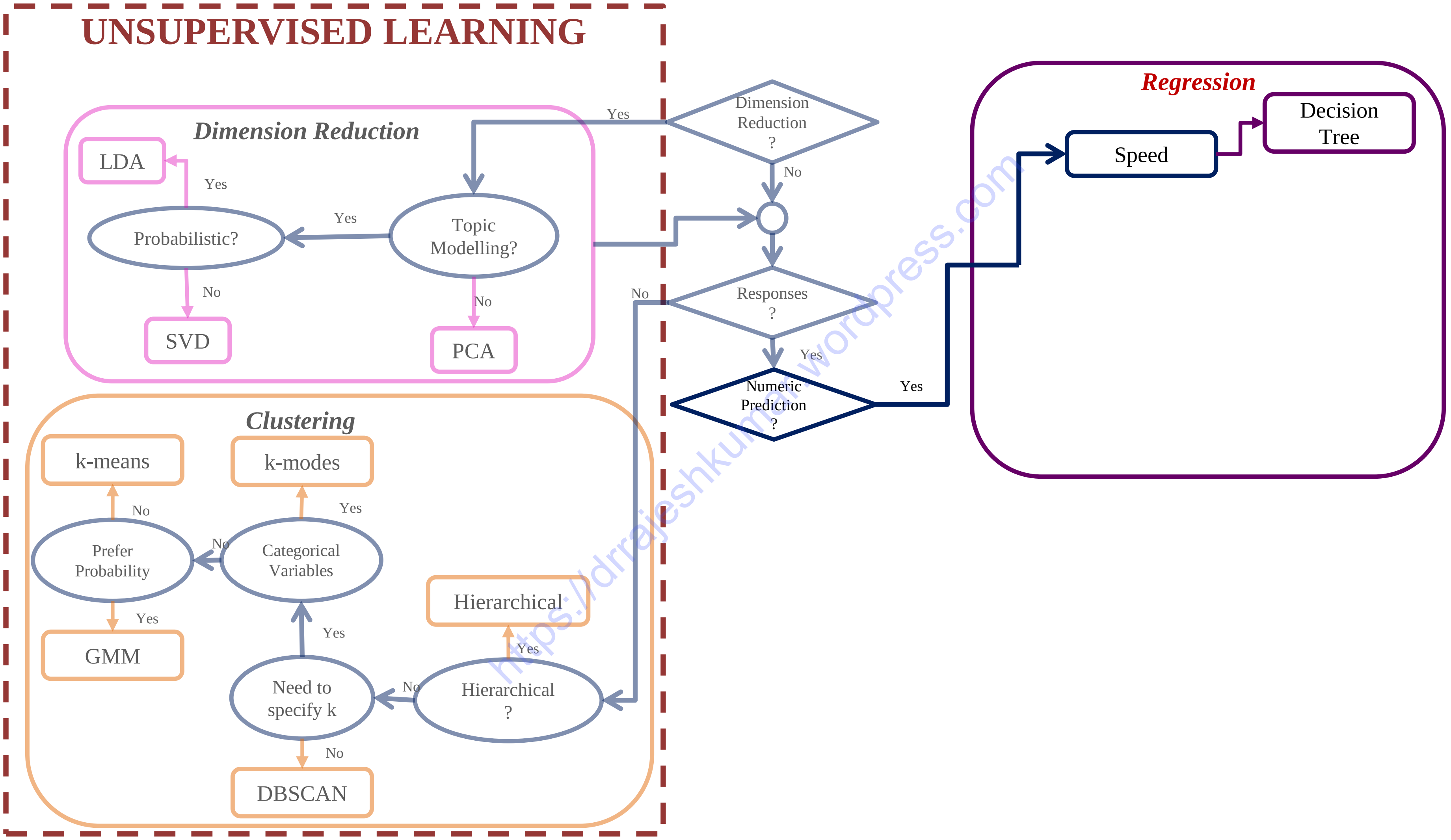
Machine Learning

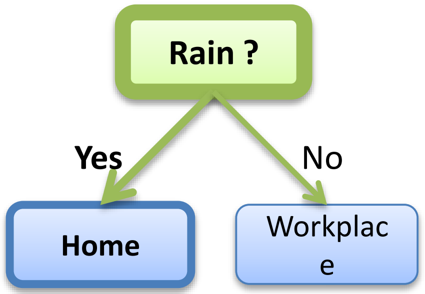
Dr. Rajesh Kumar

PhD, PDF (NUS, Singapore)
SMIEEE, FIET (UK), FIETE, FIE (I),SMIACSIT, LMISTE, MIAENG
Professor, Department of Electrical Engineering
Professor, Centre of Energy and Environment
Malaviya National Institute of Technology, Jaipur, India, 302017
Tel: (91) 9549654481
http://drrajeshkumar.wordpress.com
rkumar.ee@mnit.ac.in, rkumar.ee@gmail.com


UNSUPERVISED LEARNING
Dimension Reduction
LDA
Yes
Probabilistic?
Yes
Topic Modelling?
No
SVD
No
PCA
Yes
Dimension Reduction ?
No
Responses ?
No
Yes
Numeric Prediction ?
Yes
Regression
Speed
Decision Tree
Clustering
k-means
k-modes
No
Yes
Prefer Probability
No
Categorical Variables
Yes
GMM
Hierarchical
Yes
Yes
Need to specify k
No
Hierarchical ?
No
DBSCAN
https://drrajeshkumar.wordpress.com

Decision Tree

- Decision tree is a classifier in the form of a *tree structure*
  - Decision node:
    Specifies a test on a single attribute
  - Leaf node:
    Indicates the value of the target attribute
  - Arc/edge:
    Split of one attribute
  - Path:
    A disjunction of test to make the final decision



- Decision trees classify instances or examples by starting at the root of the tree and moving through it until a leaf node

Decision Tree

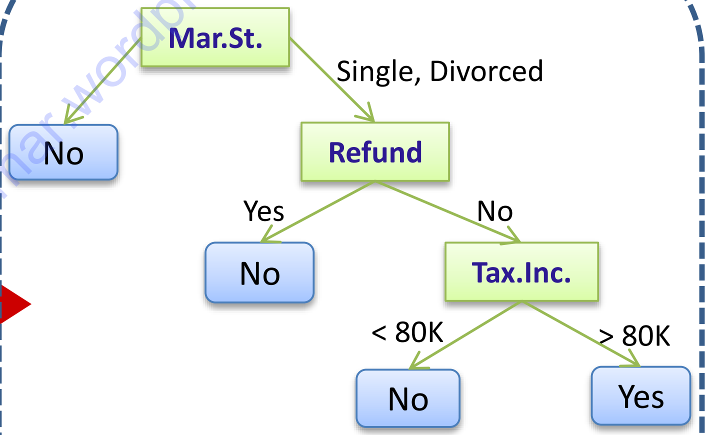
- Decision trees are powerful and popular tools for
  - Classification
  - Prediction
- Decision trees represent rules, which can be understood by humans and used in knowledge system such as database.
- Relatively fast compared to other classification models

Example of a Decision Tree

categorical categorical continuous class

Tid	Refund	Marital Status	Taxable Income	Positive Attitude
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

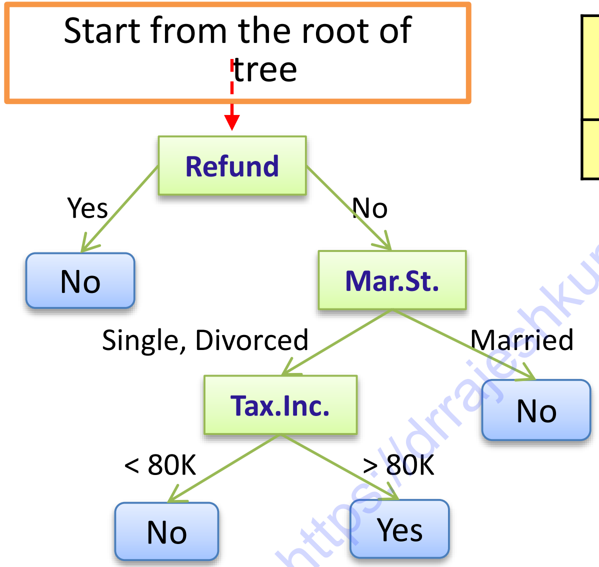
Training Data



Note : There can be more than one decision tree model

Model: Decision Tree

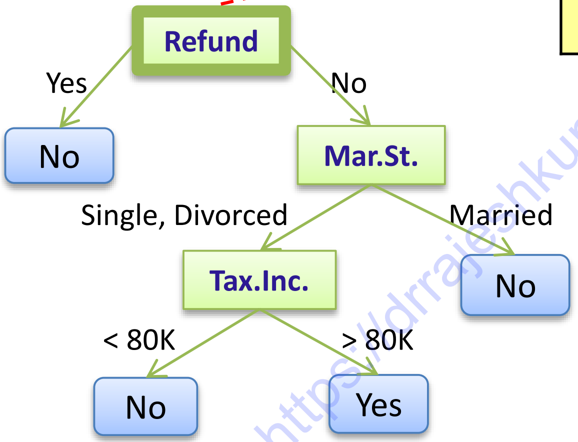
Apply Model to Test Data



Test Data

Refund	**Marital Status**	**Taxable Income**	**Positive Attitude**
No	Married	80k	?

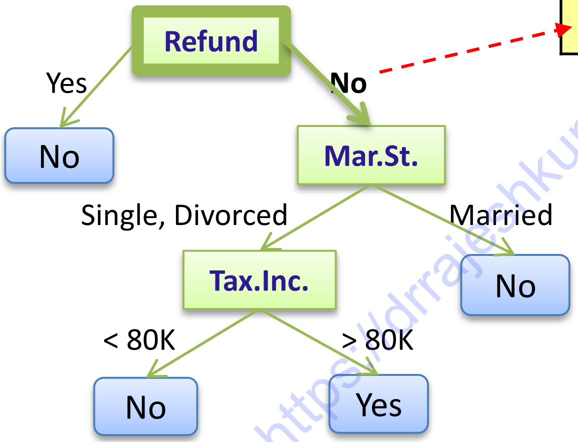
Apply Model to Test Data



Test Data

Refund	Marital Status	Taxable Income	Positive Attitude
No	Married	80k	?

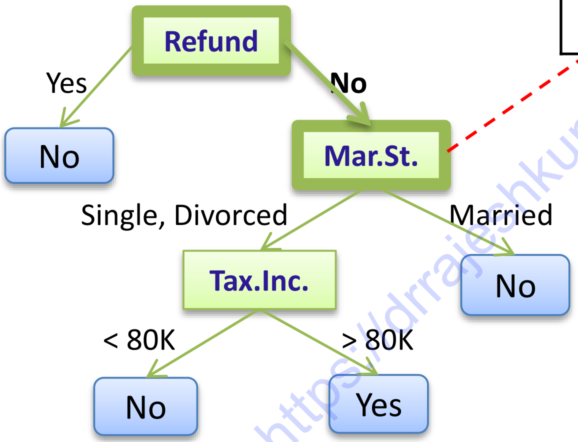
Apply Model to Test Data



Test Data

Refund	Marital Status	Taxable Income	Positive Attitude
No	Married	80k	?

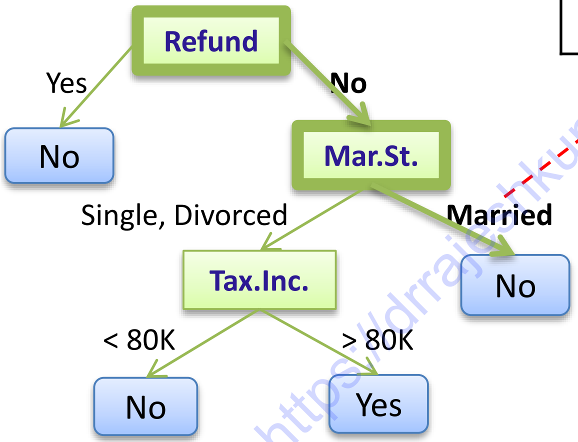
Apply Model to Test Data



Test Data

Refund	**Marital Status**	**Taxable Income**	**Positive Attitude**
No	Married	80k	?

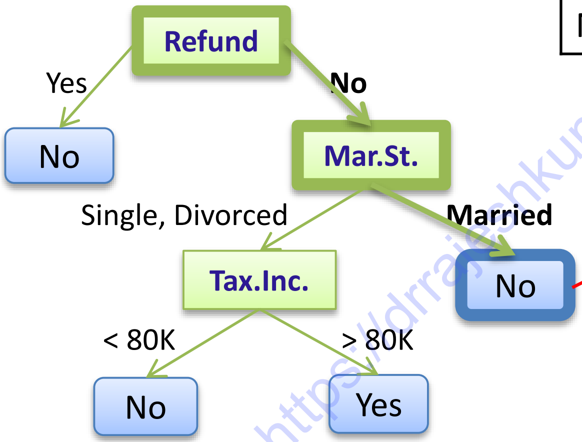
Apply Model to Test Data



Test Data

Refund	Marital Status	Taxable Income	Positive Attitude
No	Married	80k	?

Apply Model to Test Data



Test Data

Refund	Marital Status	Taxable Income	Positive Attitude
No	Married	80k	No

Predict if John will play tennis

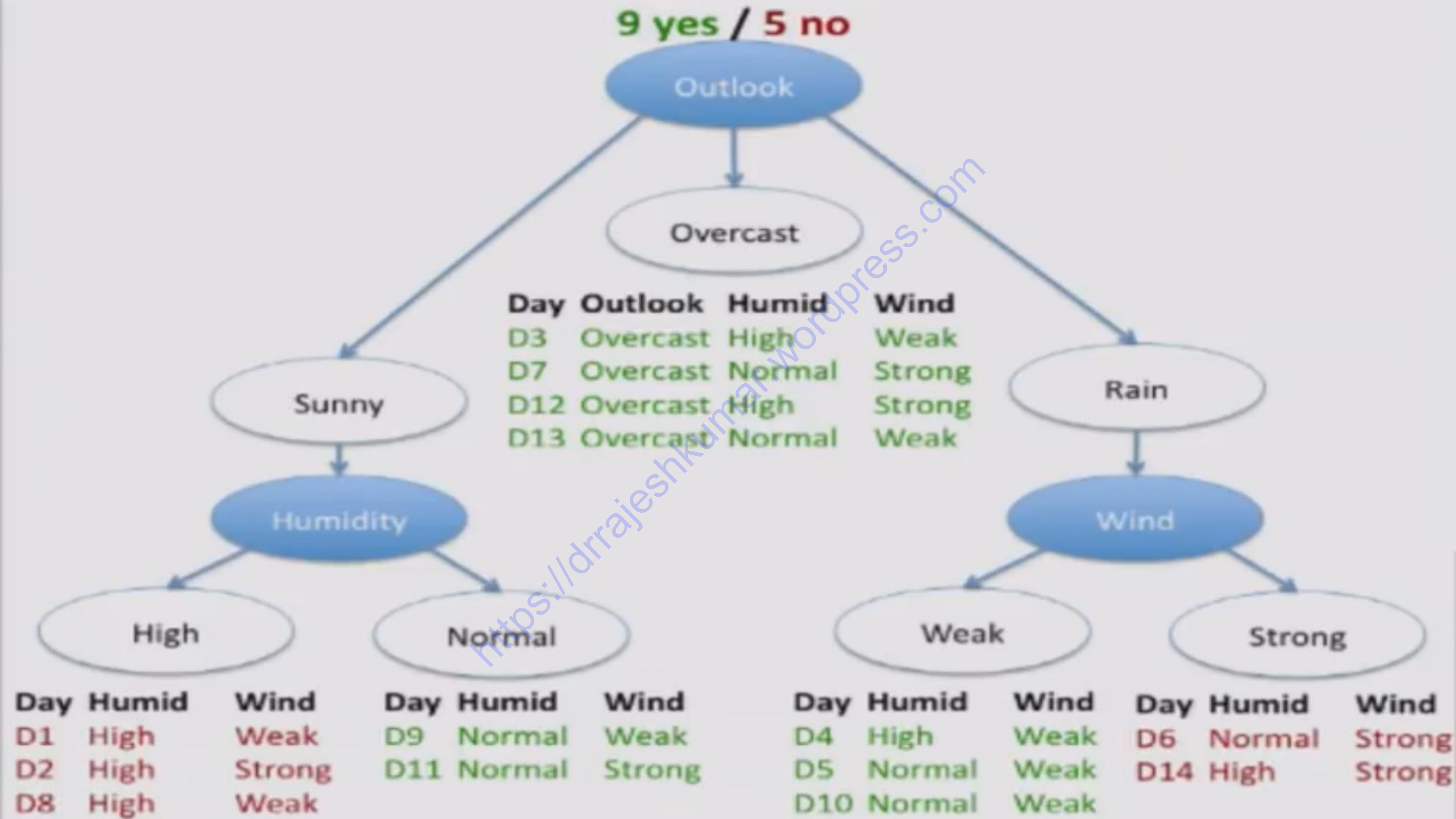
- Hard to guess
- Try to *understand* when John plays
- Divide & conquer:
  - split into subsets
  - are they pure? (all yes or all no)
  - if yes: stop
  - if not: repeat
- See which subset new data falls into

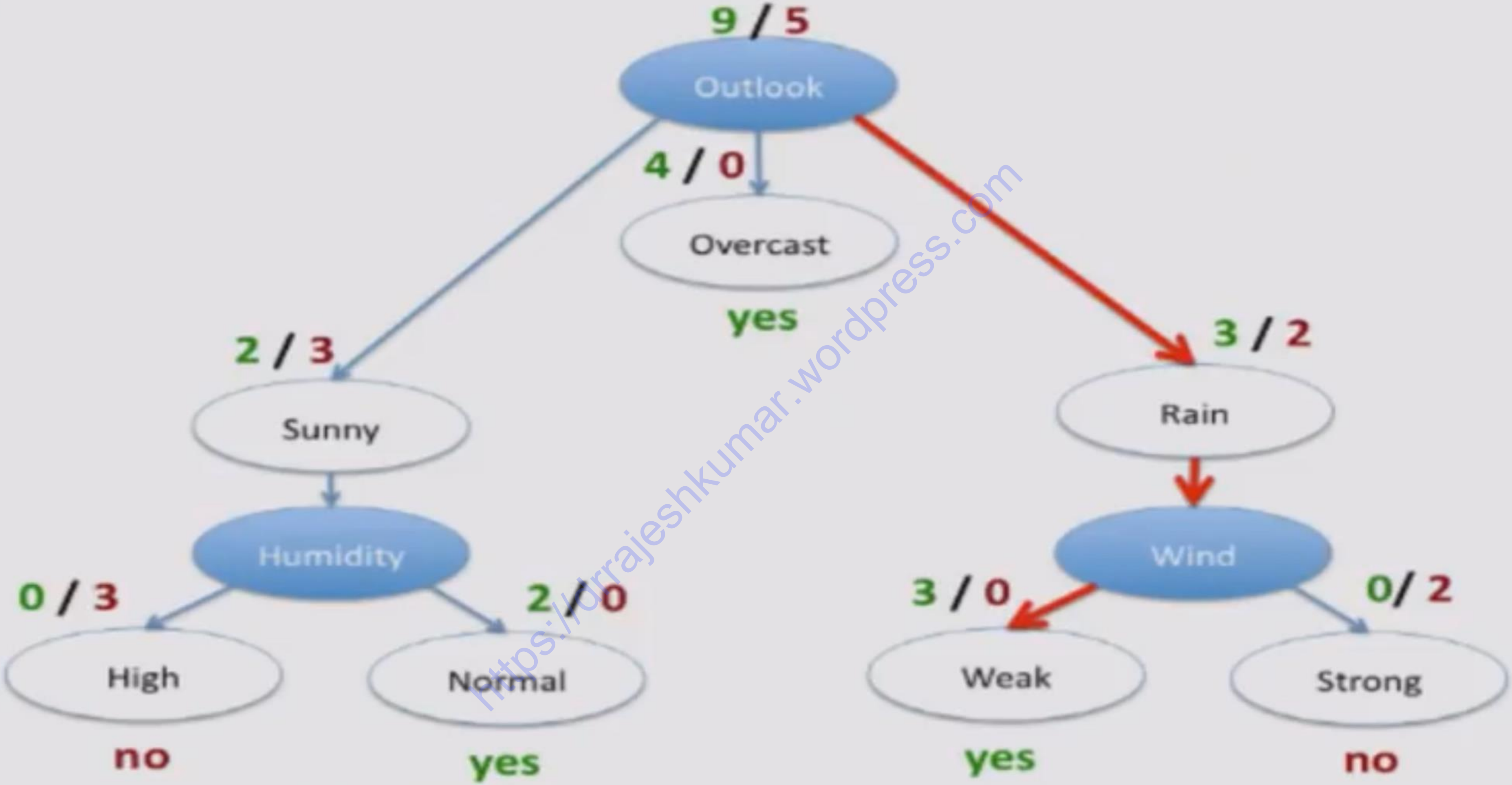
Training examples: 9 yes / 5 no

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

New data:

D15	Rain	High	Weak	?

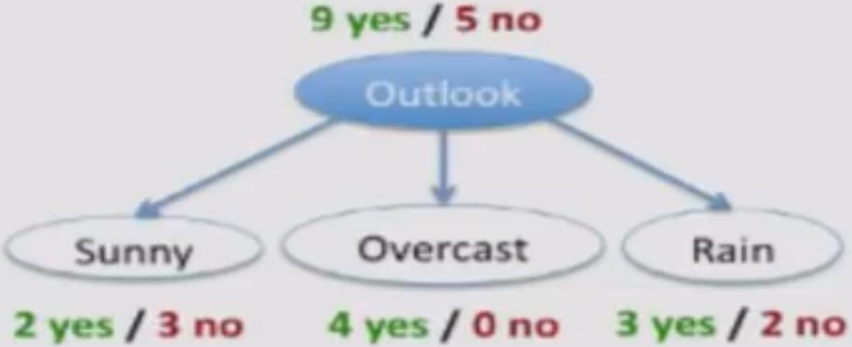

9 yes / 5 no
Outlook
Overcast
Day Outlook Humid Wind
D3 Overcast High Weak
D7 Overcast Normal Strong
D12 Overcast High Strong
D13 Overcast Normal Weak
Sunny
Rain
Humidity
Wind
High
Normal
Weak
Strong
Day Humid Wind
D1 High Weak
D2 High Strong
D8 High Weak
Day Humid Wind
D9 Normal Weak
D11 Normal Strong
Day Humid Wind
D4 High Weak
D5 Normal Weak
D10 Normal Weak
Day Humid Wind
D6 Normal Strong
D14 High Strong
https://drrajeshkumar.wordpress.com

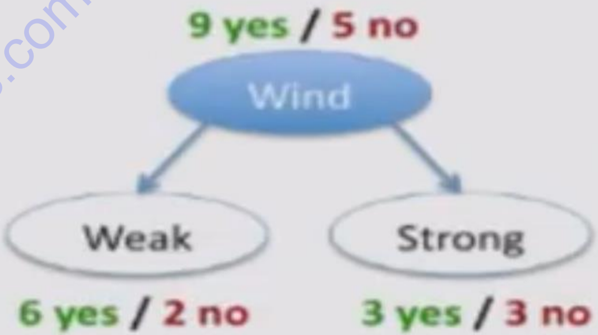


New data:

Day	Outlook	Humid	Wind	
D15	Rain	High	Weak	→ Yes

Which attribute to split on?





- Want to measure "purity" of the split
  - more certain about Yes/No after the split
    - pure set (4 yes / 0 no) => completely certain (100%)
    - impure (3 yes / 3 no) => completely uncertain (50%)
  - can't use P("yes" | set):
    - must be symmetric: 4 yes / 0 no as pure as 0 yes / 4 no

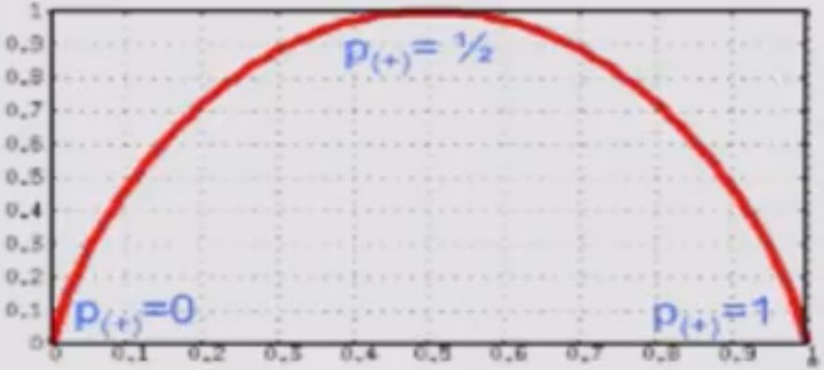
Entropy

- Entropy: $H(S) = - p_{(+)} \log_2 p_{(+)} - p_{(-)} \log_2 p_{(-)}$ bits
  - S ... subset of training examples
  - $p_{(+)}$ / $p_{(-)}$ ... % of positive / negative examples in S
- Interpretation: assume item X belongs to S
  - how many bits need to tell if X positive or negative
- impure (3 yes / 3 no):

$$H(S) = -\frac{3}{6}\log_2\frac{3}{6} - \frac{3}{6}\log_2\frac{3}{6} = 1 \text{ bits}$$

- pure set (4 yes / 0 no):

$$H(S) = -\frac{4}{4}\log_2\frac{4}{4} - \frac{0}{4}\log_2\frac{0}{4} = 0 \text{ bits}$$



Information Gain

- Want many items in pure sets
- Expected drop in entropy after split:

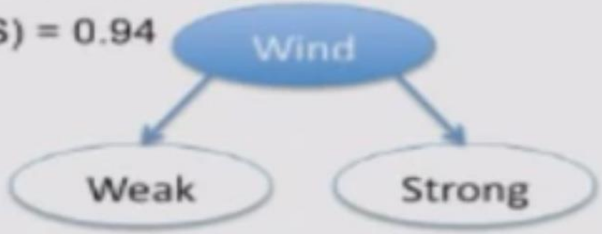
$$Gain(S,A) = H(S) - \sum_{V \in Values(A)} \frac{|S_V|}{|S|} H(S_V)$$

V ... possible values of A
S ... set of examples {X}
S_V ... subset where $X_A = V$

- Mutual Information
  - between attribute A and class labels of S

$-\frac{9}{14}\log_2\frac{9}{14} - \frac{5}{14}\log_2\frac{5}{14}$ 9 yes / 5 no

H(S) = 0.94



6 yes / 2 no

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

$H(S_{weak}) = 0.81$

3 yes / 3 no

$-\frac{3}{6}\log_2\frac{3}{6} - \frac{3}{6}\log_2\frac{3}{6}$

$H(S_{strong}) = 1.0$

Gain (S, Wind)
$= H(S) - {}^{8}/_{14}\, H(S_{weak}) - {}^{6}/_{14}\, H(S_{weak})$
$= 0.94 - {}^{8}/_{14} * 0.81 - {}^{6}/_{14} * 1.0$
$= 0.049$