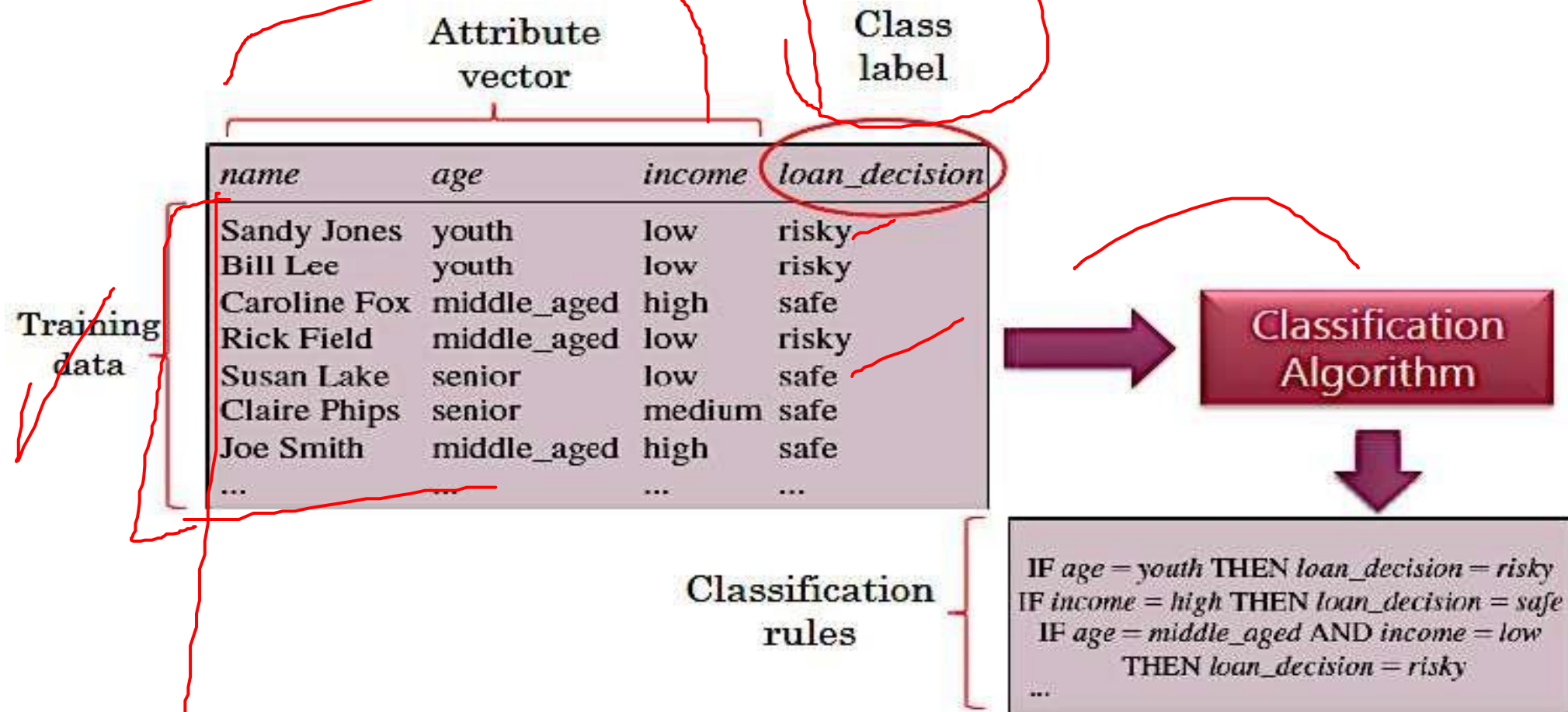
Step 1: Model construction (Learning)

- Each tuple is assumed to belong to a predefined class, as determined by one of the attributes, called the class label.
- The set of all tuples used for construction of the model is called training set.
- The model is represented in the following forms:
 1. Classification rules, (IF-THEN statements). ✓
 2. Decision tree. ✓
 3. Mathematical formula. ✓

Step 1: Model construction (Learning)



THE BASICS GENERAL APPROACH



Step 2: Model Evaluation

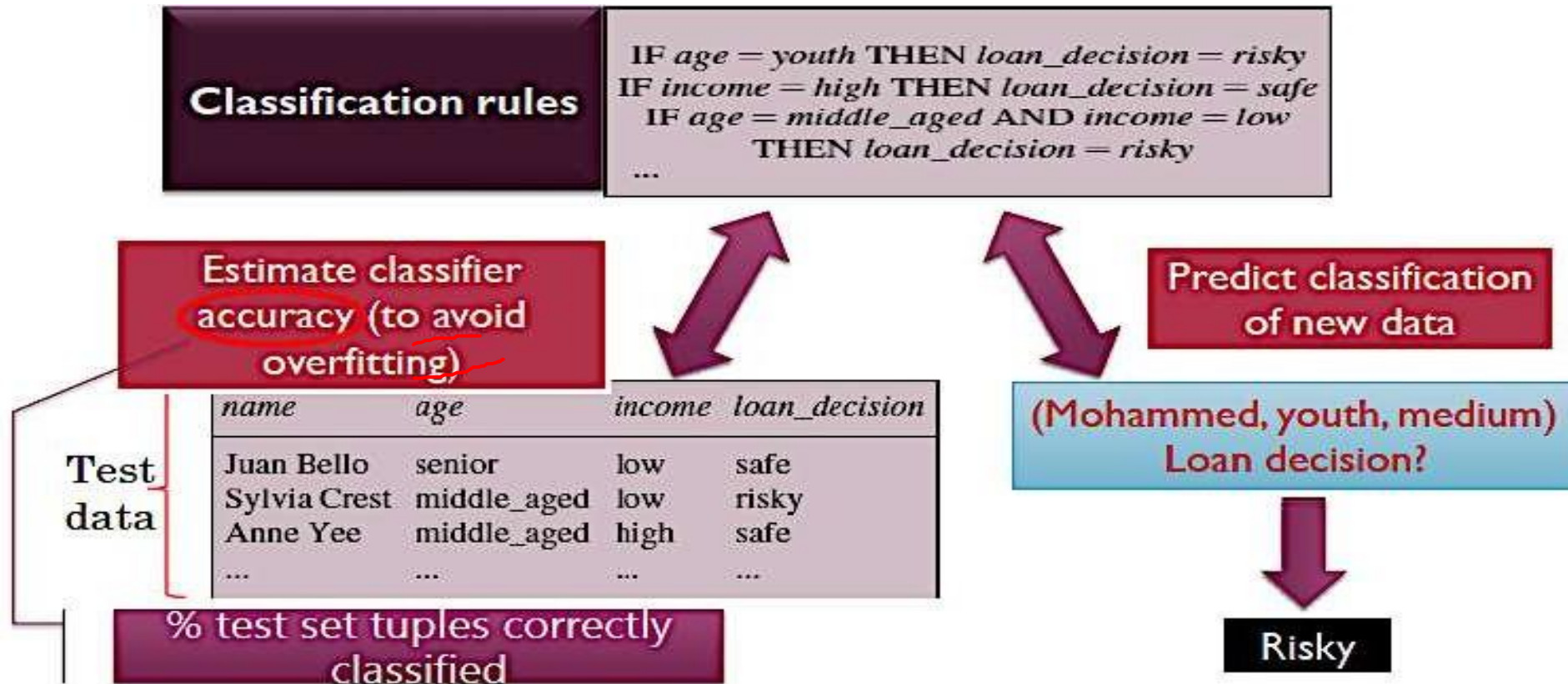
- Estimate the accuracy rate of the model based on a test set.

The known label of test sample is compared with the classified result from the model.

Accuracy rate is the percentage of test set samples that are correctly classified by the model.

Test set is independent of training set, otherwise over-fitting will occur.

- If the accuracy is acceptable, use the model to classify data tuples whose class labels are not known.



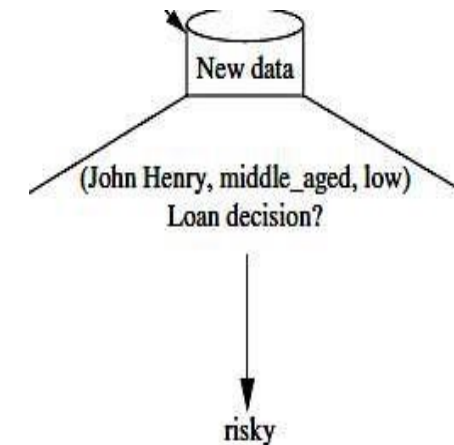
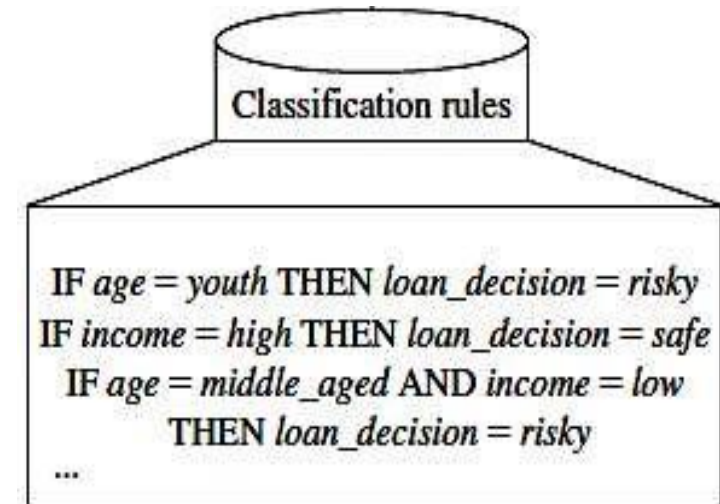
Step 3: Using the Model in Prediction

- The model is used to classify unseen objects (new data).

Give a class label to a new tuple.

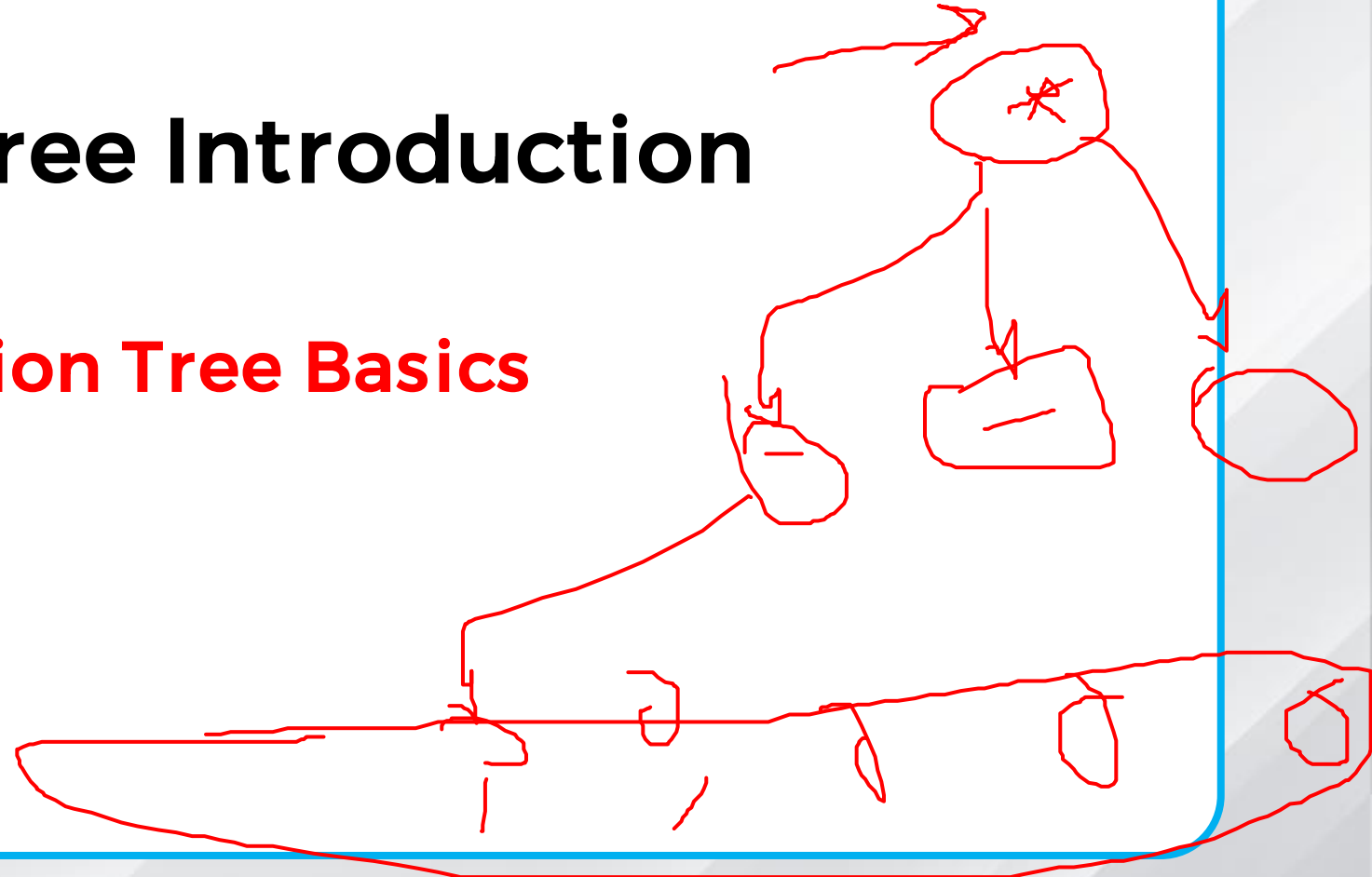
Predict the value of an actual attribute (class label).

Predict if the new data is (safe) or risky)?



Decision Tree Introduction

Decision Tree Basics

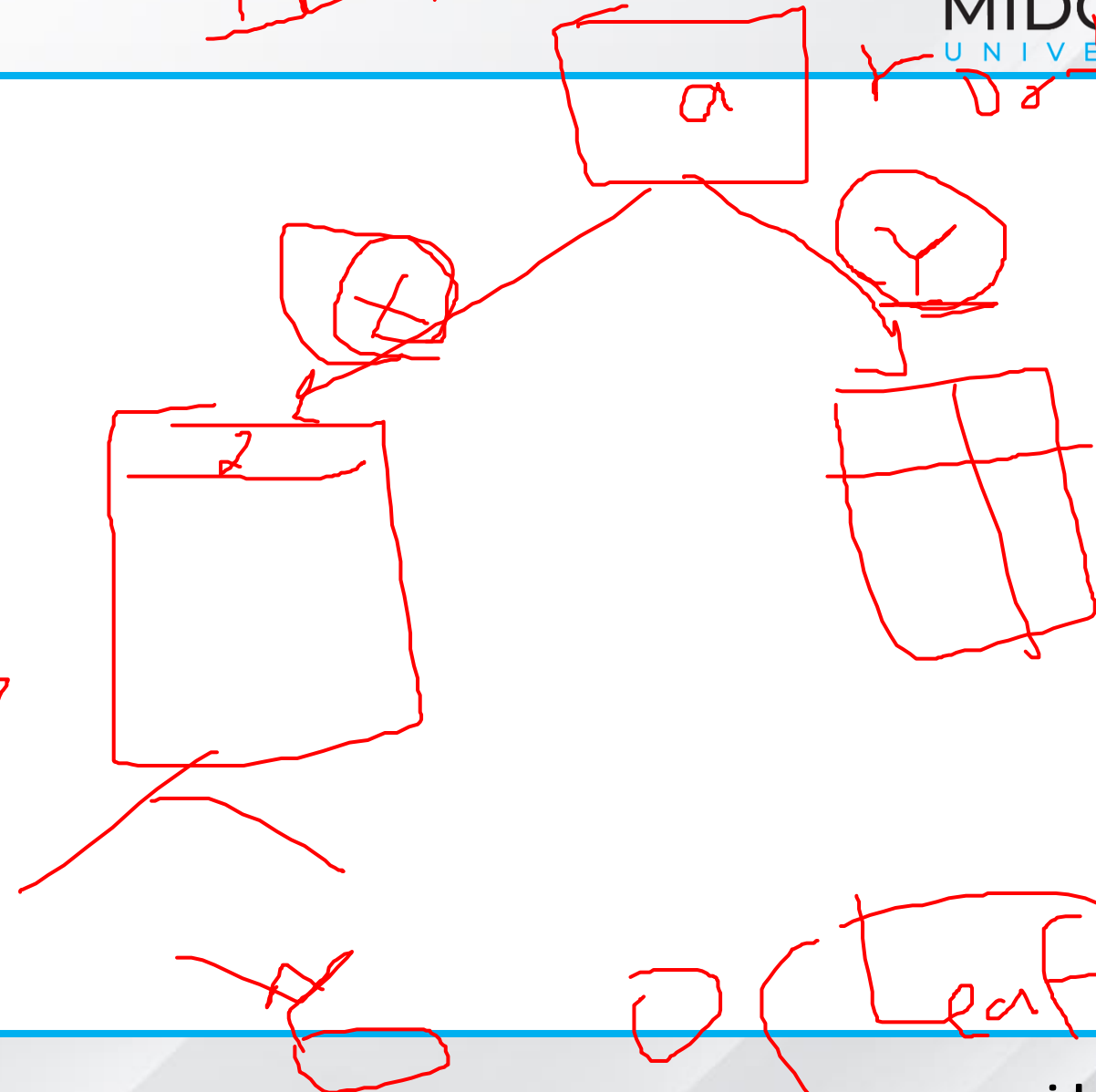


Decision Tree:

~~KDIN~~

- What is decision tree?
- Entropy.
- Information gain.

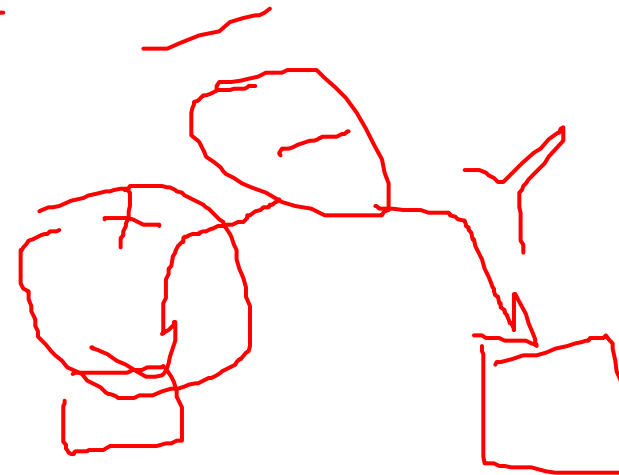
X	X	X	X	Z
X				M
Y				N



Decision Tree:

- What is decision tree?
 - A decision tree is a tree like predictive model used for classification and regression.
 - It is supervised learning method
- Learning algorithms used:
 - ID3 Algorithm: used for classification problem.

Machine Learning Decision Tree



$E(L)$
 ~~$IF(F_1)$~~
 $\checkmark (F_2) 0$
 $\checkmark (F_3) 0$

- Examples on Decision Tree
- Problems Definition
- ID3 algorithm

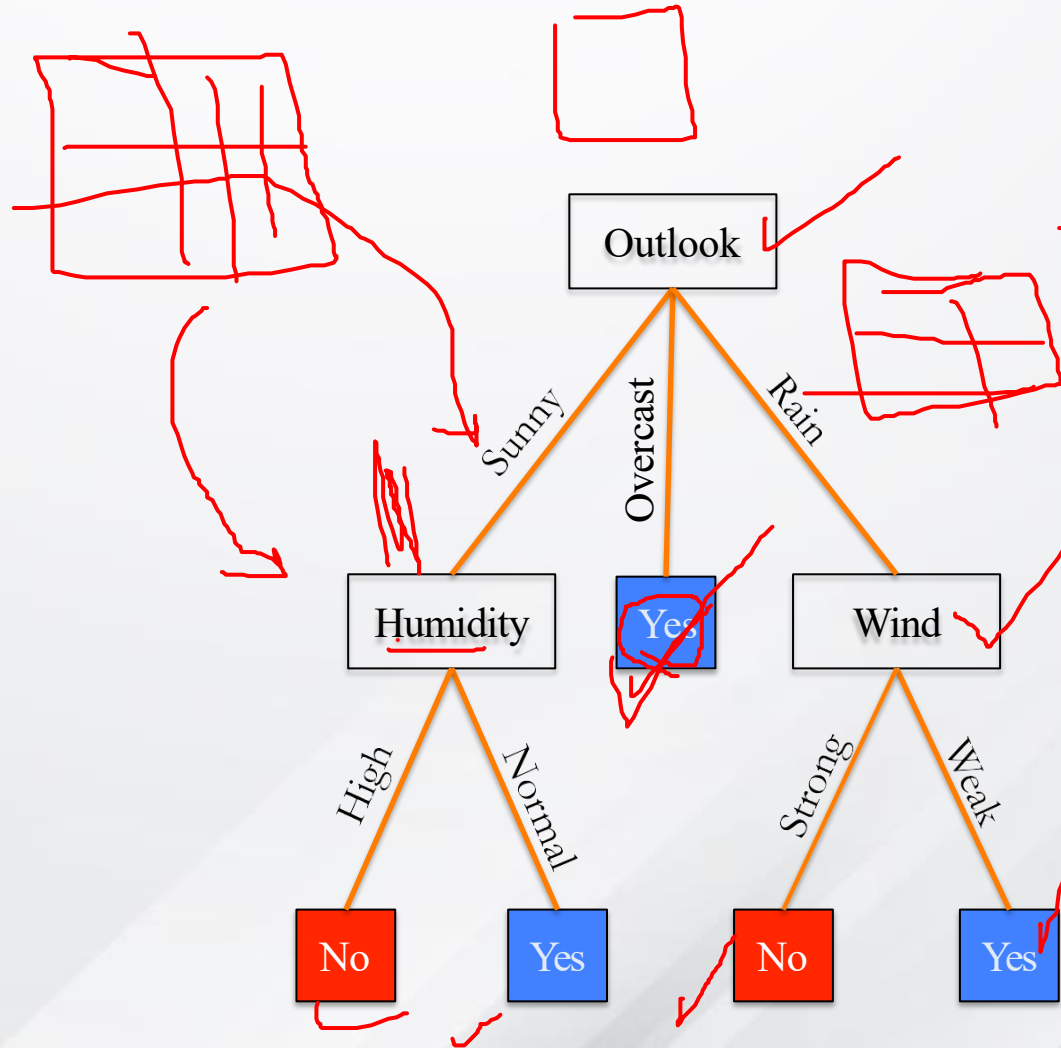
Will Nadal play Tennis?



Rafael Nadal

Day	<u>Outlook</u>	<u>Temp</u>	<u>Humidity</u>	<u>Wind</u>	Play Tennis
1 ✓	<u>Sunny</u>	<u>Hot</u>	High	<u>Weak</u> ✓	No
2	Sunny	Hot	High	Strong	No
3 ✓	<u>Overcast</u>	<u>Hot</u>	<u>High</u>	<u>Weak</u>	Yes
4	Rain	<u>Mild</u>	High	Weak	Yes
5 ✓	<u>Rain</u>	<u>Cool</u>	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Weak	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Strong	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	<u>Rain</u>	<u>Mild</u>	<u>High</u>	Strong	No

Will Nadal play Tennis?

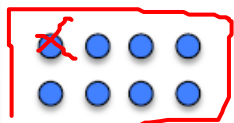


Day	Outlook	Temp	Humidity	Wind	Play Tennis
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Weak	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Strong	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

$$P(R) = \frac{3}{10}$$

$$P(0) = \frac{1}{5}$$

$$P(A) =$$



$$-\left[\left(\frac{8}{8}\right) \log_2 \left(\frac{8}{8}\right) + \left(\frac{0}{8}\right) \log_2 \left(\frac{0}{8}\right)\right] = 0$$

$$E(0) =$$

$$-\sum P_i \log_2 P_i$$

$$+ P_{-1} \log_2 P_{-1}$$

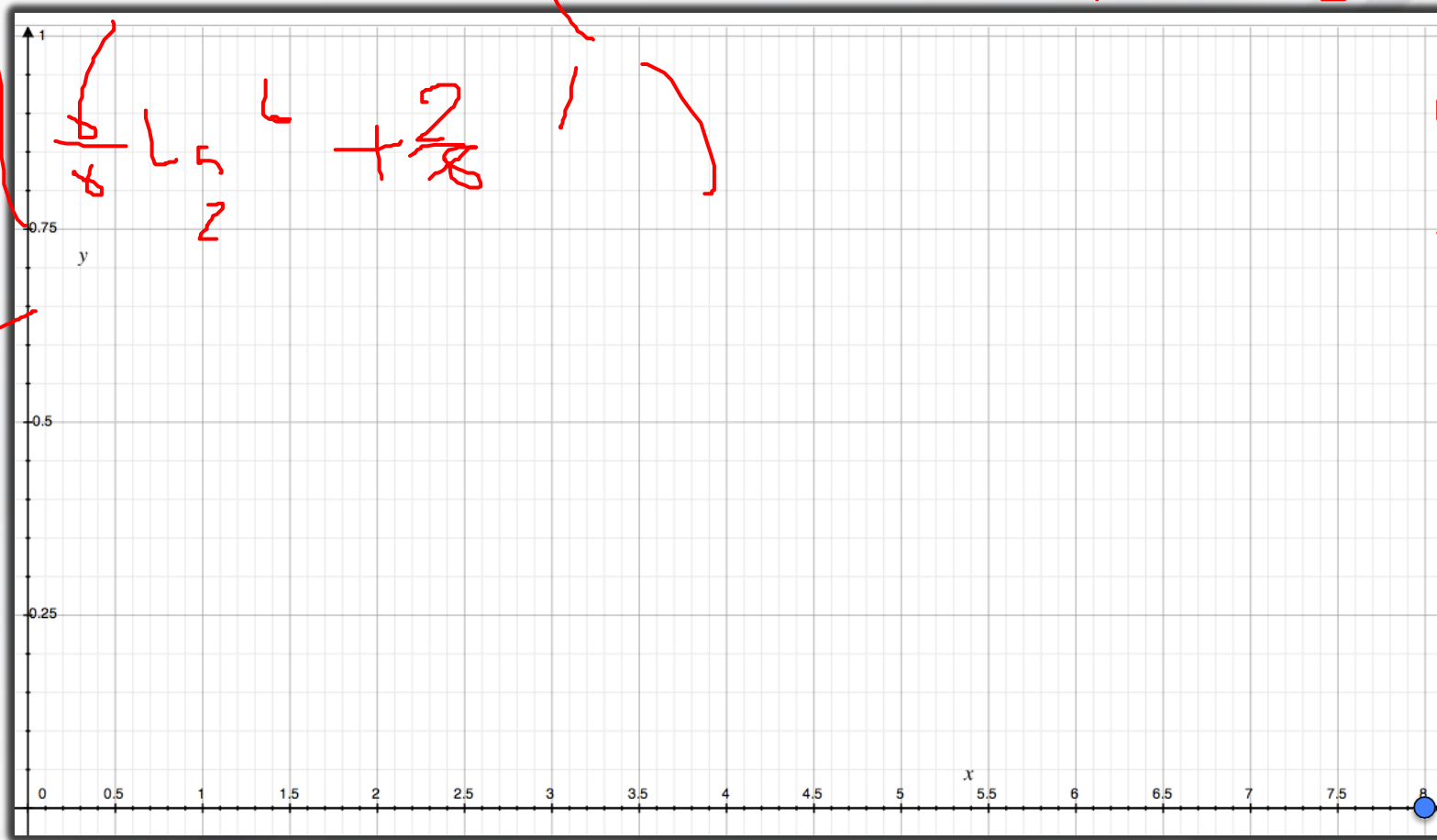
$$\frac{(6, 2)}{2}$$

$$\frac{6}{8} \log_2 \frac{6}{8} + \frac{2}{8} \log_2 \frac{2}{8}$$

$$P(1) = \frac{3}{10}$$

$$\frac{1}{6}$$

$$E(8, 6)$$



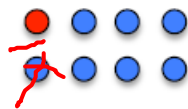
$$\{1, 2, 3\}$$

$$1, 5, 10$$

$$2, 3, 4, 5, 6, 7, 8, 9, 10$$

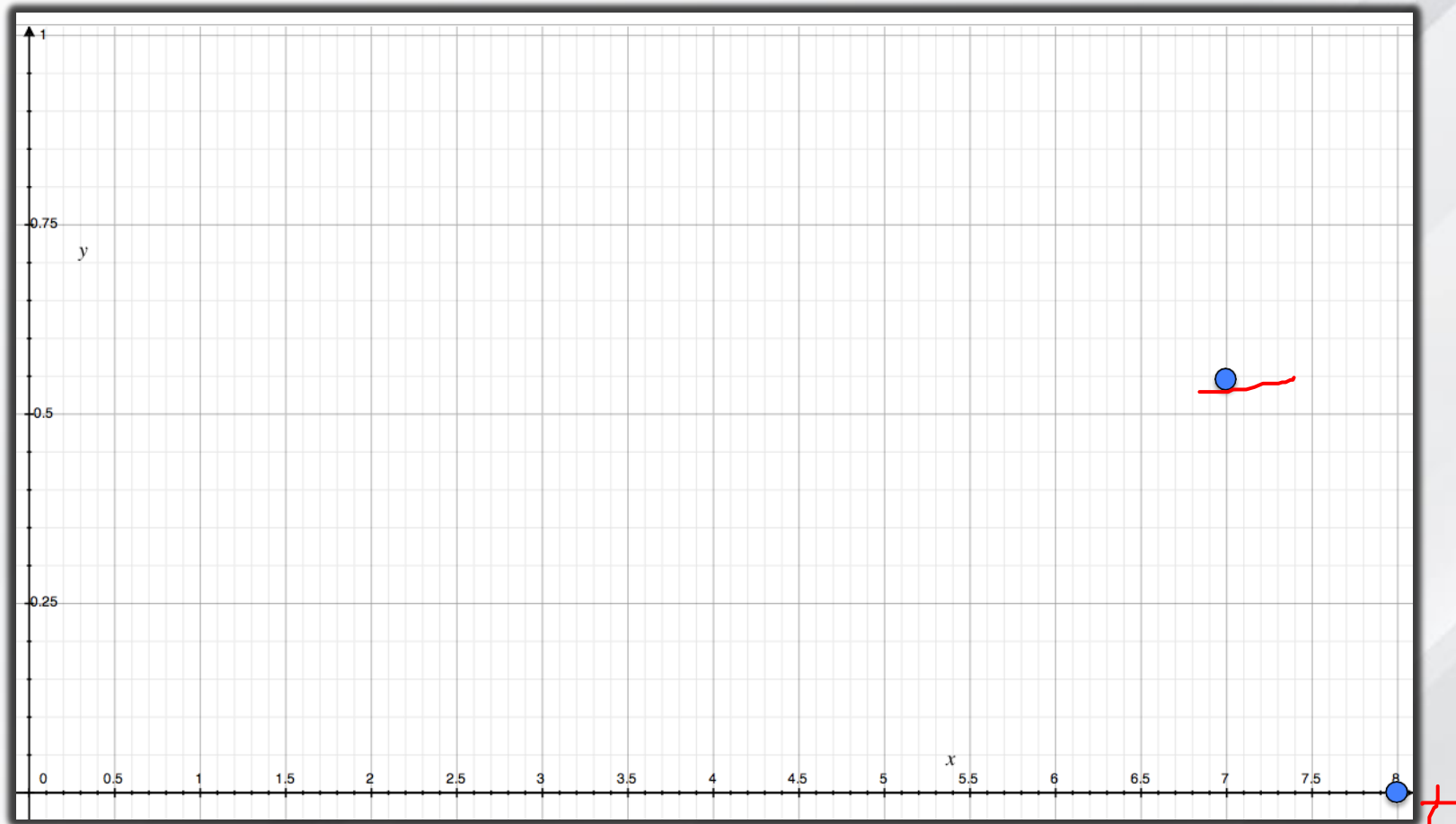
$$P(1) = \frac{1}{10}$$

$[7, 1]$



$$-\left(\frac{7}{8}\right) \log_2 \left(\frac{7}{8}\right) - \left(\frac{1}{8}\right) \log_2 \left(\frac{1}{8}\right) = \underline{0.54}$$

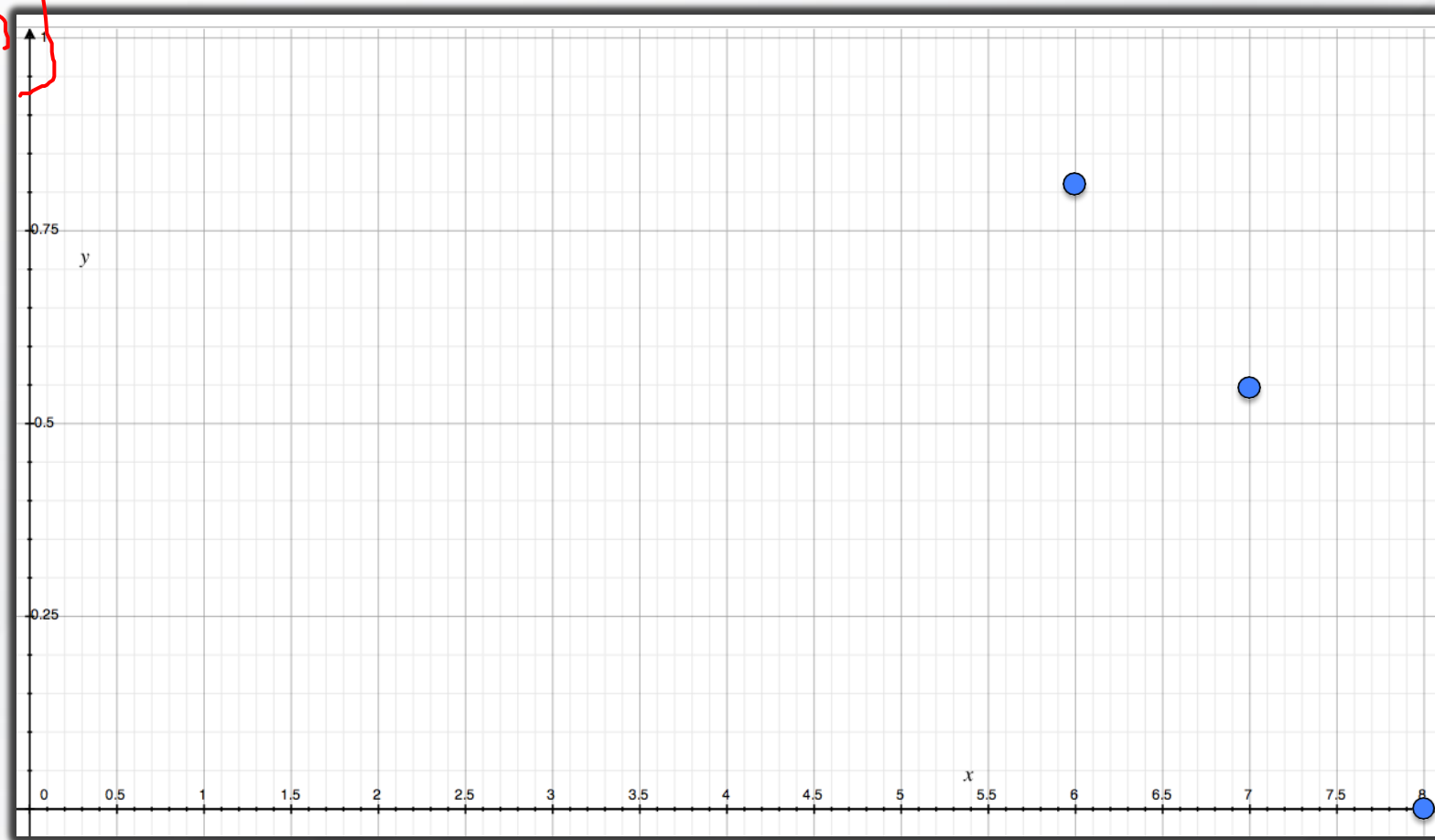
$[-\sum p_i \log_2 p_i]$





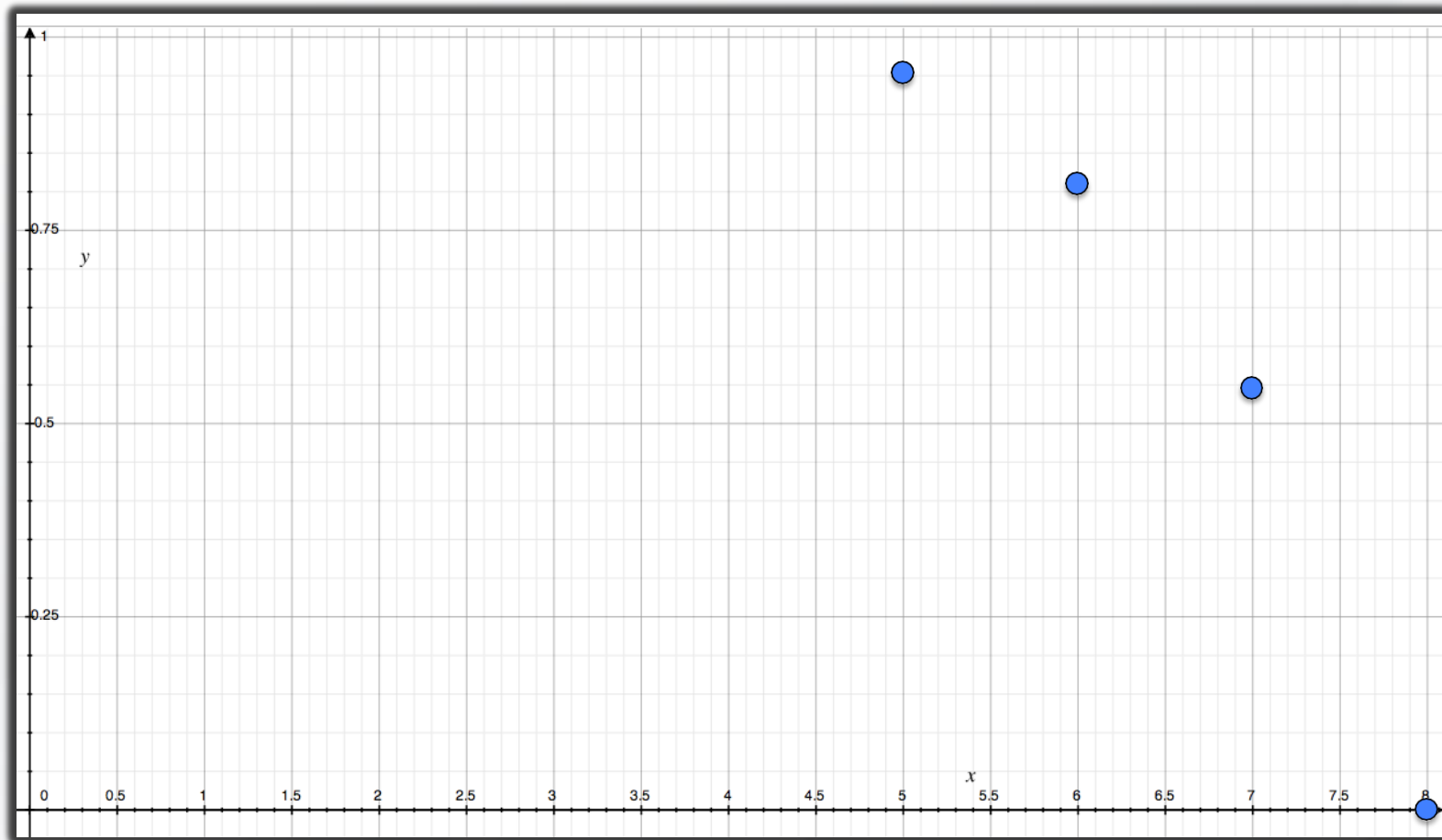
$$-\left(\frac{6}{8}\right) \log_2 \left(\frac{6}{8}\right) - \left(\frac{2}{8}\right) \log_2 \left(\frac{2}{8}\right) = \underline{0.81}$$

EL 5.3



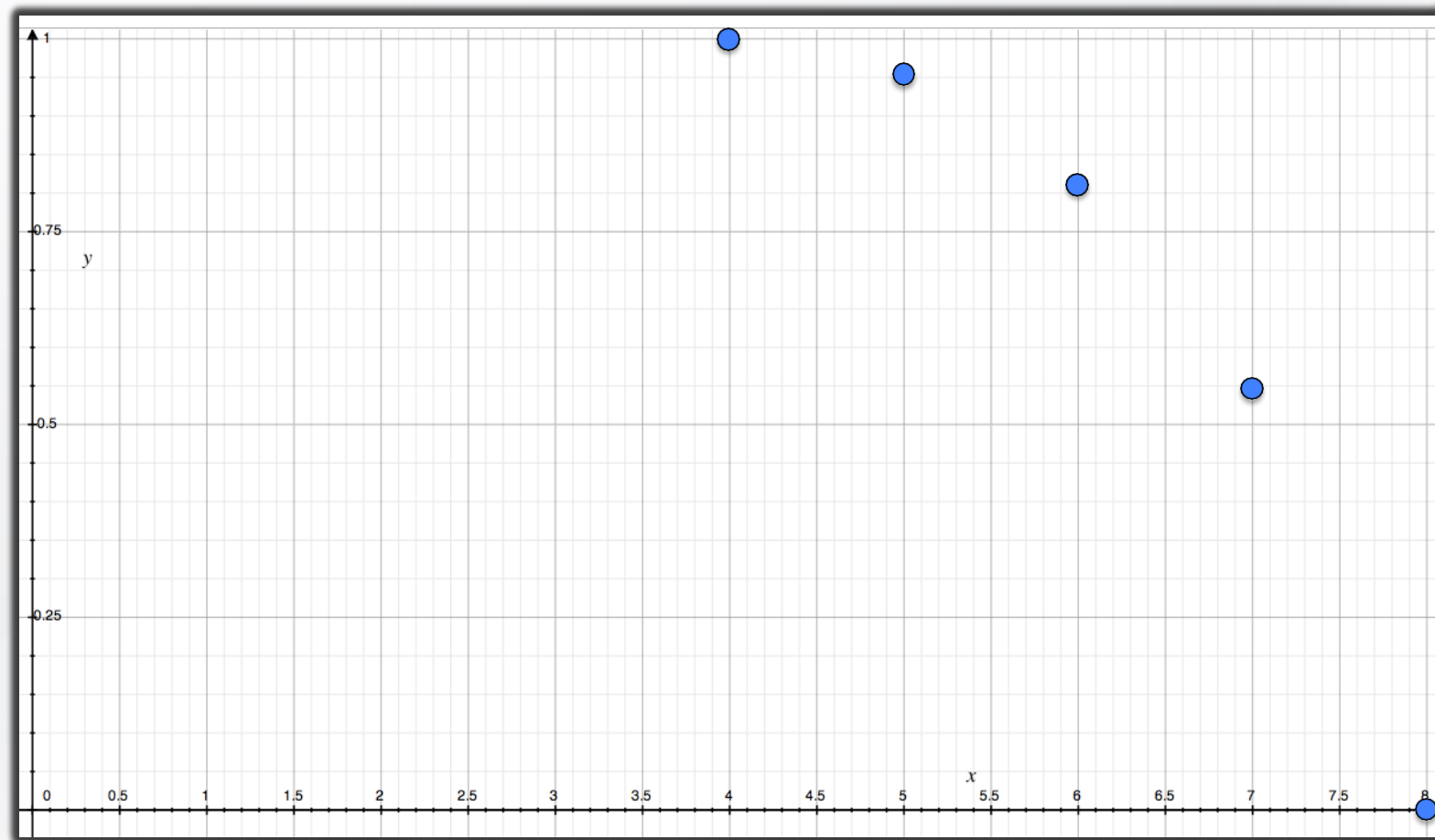


$$- \left(\frac{5}{8} \right) \log_2 \left(\frac{5}{8} \right) - \left(\frac{3}{8} \right) \log_2 \left(\frac{3}{8} \right) = \underline{0.95}$$



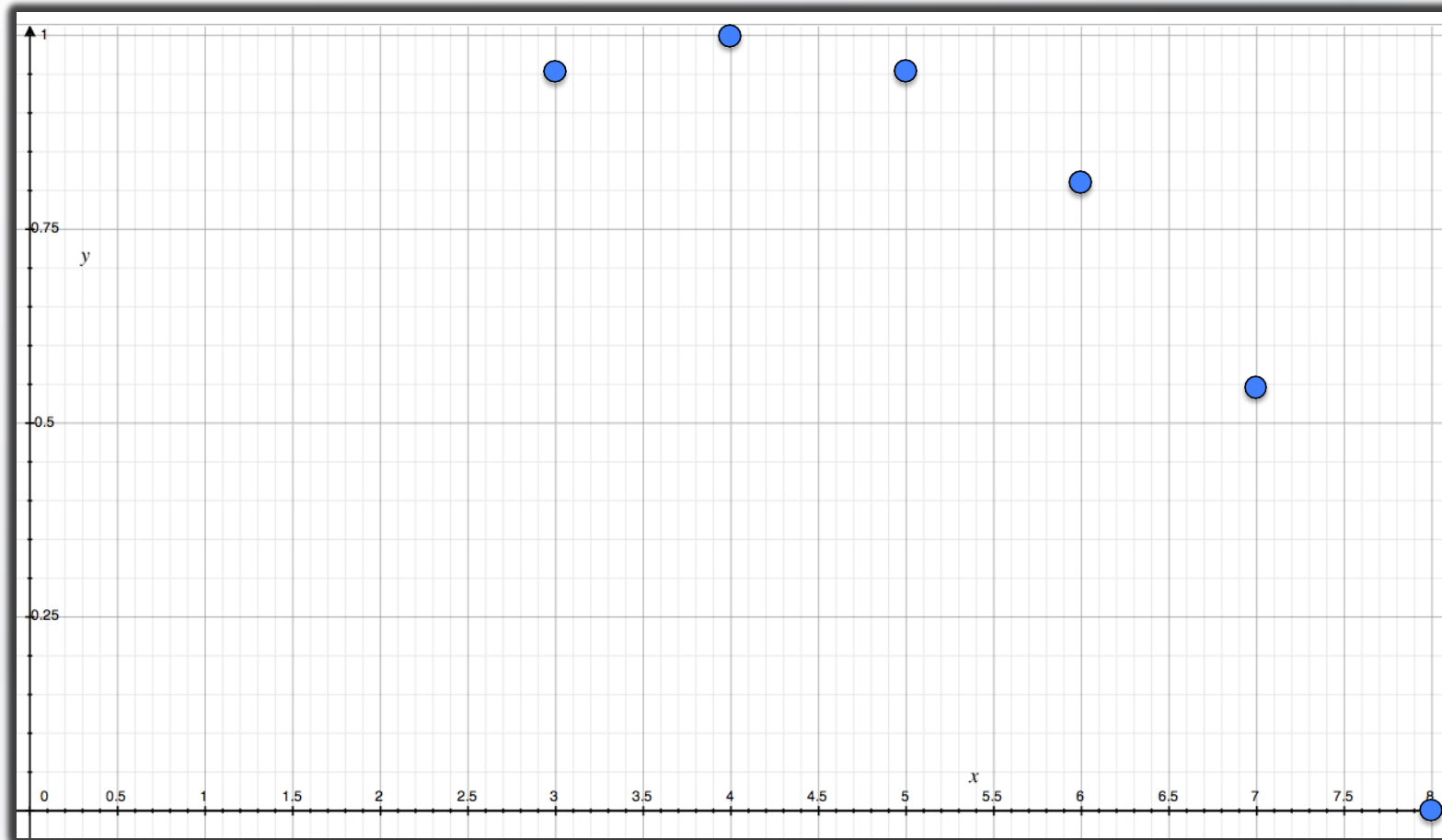


$$-\left(\frac{4}{8}\right) \log_2 \left(\frac{4}{8}\right) - \left(\frac{4}{8}\right) \log_2 \left(\frac{4}{8}\right) = 1$$



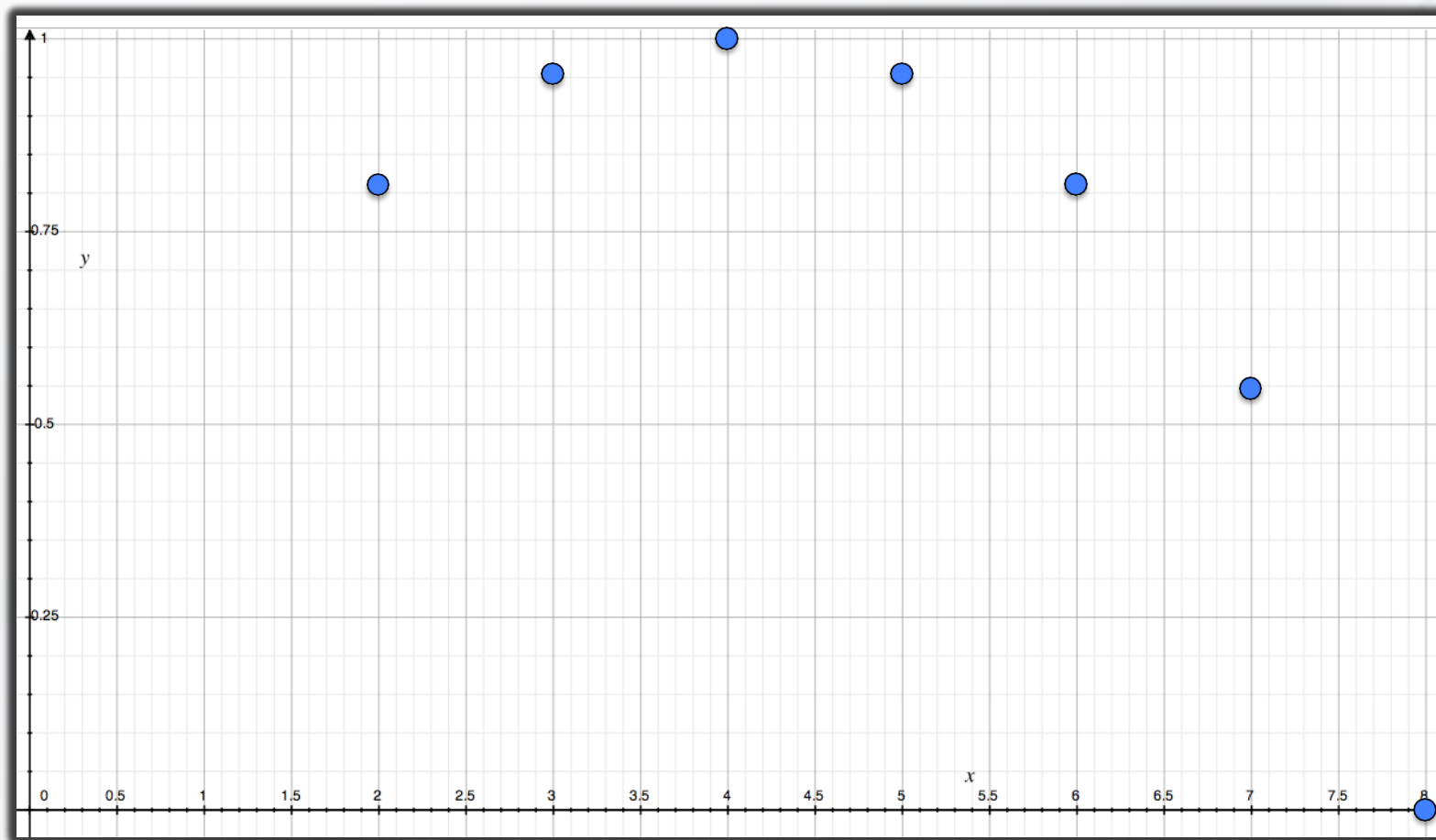


$$-\left(\frac{3}{8}\right) \log_2 \left(\frac{3}{8}\right) - \left(\frac{5}{8}\right) \log_2 \left(\frac{5}{8}\right) = 0.95$$



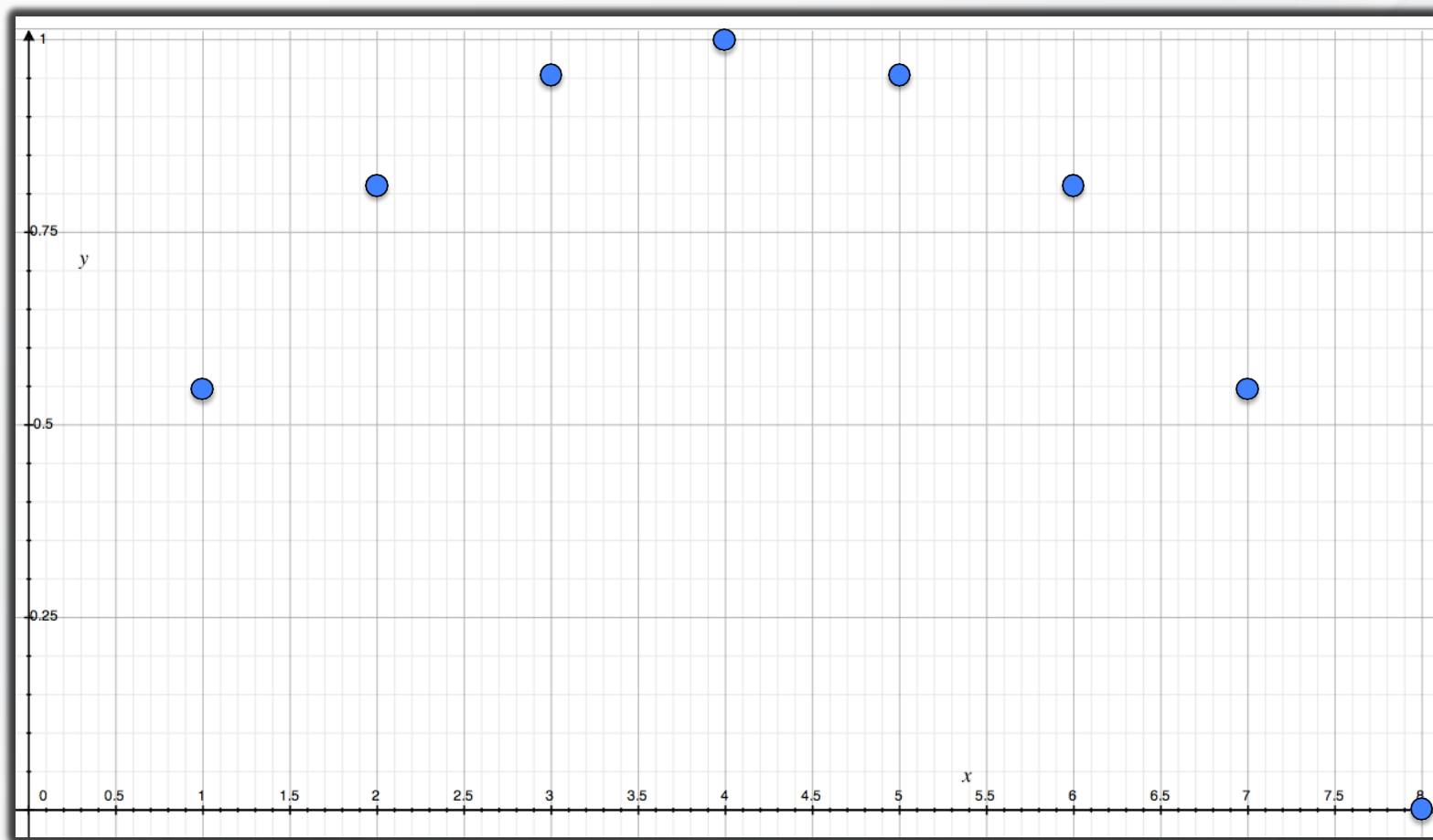


$$-\left(\frac{2}{8}\right) \log_2 \left(\frac{2}{8}\right) - \left(\frac{6}{8}\right) \log_2 \left(\frac{6}{8}\right) = 0.81$$



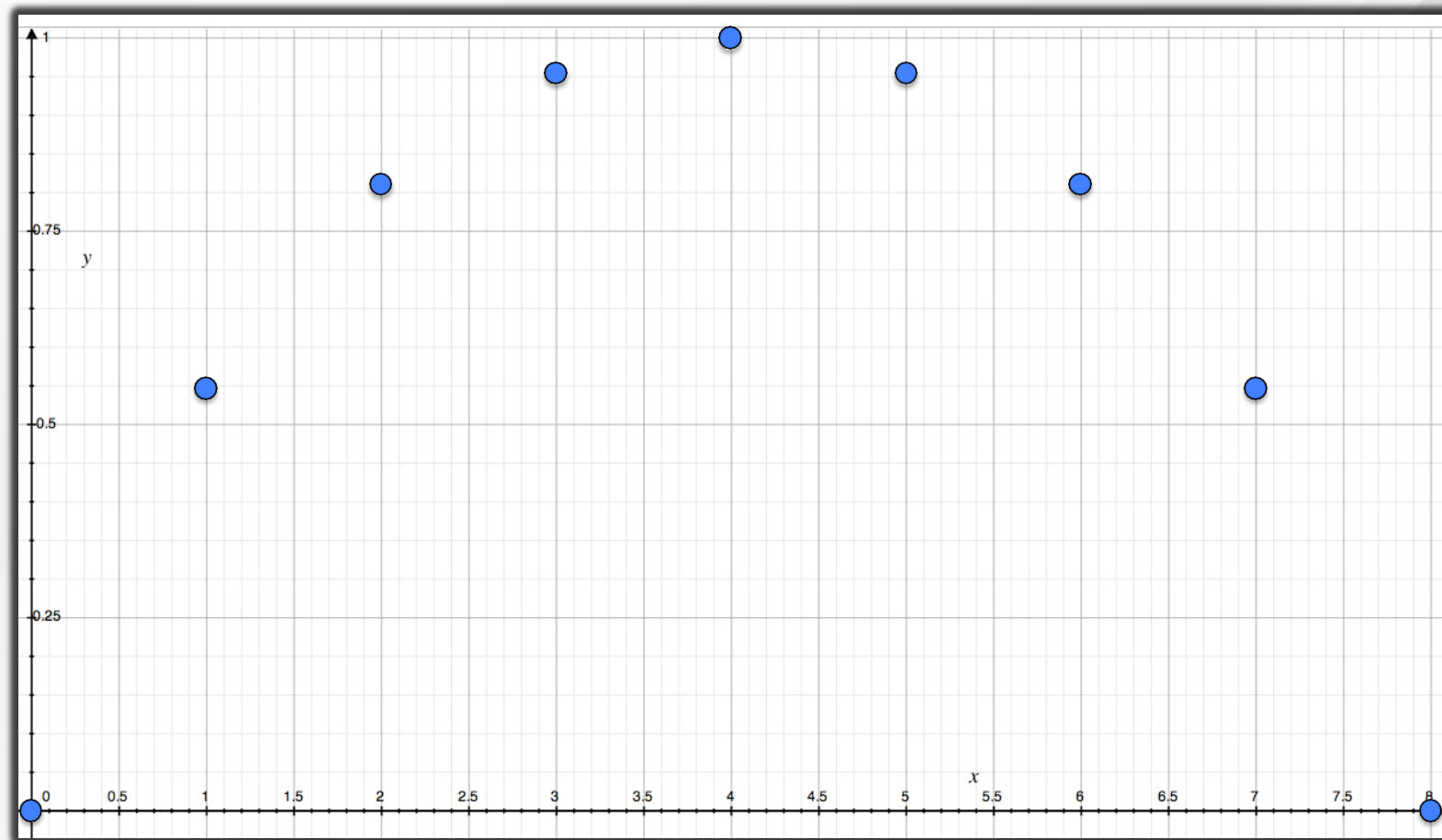


$$-\left(\frac{1}{8}\right) \log_2 \left(\frac{1}{8}\right) - \left(\frac{7}{8}\right) \log_2 \left(\frac{7}{8}\right) = 0.54$$





$$-\left(\frac{0}{8}\right) \log_2 \left(\frac{0}{8}\right) - \left(\frac{8}{8}\right) \log_2 \left(\frac{8}{8}\right) = 0$$





$$y = - \sum_{i=1}^k p_i \log_k(p_i)$$

$$y = - \underbrace{\left[\left(\frac{1}{10} \right) \log_4 \left(\frac{1}{10} \right) \right]}_{\text{Red}} - \underbrace{\left[\left(\frac{3}{10} \right) \log_4 \left(\frac{3}{10} \right) \right]}_{\text{Green}} - \underbrace{\left[\left(\frac{2}{10} \right) \log_4 \left(\frac{2}{10} \right) \right]}_{\text{Blue}} - \underbrace{\left[\left(\frac{4}{10} \right) \log_4 \left(\frac{4}{10} \right) \right]}_{\text{Yellow}}$$

$P(E|Sun) \checkmark$
(2,3)

$P(4,30) +$

$P(3,2)$

No
No

$E(Plan)$

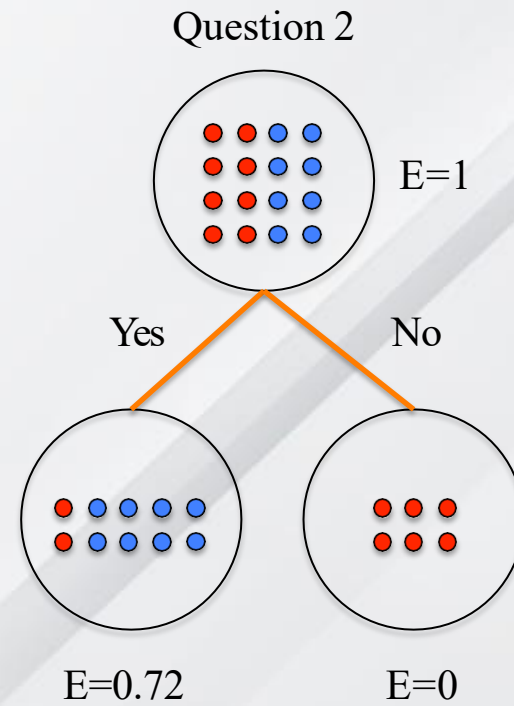
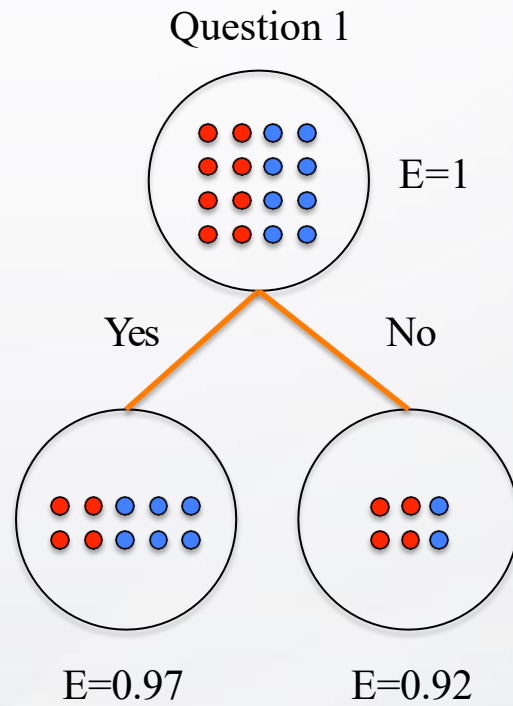
$E(9,5)$

$-\left[\frac{9}{14} + \frac{5}{14}\right]$

② $E(5)$

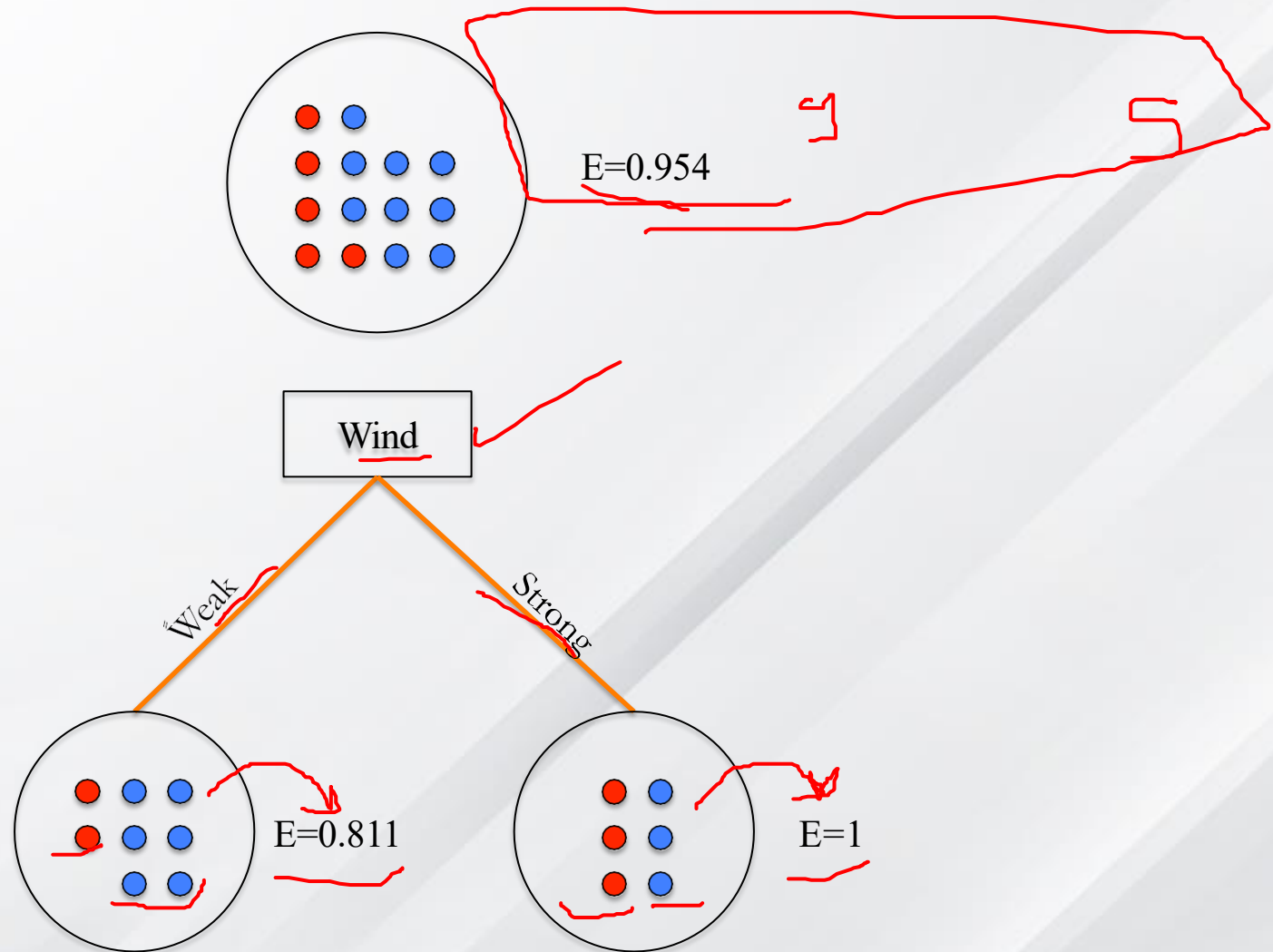
IL

Day	Outlook	Temp	Humidity	Wind	Play Tennis
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Weak	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Strong	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No



Information Gain

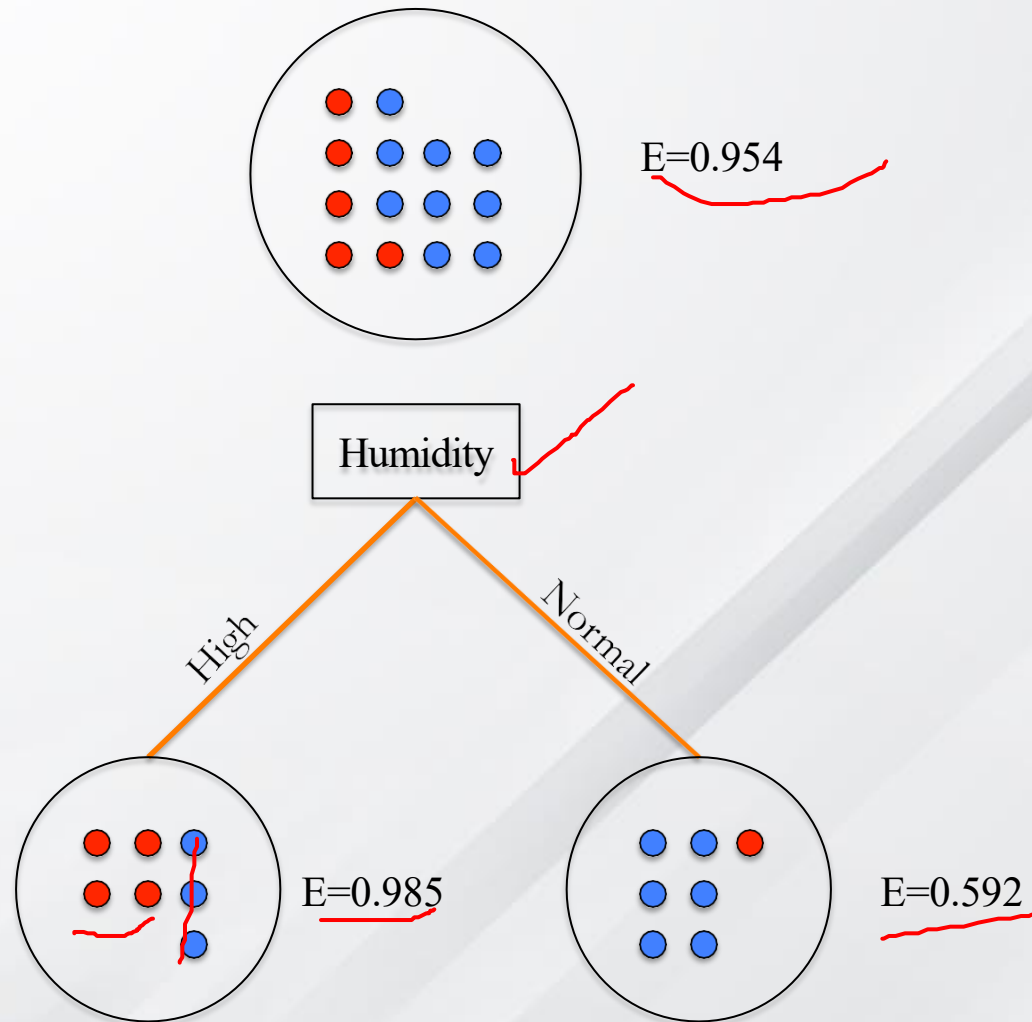
$$\underline{G(S, Q)} = \underline{E(S)} - \sum_{i=1}^k \underline{p_i E(S, Q_i)}$$



$$G(S, \text{Wind}) = 0.954 - \frac{8}{14} \cdot 0.811 - \frac{6}{14} \cdot 1$$

$$G(S, \text{Wind}) = 0.048$$

$$G(S, Wind) = 0.048$$

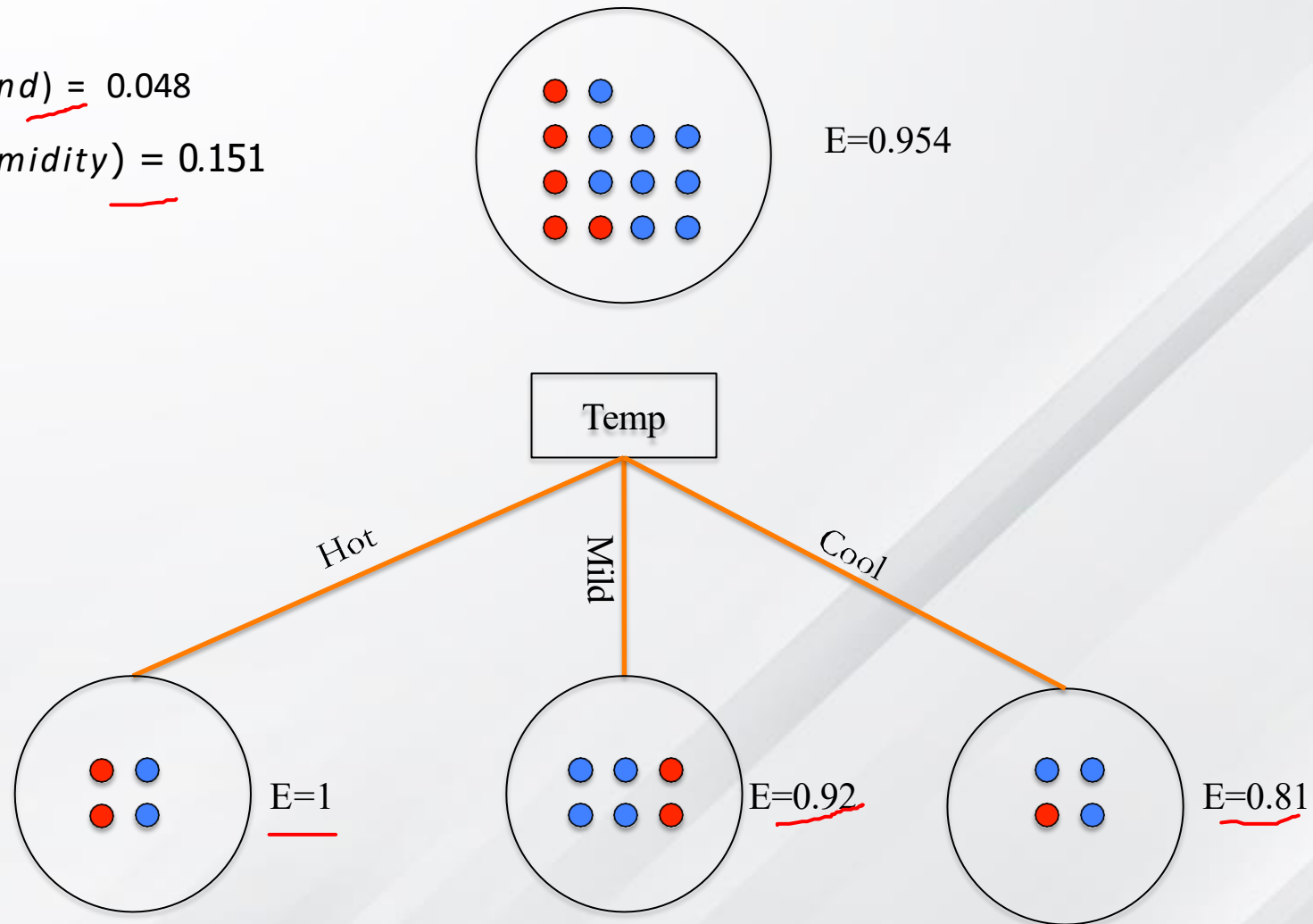


$$G(S, Humidity) = 0.954 - \frac{7}{14}0.985 - \frac{7}{14}0.592$$

$$G(S, Humidity) = 0.151$$

$$G(S, \text{Wind}) = 0.048$$

$$G(S, \text{Humidity}) = 0.151$$



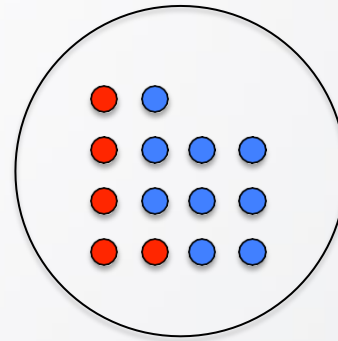
$$G(S, \text{Temp}) = 0.954 - \frac{4}{14} 1 - \frac{6}{14} 0.92 - \frac{4}{14} 0.81$$

$$G(S, \text{Temp}) = 0.042$$

$$G(S, Wind) = 0.048$$

$$G(S, Humidity) = 0.151$$

$$G(S, Temp) = 0.042$$



$$E=0.954$$

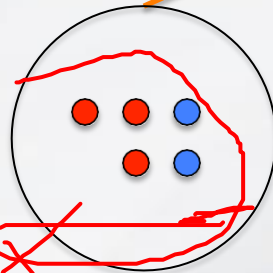
-0.247

Outlook

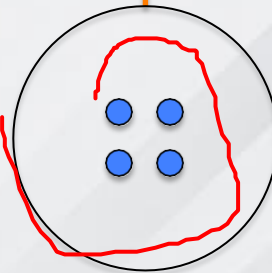
~~Sunny~~

Overcast

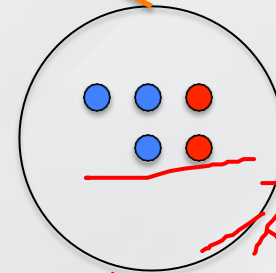
Rain



$$E=0.971$$



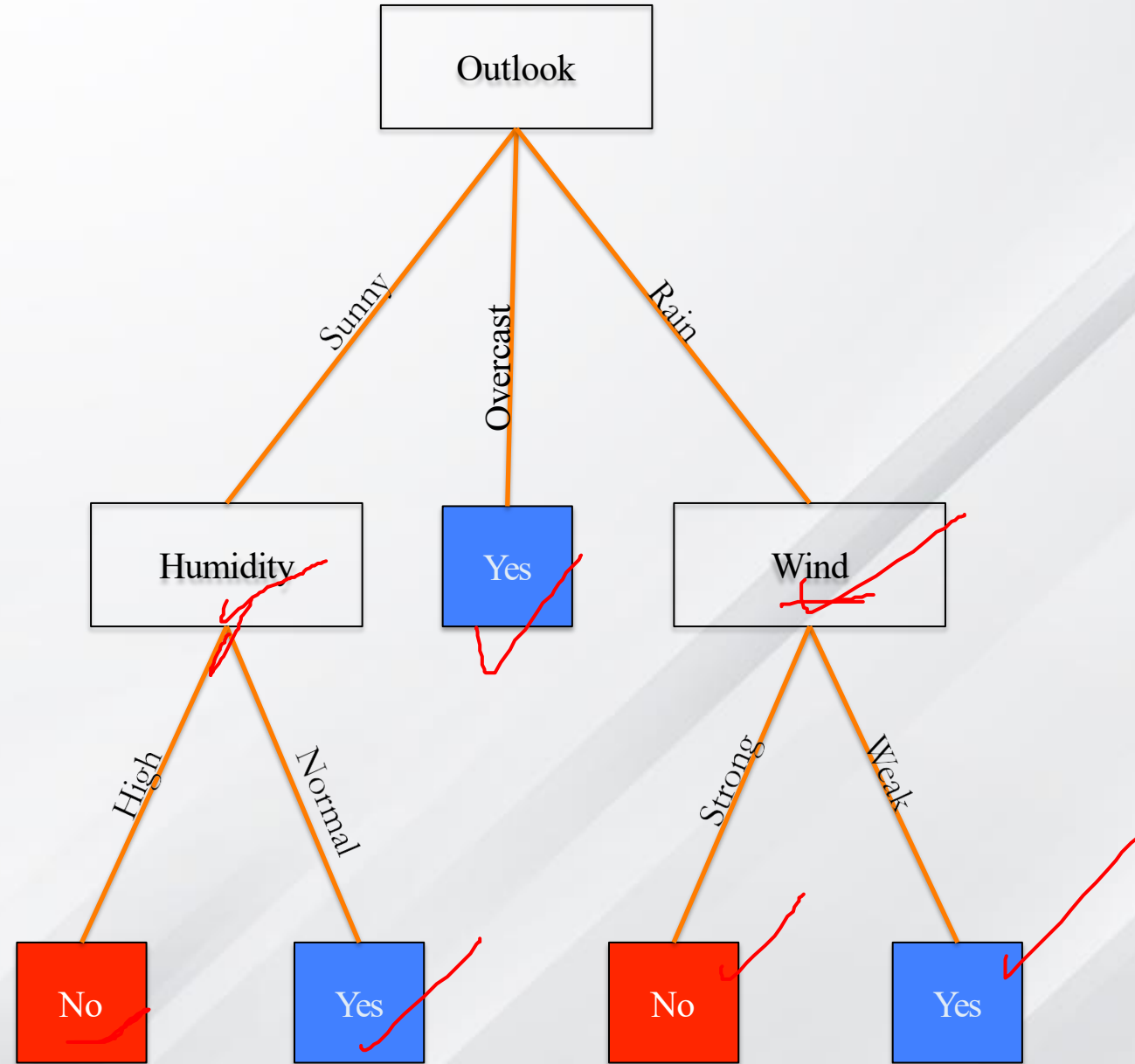
$$E=0$$



$$E=0.971$$

$$G(S, Outlook) = 0.954 - \left(\frac{5}{14} 0.971 \right) - \left(\frac{4}{14} 0 \right) - \left(\frac{5}{14} 0.971 \right)$$

$$G(S, Outlook) = 0.247$$



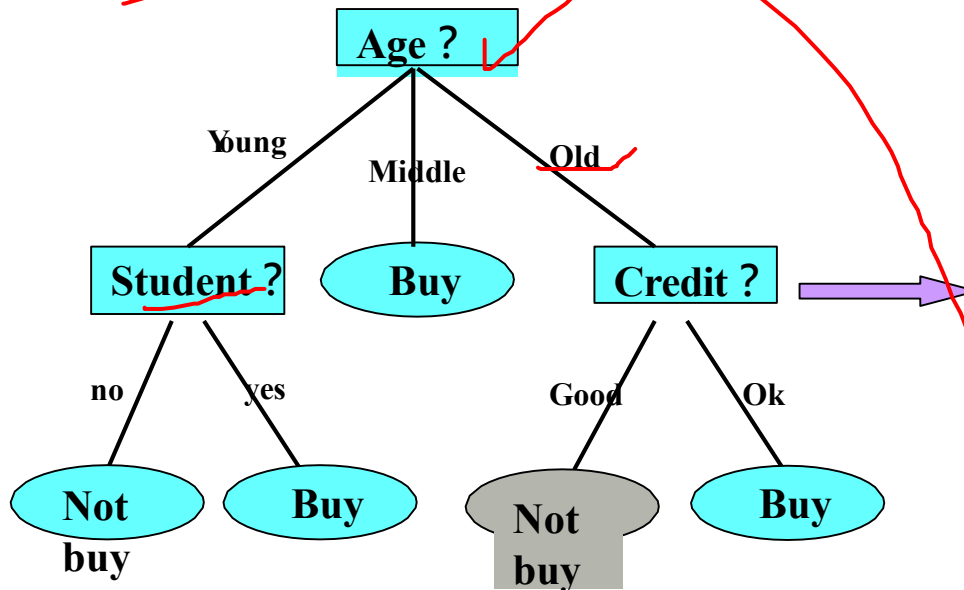
Example on Decision Tree

- If you have such a statistical table on the left, given a new customer with some information can you predict whether he or she would like to buy a computer or not.
- Decision tree methods can help you to solve this problem

Num		Salary	Student	Credit	Buy Computer or not
64		High	No	Ok	No
64		High	No	Good	No
128		High	No	Ok	Yes
60		Middle	No	Ok	Yes
64		Low	Yes	Ok	Yes
64		Low	Yes	Good	No
64		Low	Yes	Good	Yes
128		Middle	No	Ok	No
64		Low	Yes	Ok	Yes
132		Middle	Yes	Ok	Yes
64		Middle	Yes	Good	Yes
32		Middle	No	Good	Yes
32		High	Yes	Ok	Yes
63		Middle	No	Good	No
1		Middle	Yes	Good	Yes

Example on Decision Tree

- Use the decision tree below:
- Each node is an attribute, can classify the data based on the value of the attribute.
- Each node's branch corresponding to a value of this attribute.
- Each leaf node is the final decision.

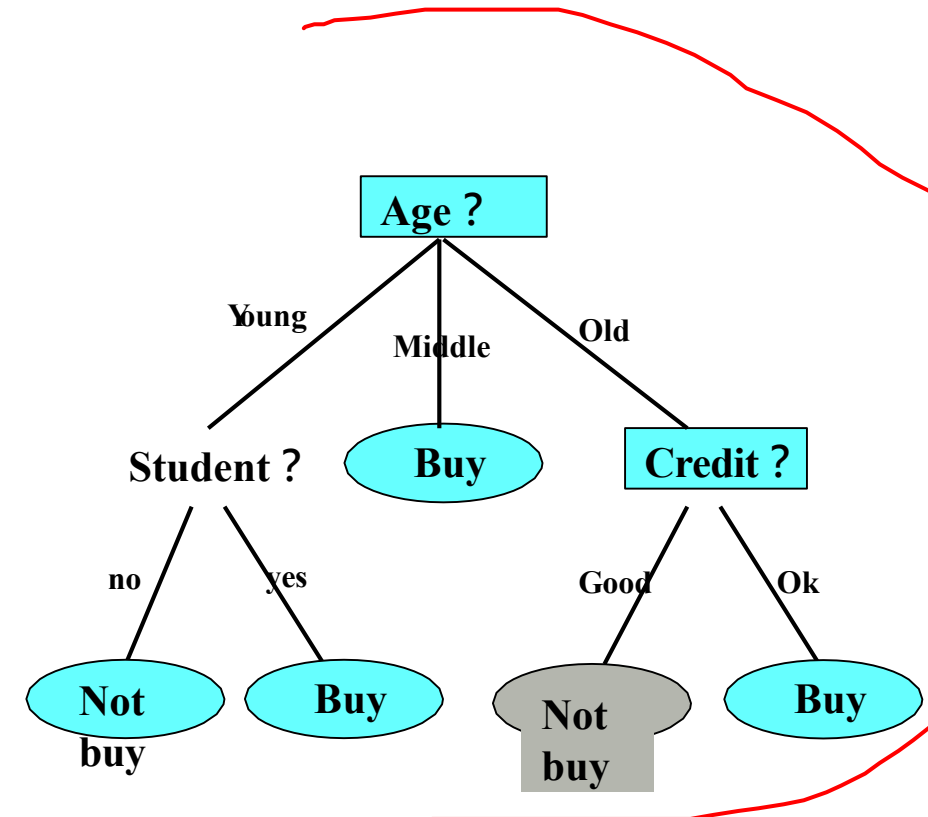


If (Age = M) or (Age = Old and Credit = Ok) or (Age = Y and Student = Y) then Buy Computer

If (Age = O and Credit = G) or (Age = Y and Student = No) then Not buy Computer

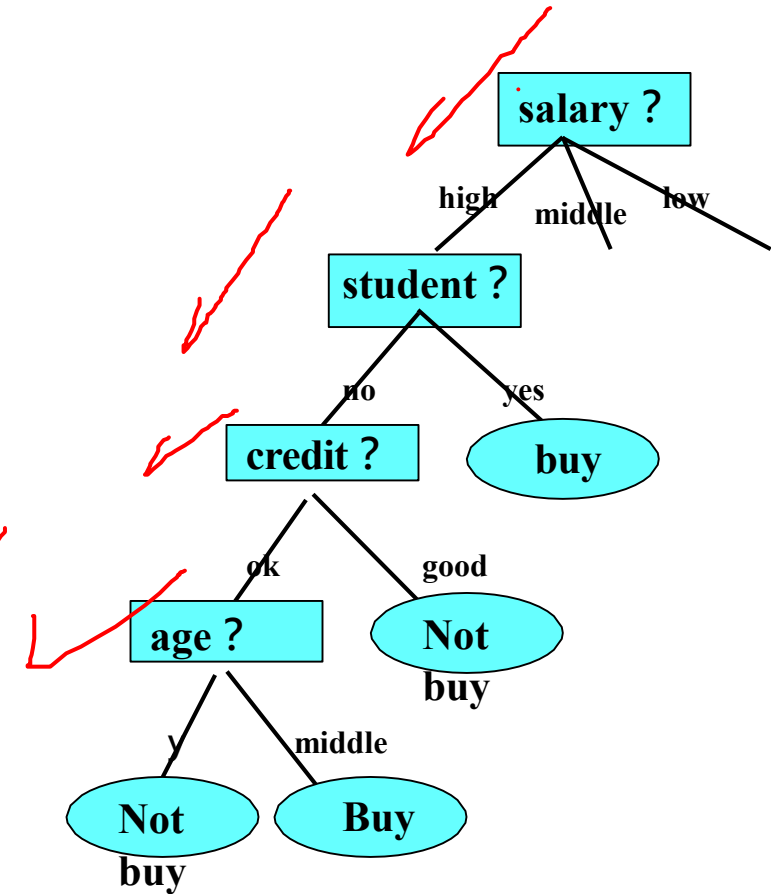
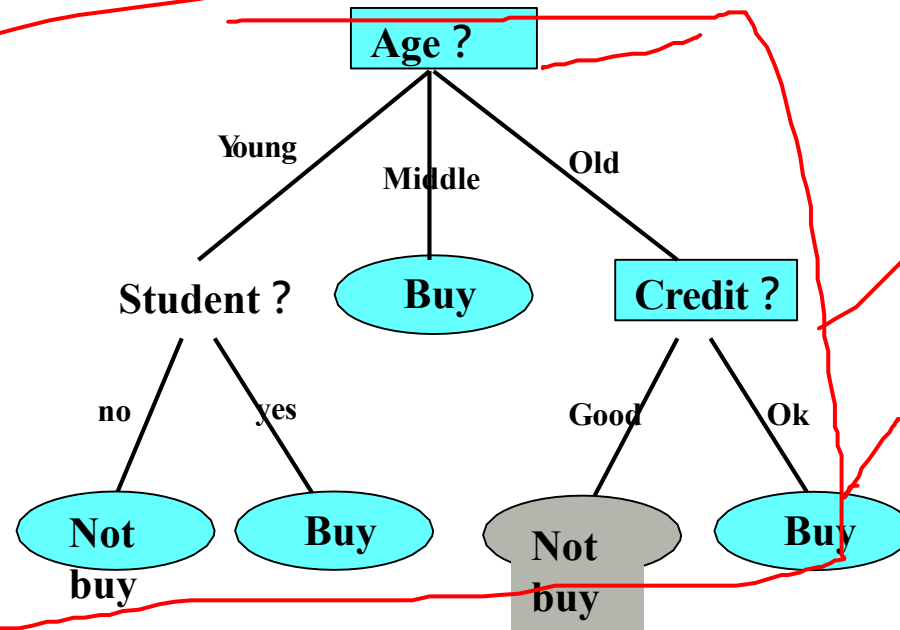
Definition on Decision Tree

- Tree.
- Decision Tree is a decision support tool that uses a tree-like graph or model of decision and their possible consequences, including chance event outcome, resource costs, and utility.
- Decision tree are commonly used in decision analysis, to help identify a strategy most likely to reach a goal.



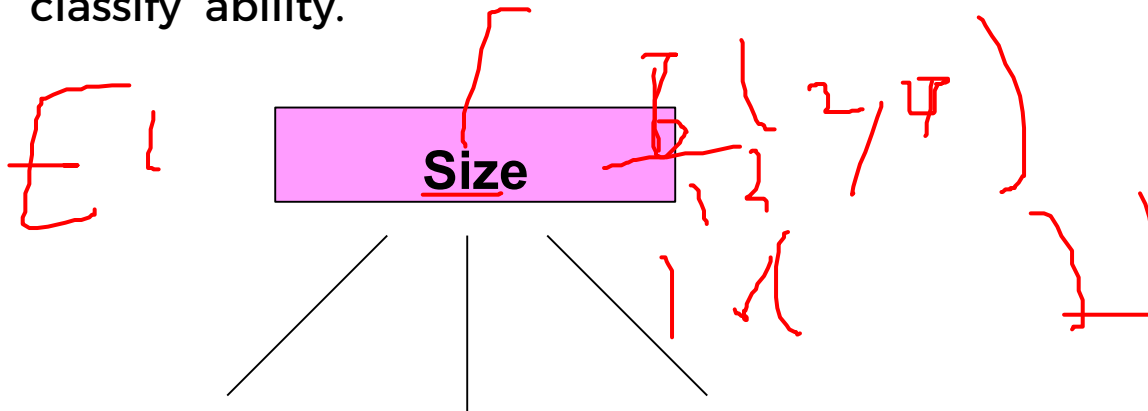
Decision Tree

- Node: attribute.
- Branch: value of the attribute.
- Leaf: final decision.
- Good or Bad: select different attributes for
- different nodes



Construct the Decision Tree

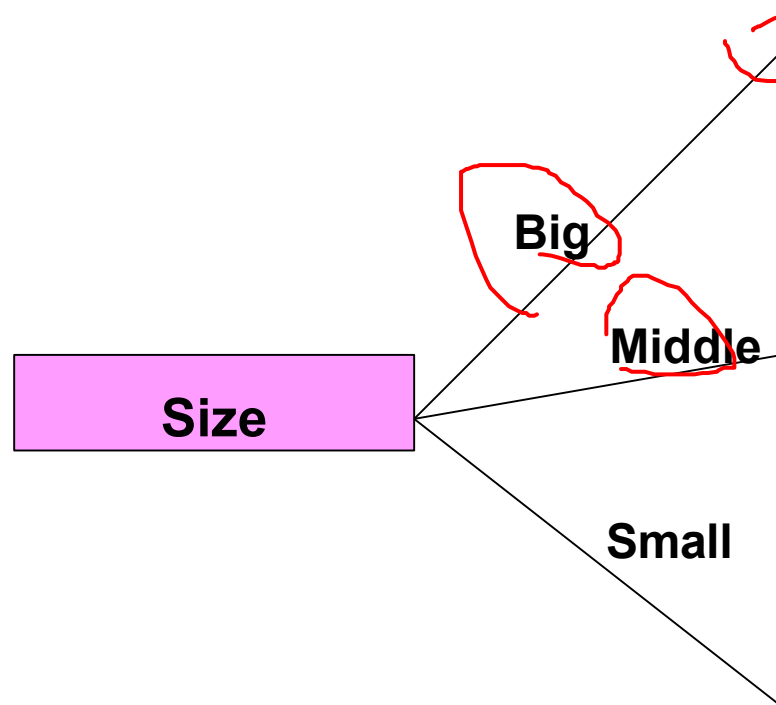
- To classify the bear dangerous or not dangerous.
- Select the attribute:
- first select the attribute which has the classify ability.



Num	Color	Size	Hair	Class
1	Black	Big	Curls	Dangerous
2	Brown	Big	Straight	Dangerous
3	Brown	Middle	Curls	No
4	Black	Small	Curls	No
5	Brown	Middle	Straight	Dangerous
6	Black	Big	Straights	Dangerous
7	Brown	Small	Curls	Dangerous
8	Brown	Small	Straights	No
9	Brown	Big	Curls	Dangerous
10	Black	Middle	Curls	No
11	Black	Middle	Straights	No
12	Black	Small	Straights	No

Construct the Decision Tree

- Divide the training sets based on the attribute value.



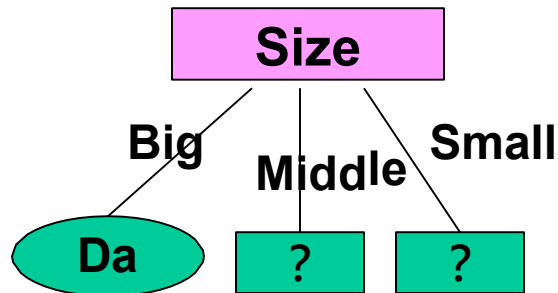
Num	Color	Size	Hair	Class
1	Black	Big	Curls	D
2	Brown	Big	Straight	D
6	Black	Big	Straight	D
9	Brown	Big	Curl	D

Num	Color	Size	Hair	Class
3	Brown	M	Curl	No
5	Brown	M	Straigh	D
10	Black	M	Curl	No
11	Black	M	Straith	No

Num	Color	Size	Hair	Class
4	Black	Small	Curl	No
7	Brown	Small	Curl	D
8	Brown	Small	Straight	No
12	Black	Small	Straight	No

Construct the Decision Tree

- continue to classify the sub-training sets until all the sub-sets belongs to the same class.



Num	Color	Size	Hair	Class
3	Br	M	C	Not
5	Br	M	S	Da
10	Bl	M	C	Not
11	Bl	M	S	Not

Br

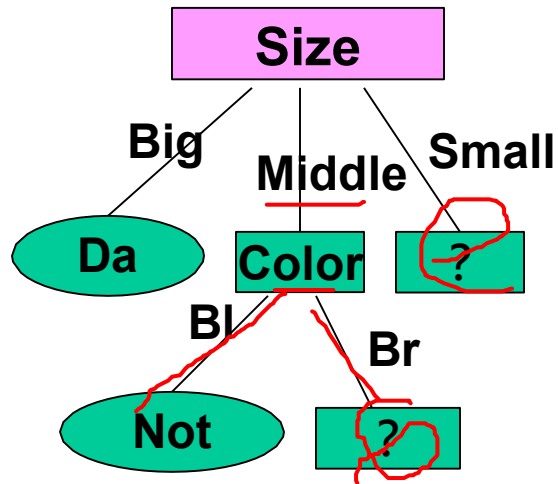
Num	Color	Size	Hair	Class
3	Br	M	C	Not
5	Br	M	S	Da

Bl

Num	Color	Size	Hair	Class
10	Bl	M	C	Not
11	Bl	M	S	Not



Construct the Decision Tree



Num	Color	Size	Hair	Class
4	Bl	S	C	Not
7	Br	S	C	Dangerous
8	Br	S	S	Not
12	Bl	S	S	Not

Color

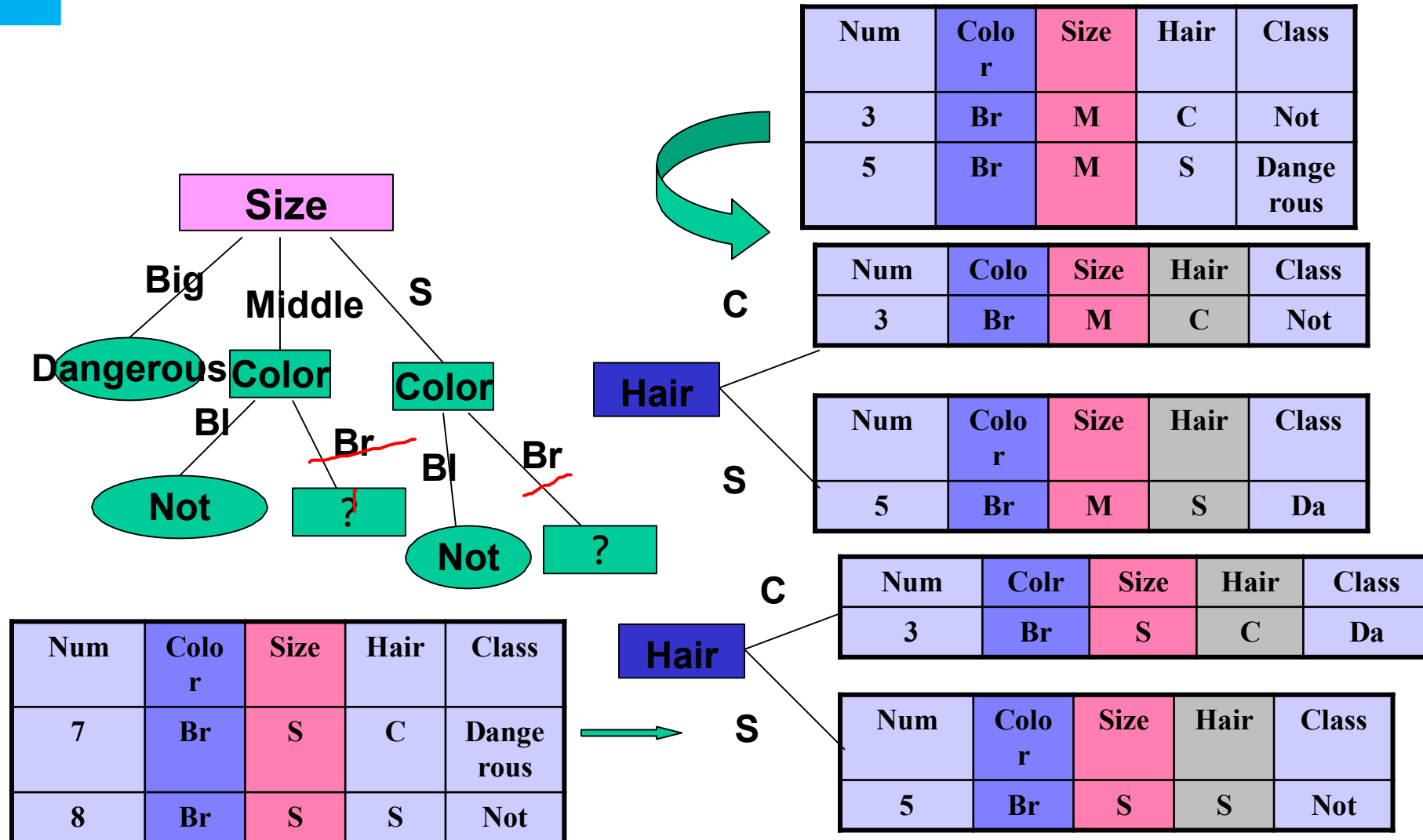
Bl

Br

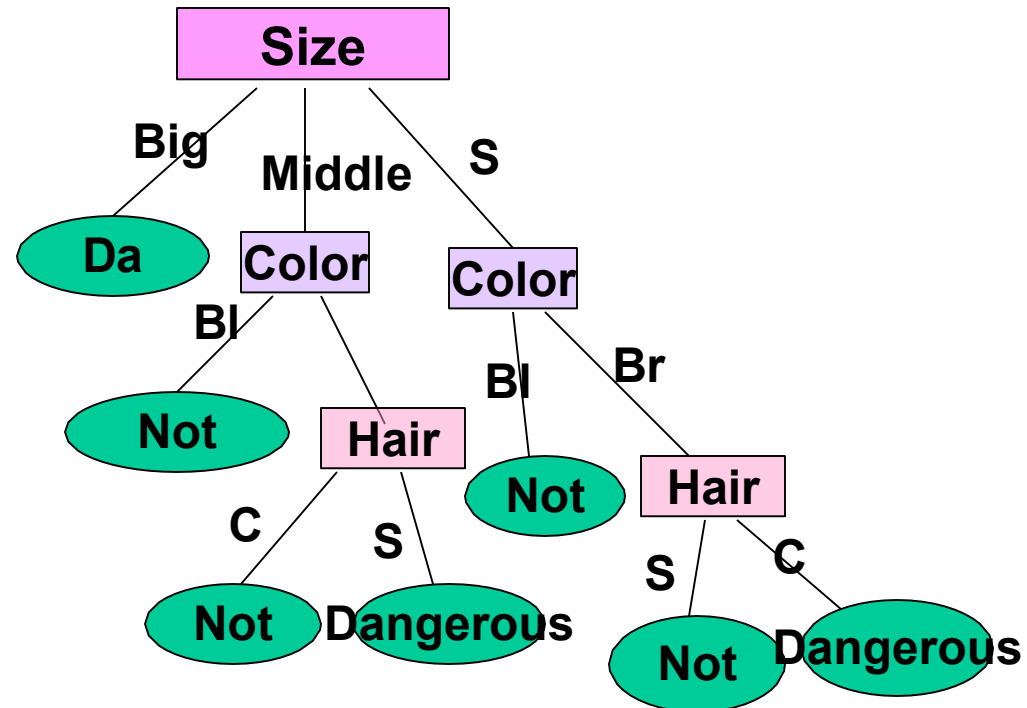
Num	Color	Size	Hair	Class
4	Bl	S	C	Not
12	Bl	S	S	Not

Num	Color	Size	Hair	Class
7	Br	S	C	Da
8	Br	S	S	Not

Construct the Decision Tree



■ Finally the Decesion Tree



Construct the Decision Tree – ID3 Algorithm

- Select the node (attribute).
- ID3(Iterative Dichotomiser 3) algorithm is an algorithm used to generate a decision tree invented by Ross Quinlan
- Summarize of the ID3 algorithm
 - Take all unused attributes and count their entropy concerning test samples
 - Choose attribute for which entropy is minimum (information gain is maximum)
 - Make node containing that attribute

Algorithm3 Construct the Decision Tree - ID

- Entropy : S is a set which contain n training set, S is divided into c classes, each class contain n_i , then the entropy of S divided by c classes is.

$$E(S) = - \sum_{i=1}^c \frac{n_i}{n} \log_2 \left(\frac{n_i}{n} \right) = - \sum_{i=1}^c P_i \log_2(P_i)$$

$$P_+ \log_2 P_+ + P_- \log_2 P_-$$

Examples – ID3 Algorithm

?Play Tennis

Day	Outlook	Temperature	Humidity	Wind	Play
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Examples – ID3 Algorithm

- The information gain of the S Sample is:

$$E(S) = -\frac{9}{14} \log_2 \left(\frac{9}{14} \right) - \frac{5}{14} \log_2 \left(\frac{5}{14} \right) = 0.94$$

- Attribute “Outlook”, compute its information gain: three properties “Sunny” “Overcast” “Rain”

Outlook	Play
Sunny	No
Sunny	No
Overcast	Yes
Rain	Yes
Rain	Yes
Rain	No
Overcast	Yes
Sunny	No
Sunny	Yes
Rain	Yes
Sunny	Yes
Overcast	Yes
Overcast	Yes
Rain	No

Examples – ID3 Algorithm

Outlook	Play
Sunny	No
Sunny	No
Sunny	No
Sunny	Yes
Sunny	Yes

$$E(S_{\text{sunny}}) = -\frac{2}{5} \log_2 \left(\frac{2}{5} \right) - \frac{3}{5} \log_2 \left(\frac{3}{5} \right) = 0.971$$

$$E(S_{\text{Overcast}}) = -\frac{4}{4} \log_2 \left(\frac{4}{4} \right) - \frac{0}{4} \log_2 \left(\frac{0}{4} \right) = 0$$

$$E(S_{\text{Rain}}) = -\frac{3}{5} \log_2 \left(\frac{3}{5} \right) - \frac{2}{5} \log_2 \left(\frac{2}{5} \right) = 0.971$$

$$E(S, \text{Outlook}) = \frac{5}{14} E(S_{\text{Sunny}}) + \frac{4}{14} E(S_{\text{Overcast}}) + \frac{5}{14} E(S_{\text{Rain}}) = 0.694$$

Outlook	Play
Overcast	Yes
Overcast	Yes
Overcast	Yes
Overcast	Yes

Outlook	Play
Rain	Yes
Rain	Yes
Rain	No
Rain	Yes
Rain	No



Examples – ID3 Algorithm

- The information gain of the attribute “Outlook” is:

$$\text{Gain}(S, \text{Outlook}) = E(S) - E(S, \text{Outlook}) = 0.94 - 0.694 = 0.246$$

- Use the same method we can compute the information gain of attributes “Humidity” “Temperature” and “Wind”

$$\text{Gain}(S, \text{Humidity}) = 0.151$$

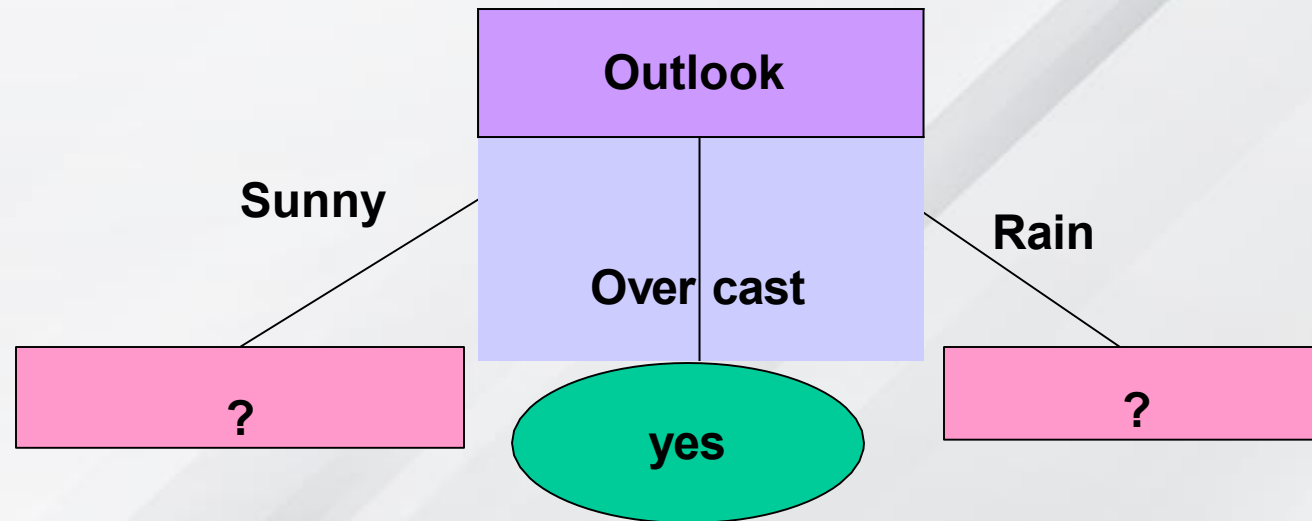
$$\text{Gain}(S, \text{Temperature}) = 0.029$$

$$\text{Gain}(S, \text{Wind}) = 0.048$$

- Based on the computation above, we choose the attribute “Outlook” as the root node.

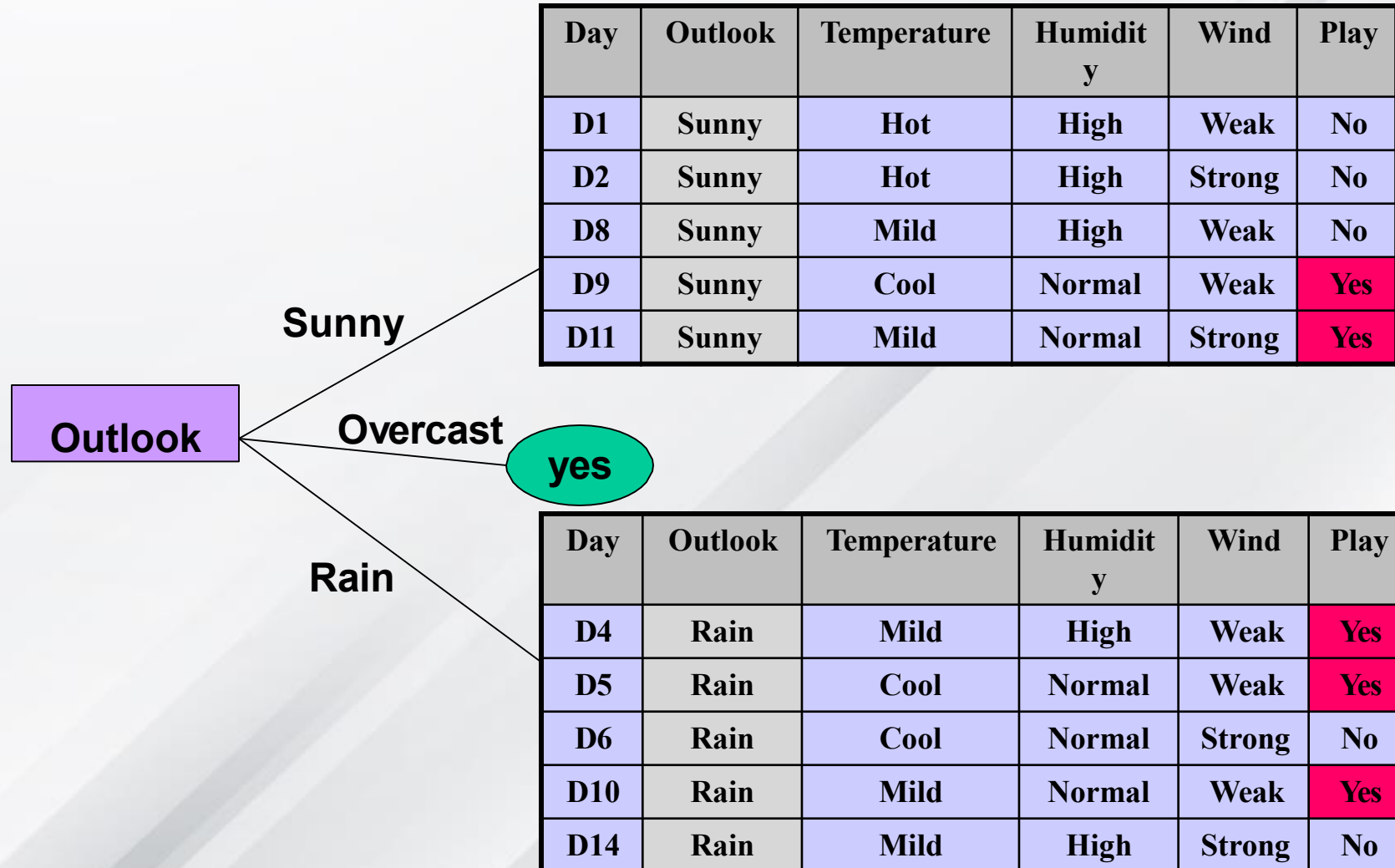
Examples – ID3 Algorithm

- The first stage decision tree



Examples – ID3 Algorithm

- The first stage decision tree



Examples – ID3 Algorithm

- Continue to construct the decision at the branch

“outlook”=“sunny.”

- Compute the entropy

$$E(S) = -\frac{3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log_2\left(\frac{2}{5}\right) = 0.971$$

Temperature	Humidity	Wind	Play
Hot	High	Weak	No
Hot	High	Strong	No
Mild	High	Weak	No
Cool	Normal	Weak	Yes
Mild	Normal	Strong	Yes

- Compute the information gain of “Temperature”
“Humidity” and “Wind”

$$Gain(S, Temperature) = E(S) - E(S, Temperature) = 0.971 - 0.4 = 0.571$$

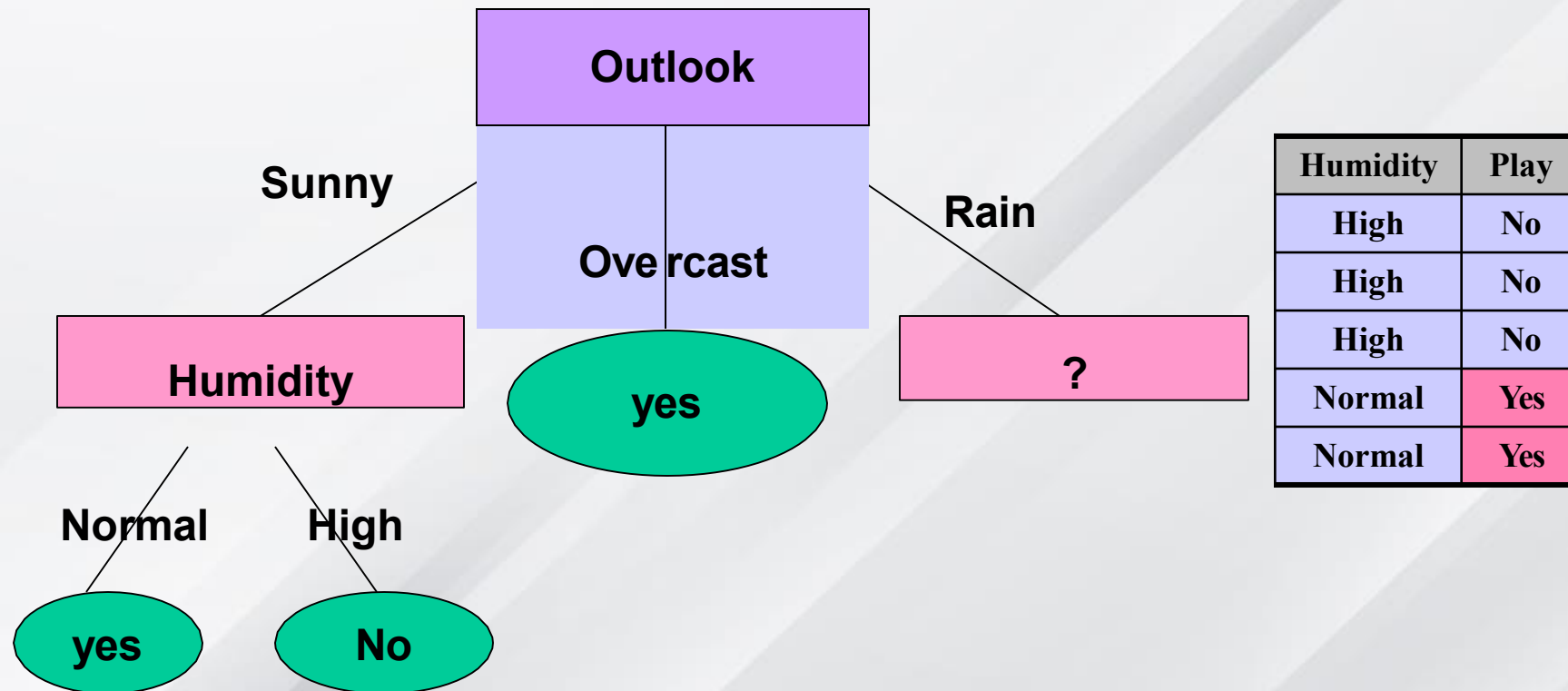
$$Gain(S, Humidity) = 0.971$$

$$Gain(S, Wind) = 0.02$$

- Select “Humidity” as current node

Examples – ID3 Algorithm

- The current stage decision tree



Examples – ID3 Algorithm

- Continue to construct the decision at the branch
“outlook”=“rain.”
- Compute the entropy

$$E(S) = -\frac{3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log_2\left(\frac{2}{5}\right) = 0.971$$

Temperature	Humidity	Wind	Play
Mild	High	Weak	Yes
Cool	Normal	Weak	Yes
Cool	Normal	Strong	No
Mild	Normal	Weak	Yes
Mild	High	Strong	No

- Compute the Entropy of “mild” “cool”

$$E(S_{Mild}) = -\frac{2}{3}\log_2\left(\frac{2}{3}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right) = 0.918$$

$$E(S_{Cool}) = -\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right) = 1$$

$$\begin{aligned}
 E(S, Temperature) &= \frac{3}{5}E(S_{Mild}) + \frac{2}{5}E(S_{Cool}) \\
 &= \frac{3}{5} \times 0.918 + \frac{2}{5} \times 1 = 0.951
 \end{aligned}$$

Temperature	Play
Mild	Yes
Mild	Yes
Mild	No

Temperature	Play
Cool	Yes
Cool	No

Examples – ID3 Algorithm

- The information gain of “temperature” is

$$Gain(S, Temperature) = E(S) - E(S, Temperature) = 0.971 - 0.951 = 0.02$$

- The information gain of “Humidity” and “Wind” is

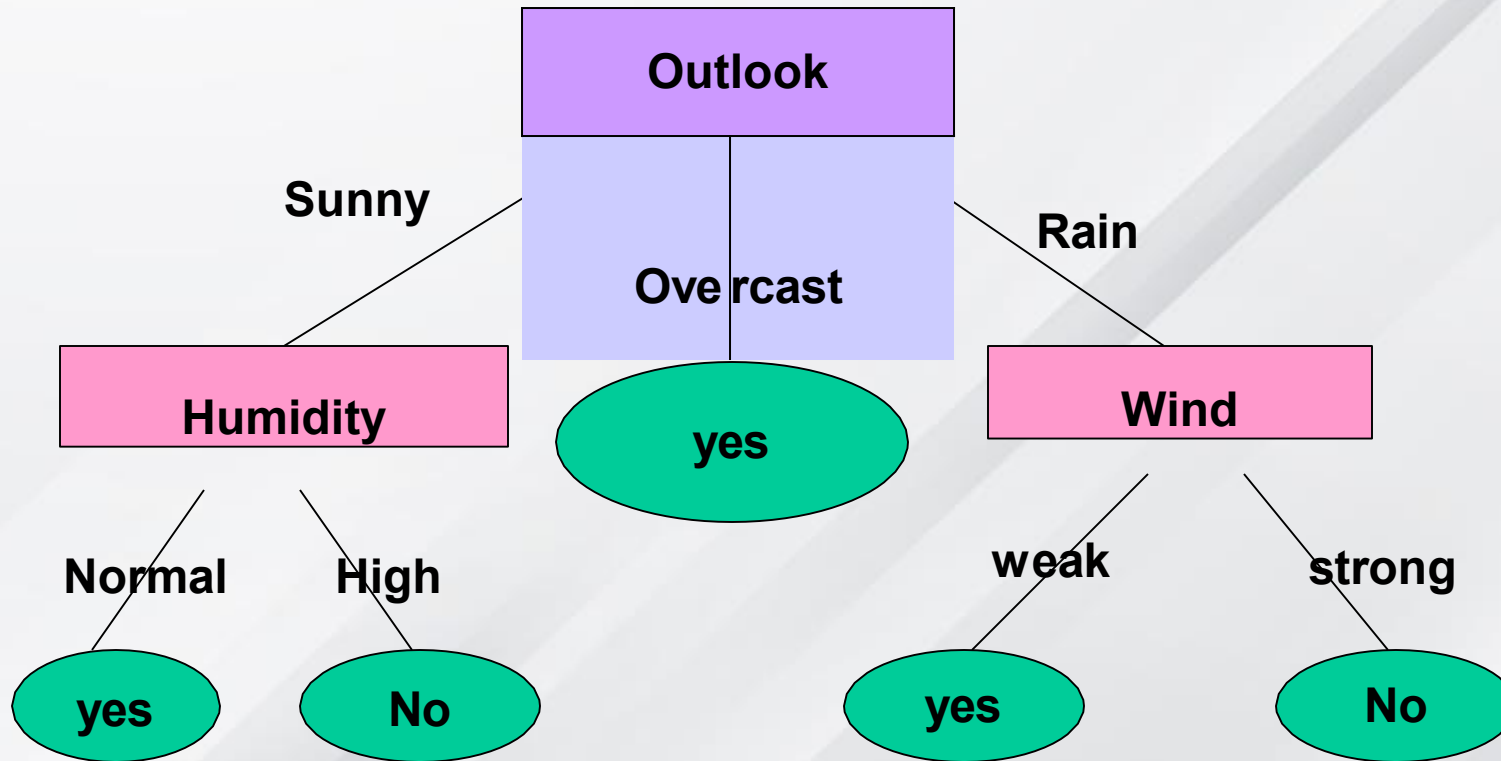
$$Gain(S, Humidity) = 0.02$$

$$Gain(S, Wind) = 0.971$$

- So select “Wind” as current node

Examples – ID3 Algorithm

- The final decision tree



Construct the Decision Tree - ID3 Algorithm

- Example: the entropy of two classes (Dangerous, Not Dangerous)
- is:

$$E(S) = -\frac{6}{12} \log_2 \frac{6}{12} - \frac{6}{12} \log_2 \frac{6}{12} (= \frac{1}{2} + \frac{1}{2} = 1)$$

Class
Da
Da
Not
Not
Da
Da
Da
Not
Da
Not
Not
Not

Construct the Decision Tree - ID3 Algorithm

- Information gain: In general terms, the information gain is the change in information entropy from a prior state to a state (new classification) that takes some information as given:
- The information gain of attribute A in according to training set S is:

$$Gain(S, A) \equiv Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

Construct the Decision Tree – ID3 Algorithm

- Based on attribute “Shape” there are three branches, each has two classes (Dangerous or Not Dangerous).

$$E(S_1) = -\frac{0}{4} \log_2\left(\frac{0}{4}\right) - \frac{4}{4} \log_2\left(\frac{4}{4}\right) = 0 + 0 = 0$$

$$E(S_2) = -\frac{1}{4} \log_2\left(\frac{1}{4}\right) - \frac{3}{4} \log_2\left(\frac{3}{4}\right) = 0.5 + 0.0637 = 0.5637$$

$$E(S_3) = -\frac{1}{4} \log_2\left(\frac{1}{4}\right) - \frac{3}{4} \log_2\left(\frac{3}{4}\right) = 0.5 + 0.0637 = 0.5637$$

$$\begin{aligned} E(X, S) &= \frac{4}{12} E(S_1) + \frac{4}{12} E(S_2) + \frac{4}{12} E(S_3) \\ &= \frac{4}{12} \times 0 + \frac{4}{12} \times 0.5637 + \frac{4}{12} \times 0.5637 = 0.3758 \end{aligned}$$

Num	Color	Size	Hair	Class
1	Black	Big	Curls	D
2	Brown	Big	Straight	D
6	Black	Big	Straight	D
9	Brown	Big	Curl	D

Num	Color	Size	Hair	Class
3	Brown	M	Curl	No
5	Brown	M	Straigh	D
10	Black	M	Curl	No
11	Black	M	Straith	No

Num	Color	Size	Hair	Class
4	Black	Small	Curl	No
7	Brown	Small	Curl	D
8	Brown	Small	Straight	No
12	Black	Small	Straight	No

Construct the Decision Tree – ID3 Algorithm

- Based on attribute “Size” the information gain is:

$$Gain(X, S) = E(S) - E(X, S) = 1 - 0.3758 = 0.6242$$

- There are also some other attributes such as: “Color” “Hair”, select the biggest information gain to be the root node... to construct the decision tree... classify.

Construct the Decision Tree – ID3 Algorithm

A decision tree can be constructed top-down using the information gain in the following way:

- begin at the root node
- determine the attribute with the highest information gain which is not already used as an ancestor node
- add a child node for each possible value of that attribute
attach all examples to the child node where the attribute values of the examples are identical to the attribute value attached to the node
- if all examples attached to the child node can be classified uniquely add that classification to that node and mark it as leaf node
- go back to step two if there is at least one more unused attribute left, otherwise add the classification of most of the examples attached to the child node

- ID3 is not optimal
 - Uses *expected* entropy reduction, not actual reduction
- Must use discrete (or discretized) attributes
 - What if the wind is just fine(neither strong nor weak ?)
 - We could break down the attributes into smaller values...

- While decision trees classify quickly, the time for building a tree may be higher than another type of classifier
- Decision trees suffer from a problem of errors propagating throughout a tree
 - A very serious problem as the number of classes increases

- Since decision trees work by a series of local decisions, what happens when one of these local decisions is wrong?
 - Every decision from that point on may be wrong
 - We may never return to the correct path of the tree

- Decision trees can be used to help predict the future
- The trees are easy to understand
- Decision trees work more efficiently with discrete attributes
- The trees may suffer from error propagation



Thank You