Logic-based Approaches

Chirayu Wongchokprasitti, PhD

University of Pittsburgh

Center for Causal Discovery

Department of Biomedical Informatics

[*chw20@pitt.edu*](mailto:chw20@pitt.edu)

[*http://www.pitt.edu/~chw20*](http://www.pitt.edu/~chw20)

Overview

- Predictive Modeling and Logic-based Approaches
- Inductive Logic Programming
- Classification and Regression Tree (CART)
- ID3 & C4.5 Tree Learning
- Tree Pruning & Model Evaluation
- Association Rule Mining

Introduction

Is logic embedded in our mind?
Innate knowledge: Socrates' dialogue with a slave boy

http://en.wikipedia.org/wiki/Meno#Dialogue_with_Meno.27s_slave
http://www.gutenberg.org/files/1643/1643-h/1643-h.htm

SOCRATES: It will be no easy matter, but I will try to please you to the utmost of my power. Suppose that you call one of your numerous attendants, that I may demonstrate on him.

MENO: Certainly. Come hither, boy.

SOCRATES: He is Greek, and speaks Greek, does he not?

MENO: Yes, indeed; he was born in the house.

SOCRATES: Attend now to the questions which I ask him, and observe whether he learns of me or only remembers.

MENO: I will.

SOCRATES: Tell me, boy, do you know that a figure like this is a square?

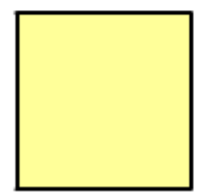
BOY: I do.

SOCRATES: And you know that a square figure has these four lines equal?

BOY: Certainly.

SOCRATES: And these lines which I have drawn through the middle of the square are also equal?

…



Valid and Invalid Arguments for Conditional Logic

http://www.psychologyinaction.org/2012/10/07/classic-psychology-experiments-wason-selection-task-part-i/

Affirming the Antecedent (modus ponens)

1. P ➔Q
2. P

---

3. Q

Denying the Consequent (modus tollens)

1. P ➔Q
2. ~Q

---

3. ~P

Affirming the Consequent (INVALID)

1. P ➔Q
2. Q

---

3. P

Denying the Consequent (INVALID)

1. P ➔Q
2. ~P

---

3. ~Q

Are there domain independent rules of logic in human reasoning?

[Wason & Johnson-Laird]

Each of the following cards has a letter on one side and a number on the other side.

E	K	4	7

Which of the above cards need to be turned over to test the following rule?

If a card has a vowel on one side then it has an even number on the other side

Rules of logic in human reasoning

Consider the following isomorph of this problem:
Each card has a name of a scientific meeting on one side and a means of transportation on the other side.

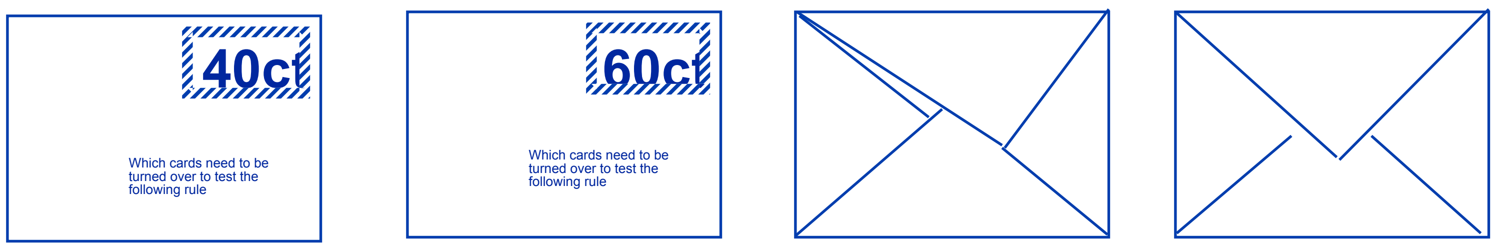
New York City	Washington DC	train	car

Which of the above cards need to be turned over to test the following rule?

Every time I go to a New York City, I travel by train

Rules of logic in human reasoning

Consider the following isomorph of this problem:



Which of the above envelopes need to be turned over to test the following rule?

If an envelope is sealed, it has a 60ct stamp on it

Rules of logic in human reasoning

The original problem (letters and numbers) is solved correctly by about 12% of the subjects, its isomorphs by over 60% of the subjects [Wason & Johnson-Laird].

The results of these (and other) experiments suggest strongly that human reasoning is not based on abstract, domain independent rules, but rather depends heavily on the content.

Predictive Modeling and Logic-based Approaches

Data Mining Tools

- GNU R project with RStudio & Rattle

  http://www.r-project.org/

  http://www.rstudio.com/ide/

  http://rattle.togaware.com/
- Mathworks MATLAB Statistical Toolbox
- Weka (University of Waikato)

  http://www.cs.waikato.ac.nz/ml/weka/

Many other tools such as SAS STAT/EM, IBM SPSS Clementine, Python SciPy/Numpy, etc.

Predictive modeling

A process to create a model that best predicts a (continuous or discrete) outcome.

e.g., use customers' gender, age, and purchase history to predict future sales.

Discover patterns

Make predictions

Identify risks and opportunities, etc.

Logic-based approach

Based on the idea of using logical sentences/rules to represent knowledge learned from data.

For example:

A set of *IF-THEN rules* ***could be*** **learned from a car dealer's marketing database, such as:**

Rule 1: IF *age >=* ***40*** AND *annual_income >=****70K***
THEN *Buy_our_SUV* = ***TRUE***

Rule 2: IF *annual_income <* ***70K***
THEN *Buy_our_SUV* = ***FALSE***

Such rules can be extracted from models created by approaches such as Decision Tree Induction, Inductive Logic Programming, and Association Rule Mining.

Inductive Logic Programming

Inductive Logic Programming (ILP)

- **Automate the induction processes using Logic Programming**
- **Try to find a theory (rules) that covers all positive examples and no negative examples (completeness & consistency)**
- **Derive hypothesis using background and examples**
- **Rules can be used for classification and prediction**

$$Positive\ examples(E^{+}) + Negative\ examples\ (E^{-}) + Background\ \ knowledge\ (B)$$
$$=> Hypothesis(H)$$

Inductive Logic Programming (ILP) (cont.)

e.g., learning the rule/hypothesis "*grandparent()*"

Background(B)

```
parent_of(charles,george)
parent_of(george,diana)
parent_of(bob,harry)
parent_of(harry,elizabeth)
...
```

Example(E)

```
grandparent_of(charles,diana)
grandparent_of(bob,elizabeth)
...
```

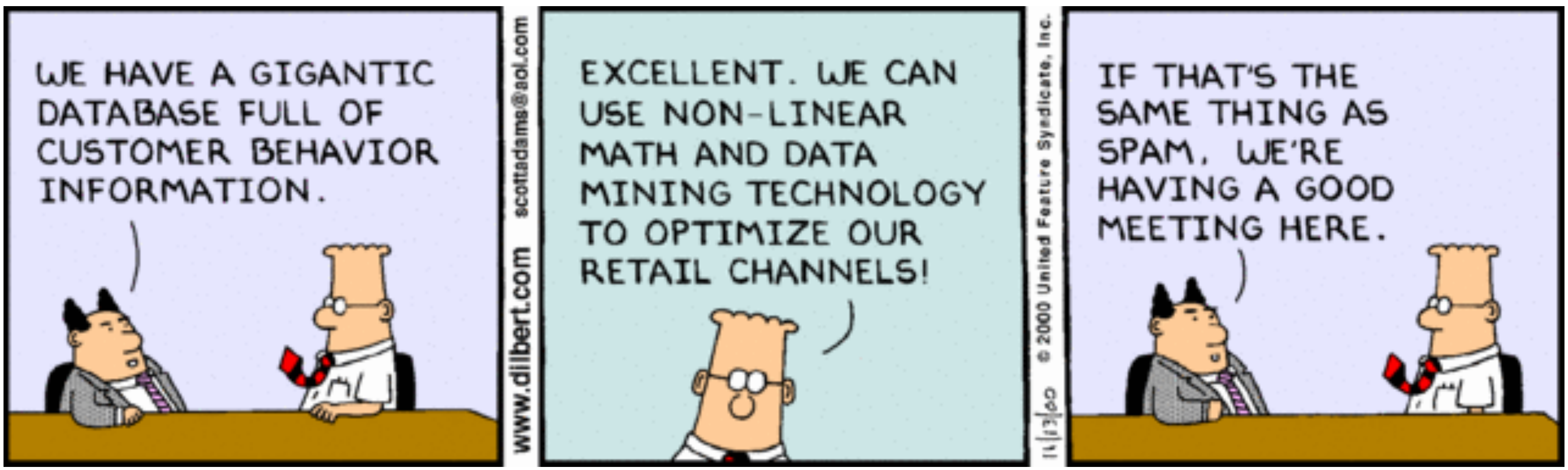
B **+** E **=>** H "*grandparent*()"

```
grandparent_of(X,Y) = parent_of(X,Z),
parent_of(Z,Y)
```

Shortcomings of ILP

- Large hypothesis spaces searched
- High computational demand
- Large numbers of trivial hypotheses derived

Shortcomings of logic-based approaches: Large number of hypotheses derived



Predictive modeling notation

Predictive modeling is about

$$[Y_1, Y_2, \ldots, Y_m; T_1, T_2, \ldots, T_n] = f(X_1, X_2, \ldots, X_i; A_1, A_2, \ldots, A_j)$$

where Y : continuous target/dependent variable

T : categorical target/dependent variable

X :continuous input/independent variable

A : categorical input/independent variable

E.g.

$Buy\_SUV = f(Age, Income)$

$Weather\_Condition = f(\ Temperature, Atmospheric\ Pressure)$

$Patient\_Location = f(Respiratory\ Rate,\ \ Diastolic\ Blood\ \ Pressure,\ \ Systolic\ Blood\ \ Pressure)$

...

Examples

- **Classification Trees**

$$T = f(X_1, X_2, \ldots, X_i; A_1, A_2, \ldots, A_j)$$

- **Regression Trees**

$$Y = f(X_1, X_2, \ldots, X_i; A_1, A_2, \ldots, A_j)$$

- **Support Vector Machines/Regression (SVM/SVR)**

$$T = f(X_1, X_2, \ldots, X_i; A_1, A_2, \ldots, A_j)$$

$$Y = f(X_1, X_2, \ldots, X_i; A_1, A_2, \ldots, A_j)$$

- **K-means Clustering**

$$[X_1, X_2, \ldots, X_i; A_1, A_2, \ldots, A_j]$$

Perspective of a statistician

- Simple/Multiple Linear Regression

$$Y = f(X_1, X_2, \ldots, X_i)$$

- 2-sample t-test , n-way ANOVA

$$Y = f(A) \quad , \quad Y = f(A_1, A_2, \ldots, A_j)$$

- MANOVA

$$[Y_1, Y_2, \ldots, Y_m] = f(A_1, A_2, \ldots, A_j)$$

ANCOVA

$$Y = f(X_1, X_2, \ldots, X_i; A)$$

MANCOVA

$$[Y_1, Y_2, \ldots, Y_m] = f(X_1, X_2, \ldots, X_i; A)$$

General Linear Model: $[Y_1, Y_2, \ldots, Y_m] = f(X_1, X_2, \ldots, X_i; A_1, A_2, \ldots, A_j)$

- Logistic Regression

$$T = f(X_1, X_2, \ldots, X_i; A_1, A_2, \ldots, A_j)$$

Classification and Regression Trees

Classification And Regression Tree (CART)

Classification Tree: $T = f(X_1, X_2, \ldots, X_i; A_1, A_2, \ldots, A_j)$

The target is a categorical variable with different classes (discrete outcomes), e.g.,

$$Weather\ Condition = f(\ Temperature,\ Atmospheric\ Pressure)$$

$$Patient\ Location = f(Respiratory\ Rate,\ \ Diastolic\ Blood\ \ Pressure,\ \ Systolic\ Blood\ \ Pressure, \ldots)$$

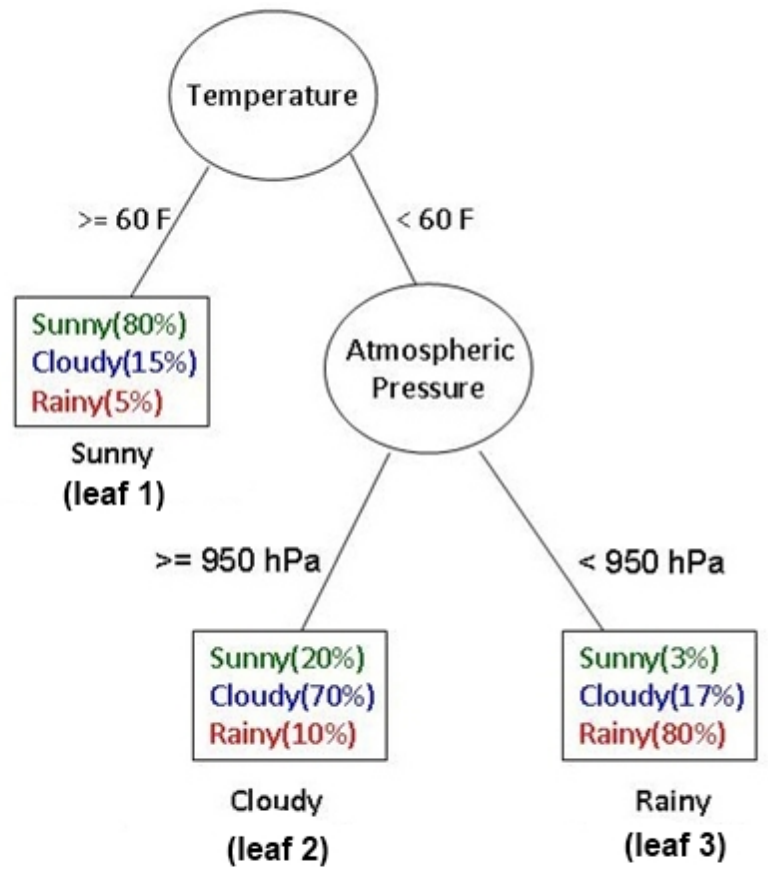
Regression Tree: $Y = f(X_1, X_2, \ldots, X_i; A_1, A_2, \ldots, A_j)$

The target is a continuous variable with real numbers (continuous outcome), e.g.,

$$Price\_of\_house = f(location,\ inflation\ rate, \ldots)$$

$$oxygen\_consumption = f(runtime, gender, age, weight, run\_pulse, rest\_pulse)$$

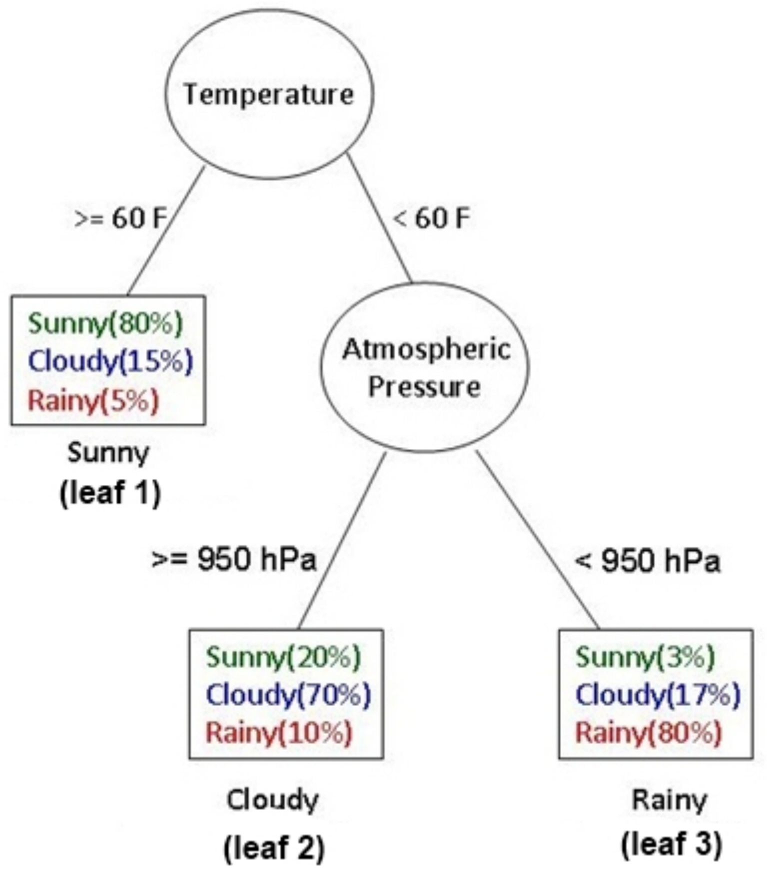
Classification and Regression Tree

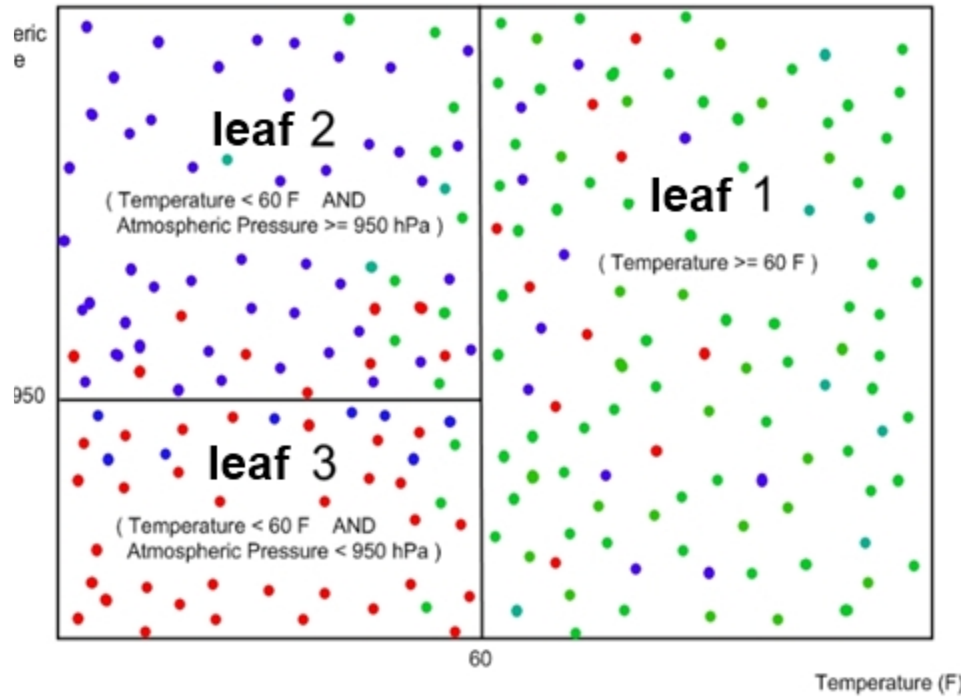


Weather Condition (Sunny, Cloudy, Rainy)	Temperature (F)	Atmospheric Pressure (hPa)
Sunny	65	980
Sunny	70	990
Sunny	55	800
Cloudy	50	960
Cloudy	72	850
Sunny	69	950
Cloudy	75	800
Rainy	60	840
Rainy	61	930
Cloudy	70	970

$$Weather\ Condition = f(\ Temperature,\ Atmospheric\ Pressure)$$

Classification and Regression Tree (cont.)





$$Weather\ Condition = f(\ Temperature,\ Atmospheric\ Pressure)$$

Classification and Regression Tree (cont.)

Gini Diversity Index:

$$Gini(D) = 1 - \sum_{i=1}^{m} p_i^2$$

$$Gini_V(D) = \sum_{i=1}^{m} \frac{|D_i|}{|D|} Gini(D_i)$$

$$\Delta Gini(V) = Gini(D) - Gini_V(D)$$

Weather Condition (Sunny, Cloudy, Rainy)	Temperature (Warm, Cool)	Atmospheric Pressure (High, Low)
Sunny	Warm	High
Sunny	Warm	High
Sunny	Cool	Low
Cloudy	Cool	High
Cloudy	Warm	Low
Sunny	Warm	High
Cloudy	Warm	Low
Rainy	Cool	Low
Rainy	Cool	High
Cloudy	Warm	High

where

m **: number of possible classes/real values of the target variable (e.g., 3 weather conditions,** *m* **= 3)**

p_i **: percentage of tuples/observations that belongs to the class/real value** ***i***

$|D_i|$ **: number of tuples/observations that belongs to the class/real value** ***i*** **in the node**

V **: an independent variable as a predictor**

Note: Gini index increases both when the number of types increases and when evenness increase (similar to entropy).

Classification and Regression Tree (cont.)

The split criterion is to choose a variable *V* and a nominal value (categorical variable) or a split point (continuous variable) that maximize $\Delta Gini(V)$, i.e., to find the minimum $Gini_V(D)$.

$$Gini(D) = 1 - \sum_{i=1}^{m} p_i^2$$

$$Gini_V(D) = \sum_{i=1}^{m} \frac{|D_i|}{|D|} Gini(D_i)$$

$$\Delta Gini(V) = Gini(D) - Gini_V(D)$$

Take the weather data as example

Weather Condition	Tempe-rature	Atmospheric Pressure
Sunny	Warm	High
Sunny	Warm	High
Sunny	Cool	Low
Cloudy	Cool	High
Cloudy	Warm	Low
Sunny	Warm	High
Cloudy	Warm	Low
Rainy	Cool	Low
Rainy	Cool	High
Cloudy	Warm	High

$$Gini(Weather) = 1 - \left(\frac{4}{10}\right)^2 - \left(\frac{4}{10}\right)^2 - \left(\frac{2}{10}\right)^2 = 0.64$$

$$Gini_{Temperature}(Weather) = \frac{4}{10}\left(1 - \left(\frac{3}{4}\right)^2 - \left(\frac{1}{4}\right)^2\right) + \frac{4}{10}\left(1 - \left(\frac{3}{4}\right)^2 - \left(\frac{1}{4}\right)^2\right) + \frac{2}{10}\left(1 - \left(\frac{2}{2}\right)^2 - \left(\frac{0}{2}\right)^2\right)$$

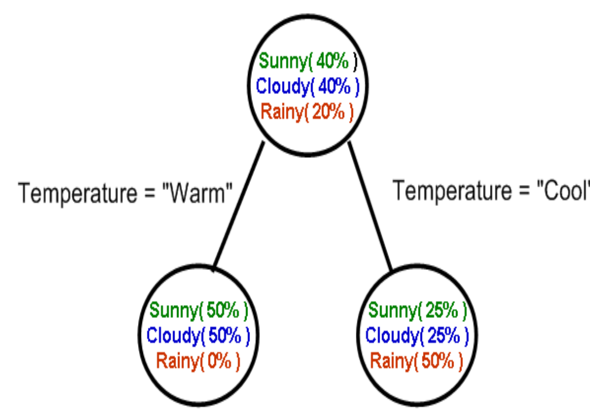
$$= 0.15 + 0.15 + 0 = 0.3$$

$$\Delta Gini(Temperature) = 0.64 - 0.3 = 0.34$$

$$Gini_{Atmospheric\ Pressure}(Weather) = \frac{4}{10}\left(1 - \left(\frac{3}{4}\right)^2 - \left(\frac{1}{4}\right)^2\right) + \frac{4}{10}\left(1 - \left(\frac{2}{4}\right)^2 - \left(\frac{2}{4}\right)^2\right) + \frac{2}{10}\left(1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)\right.$$

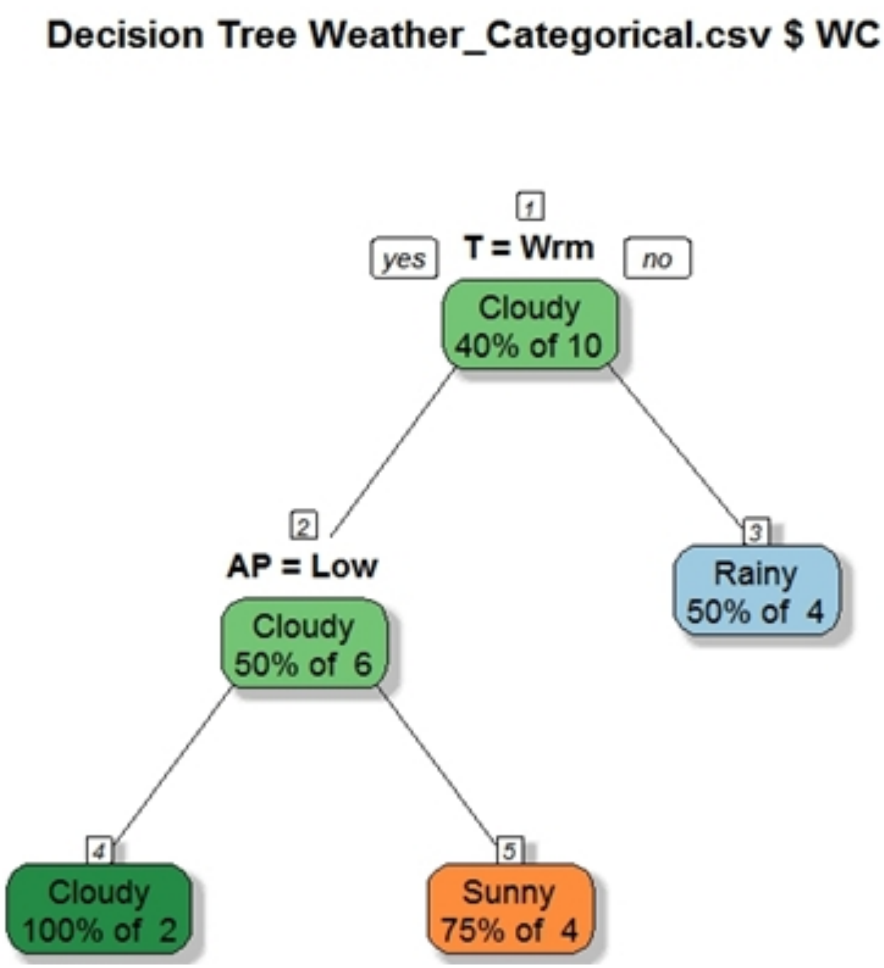
$$= 0.15 + 0.2 + 0.1 = 0.5$$

$$\Delta Gini(Atmospheric\ Pressure) = 0.64 - 0.5 = 0.14$$



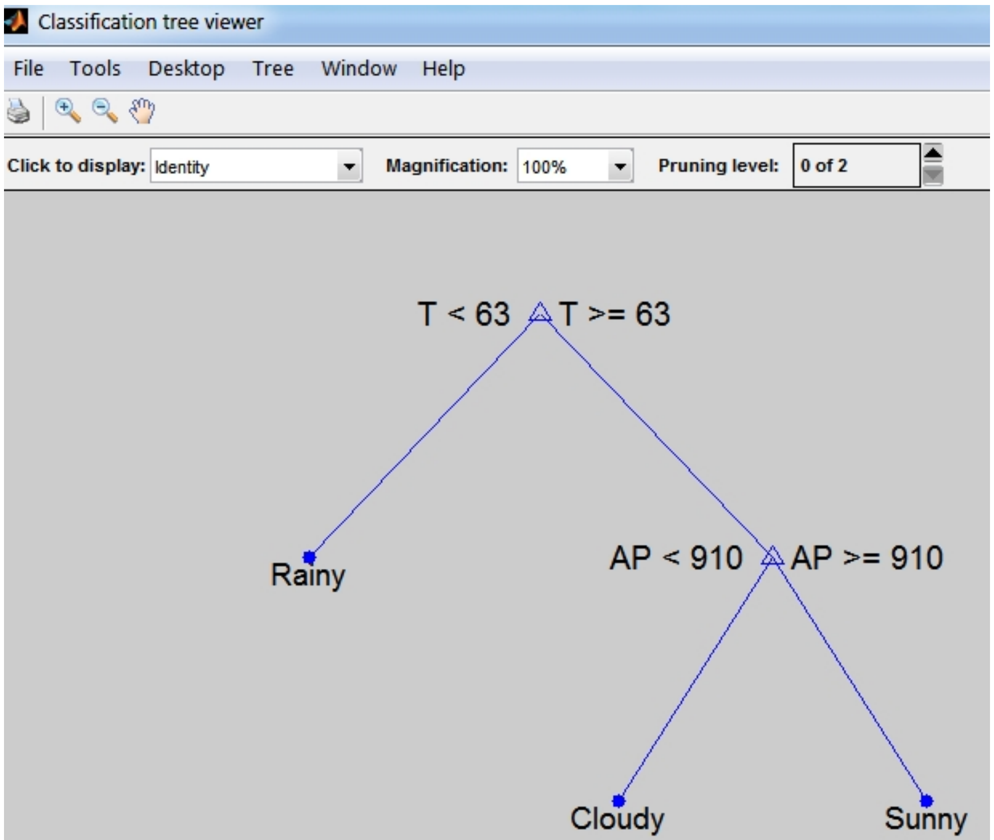
Therefore, we choose *Temperature*, not the *Atmospheric Pressure* as the first split to create the tree.

Classification Tree using R/Rattle with Weather data



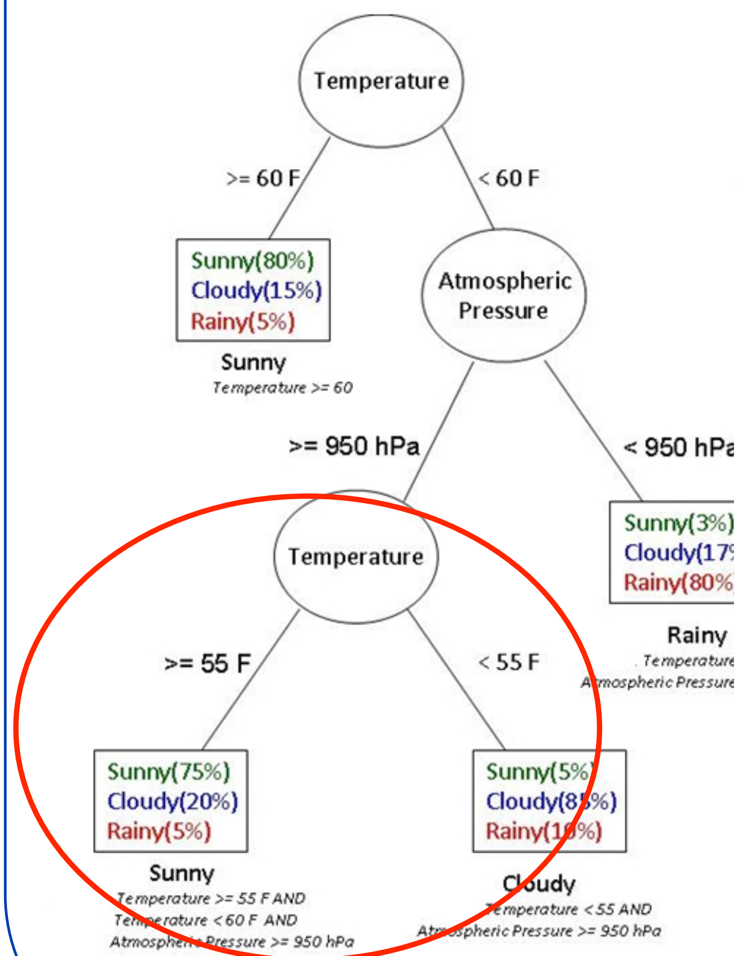
Weather Condition (Sunny, Cloudy, Rainy)	**Temperature (Warm, Cool)**	**Atmospheric Pressure (High, Low)**
Sunny	Warm	High
Sunny	Warm	High
Sunny	Cool	Low
Cloudy	Cool	High
Cloudy	Warm	Low
Sunny	Warm	High
Cloudy	Warm	Low
Rainy	Cool	Low
Rainy	Cool	High
Cloudy	Warm	High

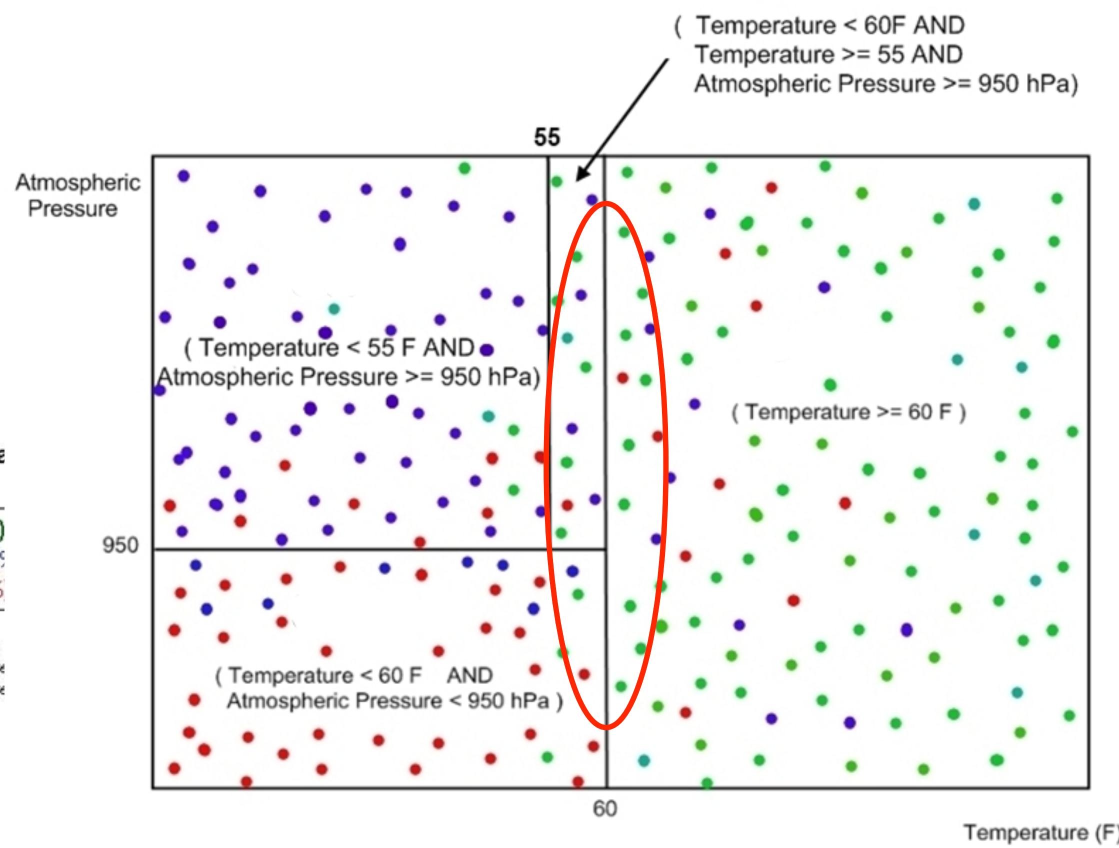
Classification Tree using MATLAB



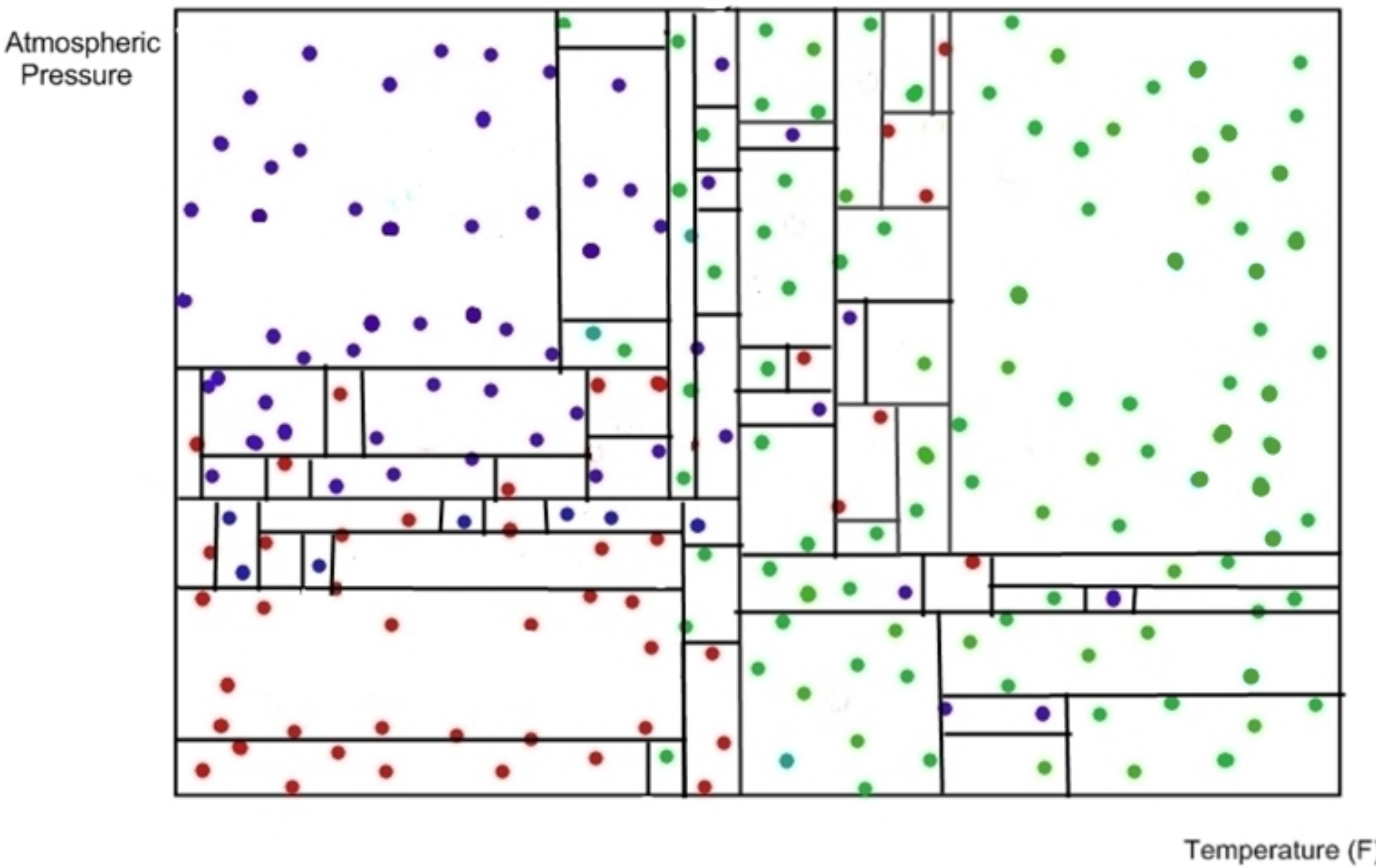
WC	T	AP
'Sunny'	65	980
'Sunny'	70	990
'Sunny'	55	800
'Cloudy'	50	960
'Cloudy'	72	850
'Sunny'	69	950
'Cloudy'	75	800
'Rainy'	60	840
'Rainy'	61	930
'Cloudy'	70	970
'Sunny'	80	950
'Sunny'	81	930
'Sunny'	80	940
'Rainy'	60	920
'Rainy'	78	890
'Rainy'	60	810

Tree splitting





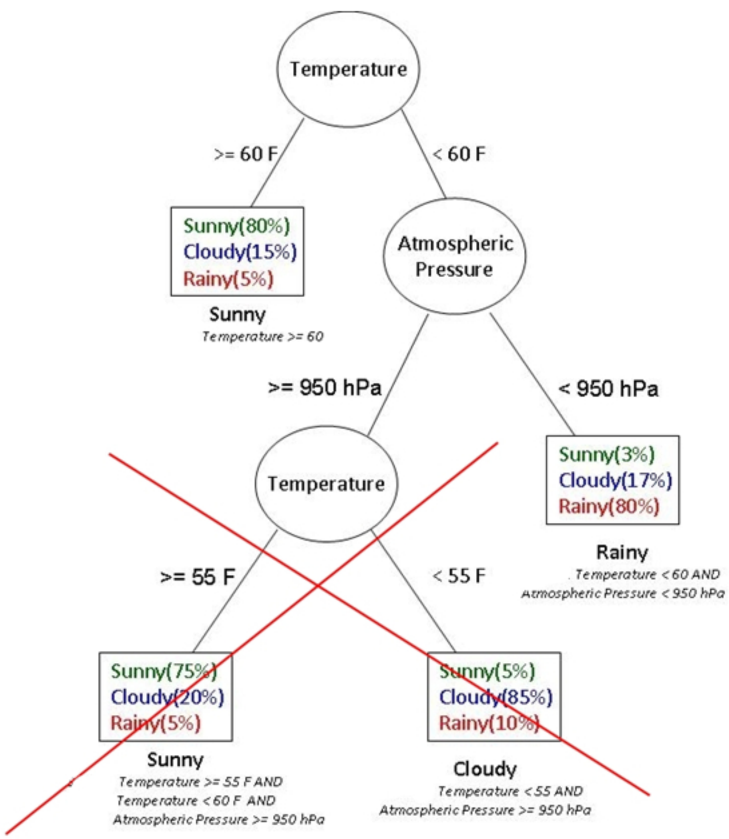
Tree “over”- splitting

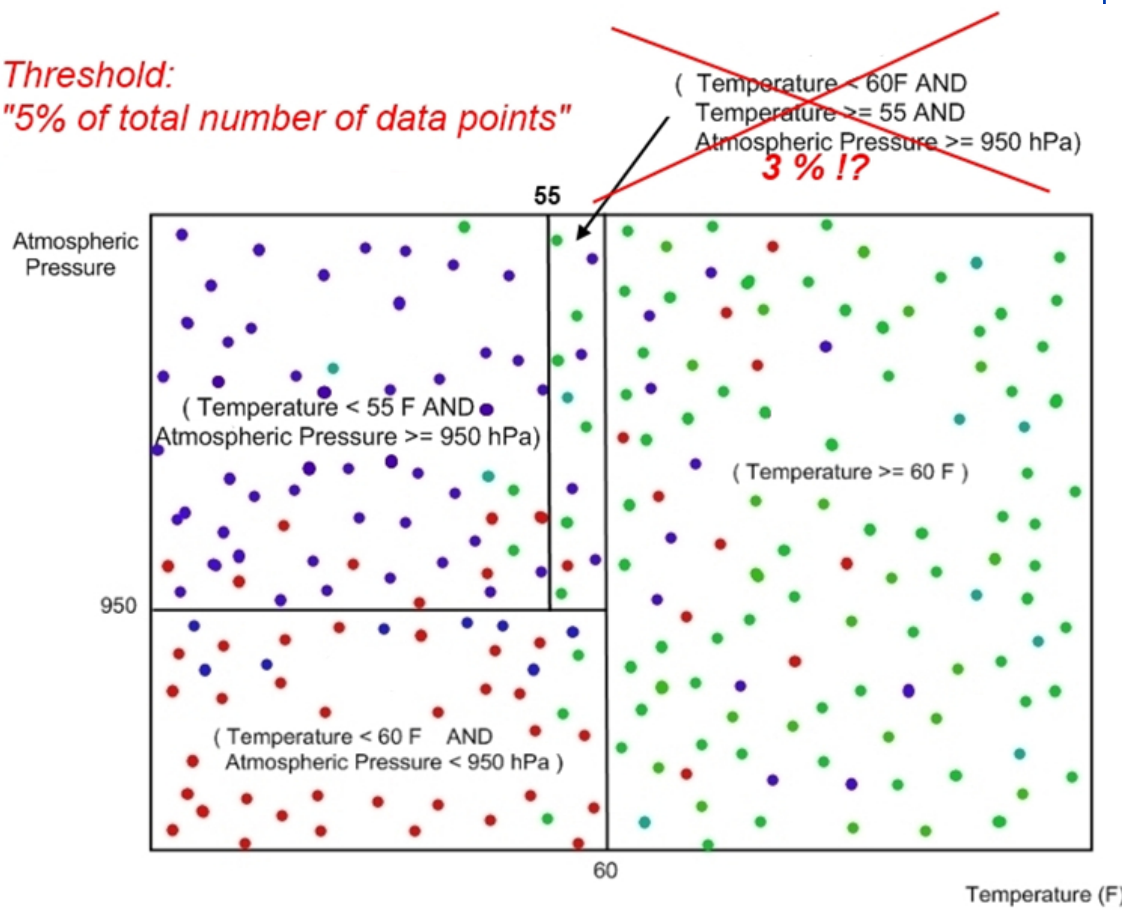


Tree pruning

Pre-pruning by halting the tree construction early

Threshold:
"5% of total number of data points"





Tree pruning (cont.)

$$Cost_A = \frac{\frac{(3+20)}{100} - \frac{(3+2+8+5+1)}{100}}{5 - 2} = 0.0166$$

$$Cost_B = \frac{\frac{(3+12+5+1)}{100} - \frac{(3+2+8+5+1)}{100}}{5 - 4} = 0.02$$

$$Cost_C = \frac{\frac{(3+2+8+8)}{100} - \frac{(3+2+8+5+1)}{100}}{5 - 4} = 0.02$$

Post-pruning **by pruning subtrees from a fully-grown tree using *cost-complexity pruning algorithm*.**

$$Cost = \frac{MR_{pruned} - MR_{orig}}{|leaves_{orig}| - |leaves_{pruned}|}$$

where
MR : error/misclassified rate
|***leaves***| : number of leaf nodes of a tree

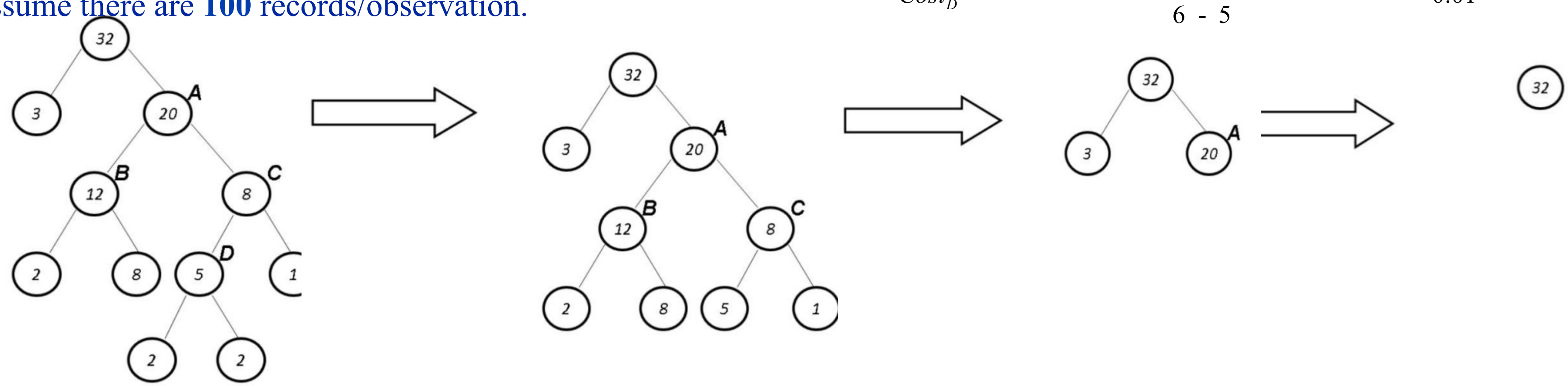
Each node contains the number of misclassified observations.
Assume there are **100** records/observation.

$$Cost_A = \frac{\frac{(3+20)}{100} - \frac{(3+2+8+2+2+1)}{100}}{6 - 2} = 0.0125$$

$$Cost_B = \frac{\frac{(3+12+2+2+1)}{100} - \frac{(3+2+8+2+2+1)}{100}}{6 - 5} = 0.02$$

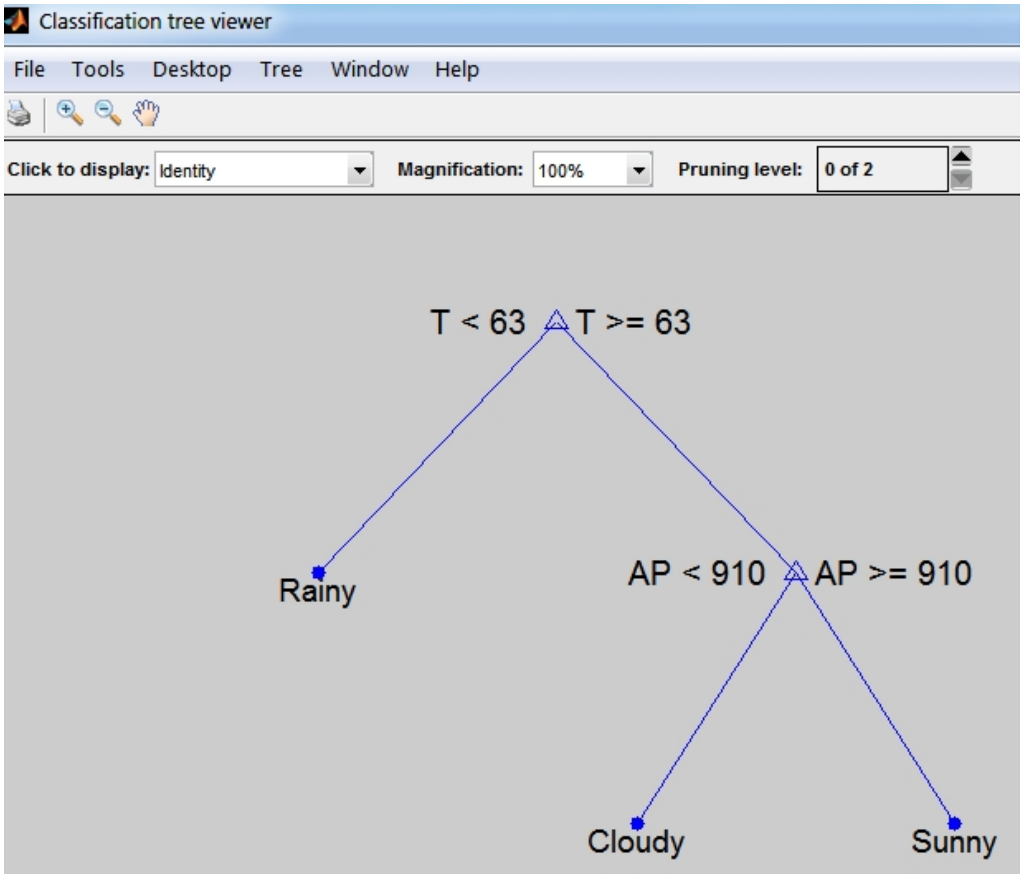
$$Cost_C = \frac{\frac{(3+2+8+8)}{100} - \frac{(3+2+8+2+2+1)}{100}}{6 - 4} = 0.015$$

$$Cost_D = \frac{\frac{(3+2+8+5+1)}{100} - \frac{(3+2+8+2+2+1)}{100}}{6 - 5} = 0.01$$



Tree 1 Tree 2 Tree 3 Tree 4

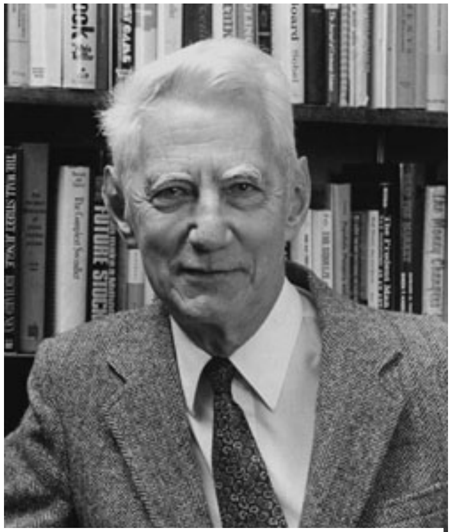
Tree Pruning using MATLAB with Weather data



WC	T	AP
'Sunny'	65	980
'Sunny'	70	990
'Sunny'	55	800
'Cloudy'	50	960
'Cloudy'	72	850
'Sunny'	69	950
'Cloudy'	75	800
'Rainy'	60	840
'Rainy'	61	930
'Cloudy'	70	970
'Sunny'	80	950
'Sunny'	81	930
'Sunny'	80	940
'Rainy'	60	920
'Rainy'	78	890
'Rainy'	60	810

Iterative Dichotomiser (ID3) and C4.5 Tree Learning

Entropy

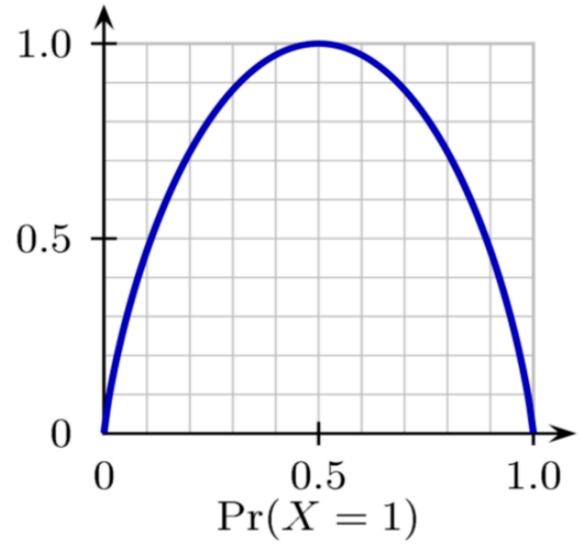


- Entropy: $H(X) = - \sum p(X = i)\log_2 p(X = i)$
- Introduced by Claude Shannon (1948), "A Mathematical Theory of Communication"
- Entropy measures the uncertainty in a specific distribution.

- If $p(X = 1) = 1$ then:
  $H(X) = -(1)(\log_2 1) - (0)(\log_2 0) = 0$

- If $p(X = 1) = .5$ then:
  $H(X) = -(.5)(\log_2 .5) - (.5)(\log_2 .5) = 1$

- Which attribute to split on?
  - The one that is best to reduce uncertainty (Entropy)
  - In the other word, the one that gives the most information gain



Iterative Dichotomiser (ID3) and C4.5 Tree Learning

ID3

$$Info(D) = -\sum_{i=1}^{m} p_i \log_2(p_i)$$

$$Info_V(D) = \sum_{i=1}^{m} \frac{|D_i|}{|D|} \times Info(D_i)$$

$$Gain(V) = Info(D) - Info_V(D)$$

C4.5

$$SplitInfo_V(D) = -\sum_{i=1}^{m} \frac{|D_i|}{|D|} \times \log_2\left(\frac{|D_i|}{|D|}\right)$$

$$GainRatio(V) = \frac{Gain(V)}{SplitInfo_V(D)}$$

m : the number of possible classes/real values of the target variable (e.g. 3 weather conditions, m = 3)

p_i : the percentage of tuples/observations that belongs to the class/real value *i*.

$|D_i|$: the number of tuples/observations that belongs to the class/real value *i* in the node

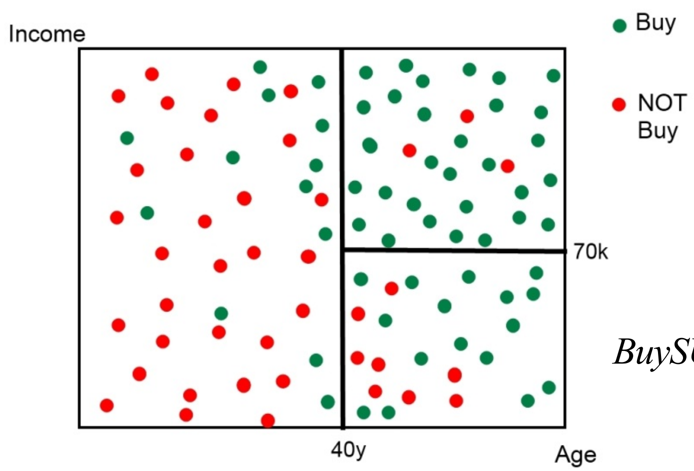
V : a independent variable as a predictor

Differences between CART & ID3 (C4.5)

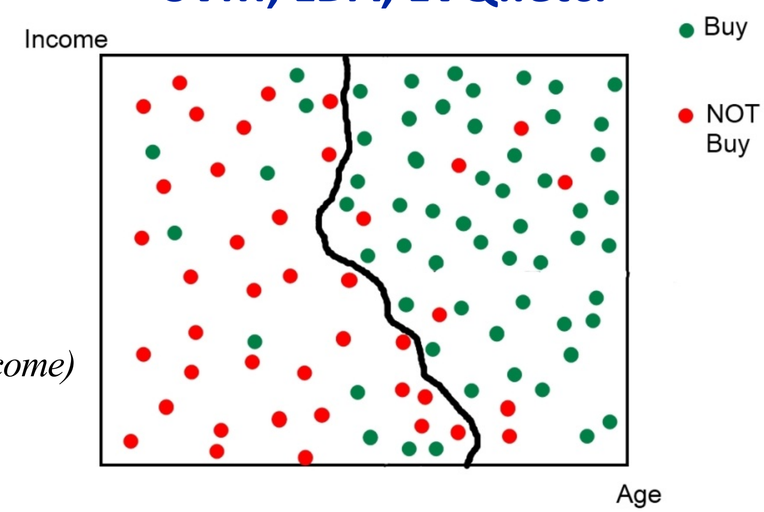
- Gini Diversity Index vs. Binary Shannon Entropy
- Binary trees vs. N-ary trees
- Different tree pruning algorithms

Why bother with "the Rules"?

Classification Tree



SVM, LDA, LVQ..etc.



BuySUV = f(Age, Income)

Misclassification Rate = 15% ??

Misclassification Rate = 10% ?? (usually lower)

Rules:

1: IF Age>= 40y AND Income >= 70k
   THEN Buy = True
2: IF Age>= 40y AND Income < 70k
   THEN Buy = True
3: IF Age < 40y
   THEN Buy = False

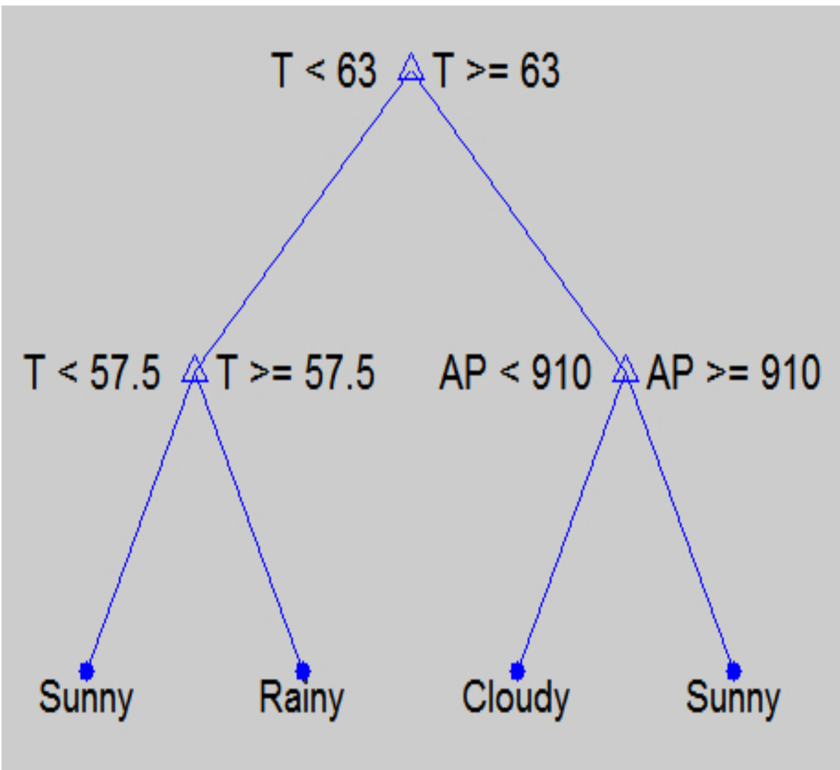
Simple!!

Rules: ??

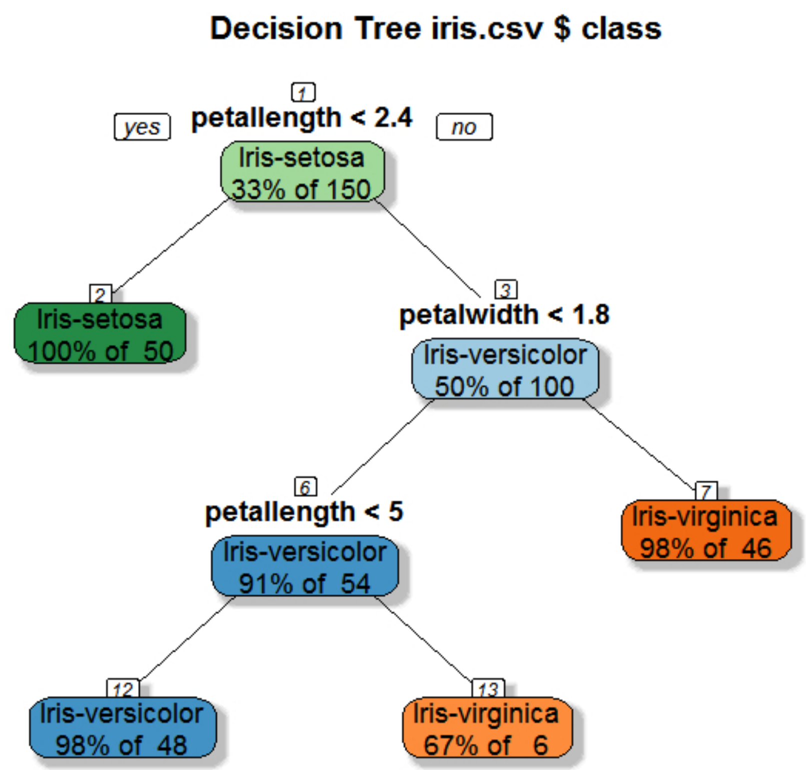
Obscure!!
Rule extraction?!

Rule Extraction from Tree Model using MATLAB, Rattle and Weka

Weather data



IRIS data



Tree Pruning and Model Evaluation

Tree model evaluation

Confusion Matrix and Misclassification Rate (*MR*)

		Actual Class	
		C	$\overline{C}$
Predicted Class	C	True Positive (**TP**)	False Positive (**FP**)
	$\overline{C}$	False Negative (**FN**)	True Negative (**TN**)

Confusion Matrix

$$T = TP + FP + FN + TN$$

$$Accuracy = \frac{TP + TN}{T}$$

$$MR = 1 - Accuracy = 1 - \frac{TP + TN}{T}$$

$K = 0$

		Predicted Class		
		A	***B***	
Actual Class	***A***	25	25	50
	B	25	25	50
		50	50	100

Matrix A

$$MR = 1 - \frac{25 + 25}{100} = 0.5$$

$K = 0.1873$

		Predicted Class		
		A	***B***	
Actual Class	***A***	25	1	26
	B	49	25	74
		74	26	100

Matrix B

$$MR = 1 - \frac{25 + 25}{100} = 0.5$$

Cohen's Kappa Statistics (*K*)

***K* takes into account the chance agreement**

(correctly classified by chance)

$$K = \frac{P_o - P_e}{1 - P_e}$$

$$P_o = \text{observed accuracy} = 1 - MR = \frac{TP + TN}{T}$$

$$P_e = \text{expected (chance) accuracy}$$

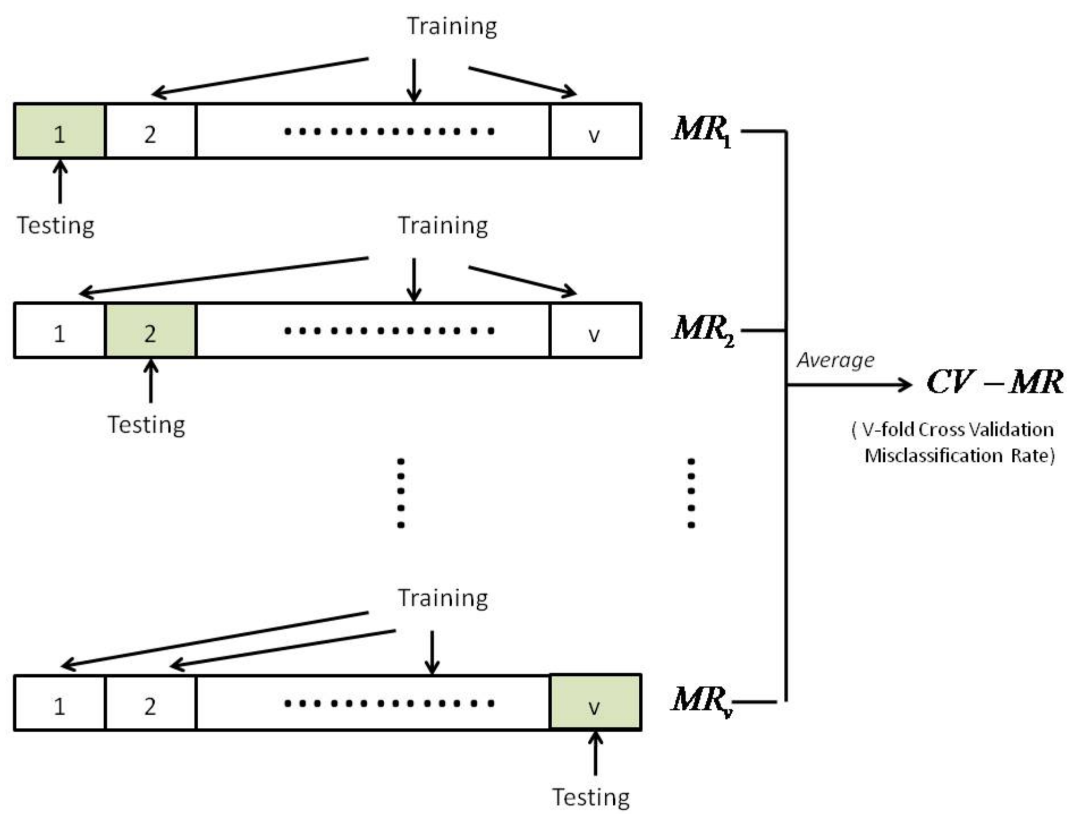
$$= \frac{\frac{(TP + FN) * (TP + FP)}{T} + \frac{(FP + TN) * (FN + TN)}{T}}{T}$$

K	Interpretation
< 0	**Poor agreement**
0.0 — 0.20	**Slight agreement**
0.21 — 0.40	**Fair agreement**
0.41 — 0.60	**Moderate agreement**
0.61 — 0.80	**Substantial agreement**
0.81 — 1.00	**Almost perfect agreement**

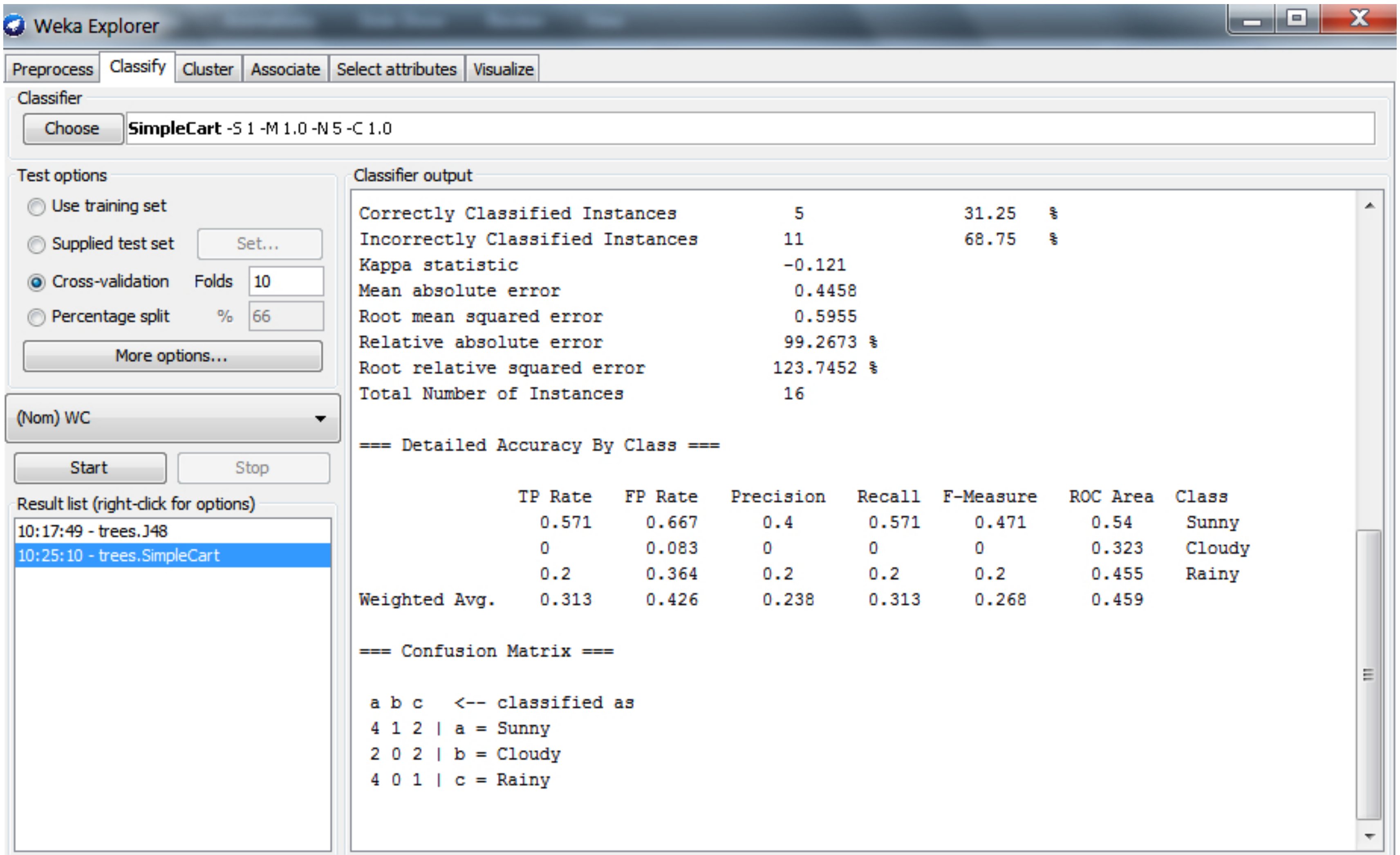
Tree model evaluation (cont.)

V-fold Cross Validation MR (CV-MR)

A fair MR to measure the performance of a tree applied to both training and testing dataset (other independent datasets).



Tree model evaluation using Weka



Association Rule Mining

Association Rule Mining

Discover interesting association rules between item(set)s in large databases

- The ***Support*** of an itemset ***X*** is the percentage of transactions in the dataset ***T*** which contains ***X***

$$\mathrm{supp}(X) = \frac{|X|}{|T|}$$

e.g. Supp({milk, bread}) = 2/5 = 0.4

Transaction ID	Item1	Item2	Item3
1	**milk**	**bread**	
2	**butter**		
3	**beer**		
4	**milk**	**bread**	**butter**
5	**bread**		

- The ***Confidence*** of a rule (***X*** => ***Y***) is the percentage that itemset ***X*** and ***Y*** both appear together. e.g. Conf({milk, bread} => {butter}) = 0.2/0.4 = 0.5

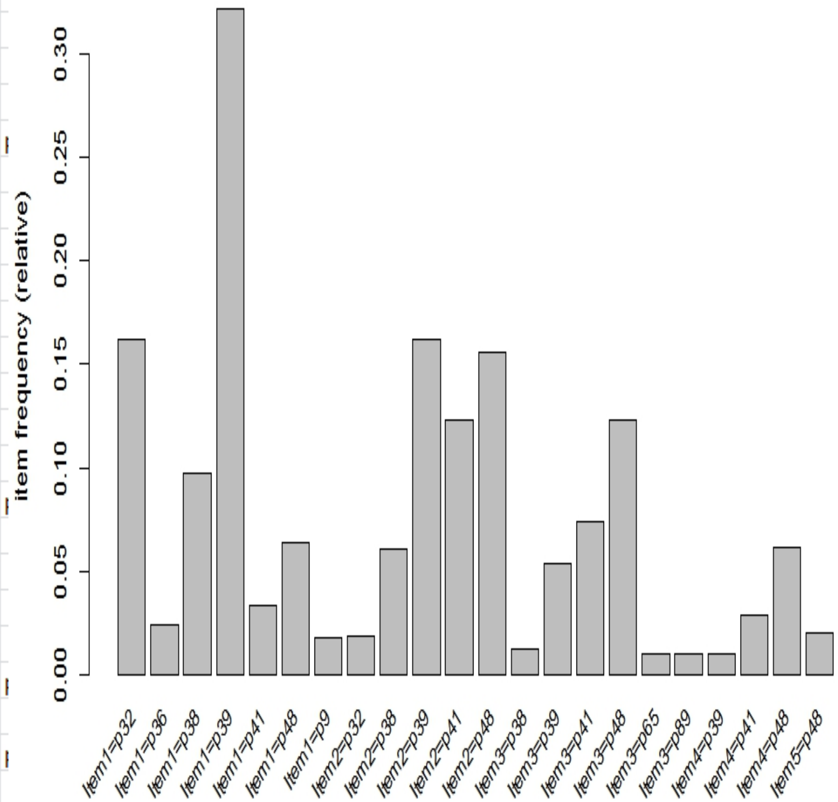
$$\mathrm{conf}(X \Rightarrow Y) = \mathrm{P}(Y \mid X) = \frac{\mathrm{supp}(X \cup Y)}{\mathrm{supp}(X)}$$

- The ***Lift*** of a rule (***X*** => ***Y***) is the ratio of the observed support to the expected support. It is to measure the degree of independence between itemset ***X*** and ***Y*** e.g. Lift({milk, bread} => {butter}) = 0.2/ (0.4 * 0.4) = 1.25

$$\mathrm{lift}(X \Rightarrow Y) = \frac{\mathrm{supp}(X \cup Y)}{\mathrm{supp}(X) \times \mathrm{supp}(Y)}$$

Association Rule Mining using R/Rattle with Retail data

Item1	Item2	Item3	Item4	Item5	Item6	Item7	Item8	Item9	Item10	Item11
p30	p31	p32								
p33	p34	p35								
p36	p37	p38	p39	p40	p41	p42	p43	p44	p45	p46
p38	p39	p47	p48							
p38	p39	p48	p49	p50	p51	p52	p53	p54	p55	p56
p32	p41	p59	p60	p61	p62					
p3	p39	p48								
p63	p64	p65	p66	p67	p68					
p32	p69									
p48	p70	p71	p72							
p39	p73	p74	p75	p76	p77	p78	p79			
p36	p38	p39	p41	p48	p79	p80	p81			
p82	p83	p84								
p41	p85	p86	p87	p88						
p39	p48	p89	p90	p91	p92	p93	p94	p95	p96	p97
p36	p38	p39	p48	p89						
p39	p41	p102	p103	p104	p105	p106	p107	p108		
p38	p39	p41	p109	p110						
p39	p111	p112	p113	p114	p115	p116	p117	p118		
p119	p120	p121	p122	p123	p124	p125	p126	p127	p128	p129
p48	p134	p135	p136							
p39	p48	p137	p138	p139	p140	p141	p142	p143	p144	p145
p39	p150	p151	p152							
p38	p39	p56	p153	p154	p155					



Retail market basket data from anonymous Belgian retail store, http://fimi.ua.ac.be/data/

References

[1] "Predictive modeling," *Wikipedia, the free encyclopedia*. 21-Sep-2012.

[2] "Inductive logic programming," *Wikipedia, the free encyclopedia*. 13-Oct-2012.

[3] T. K. Prasad, "CS 774 Logic Programming." [Online]. Available: http://www.cs.wright.edu/~tkprasad/courses/cs774/cs774.html.

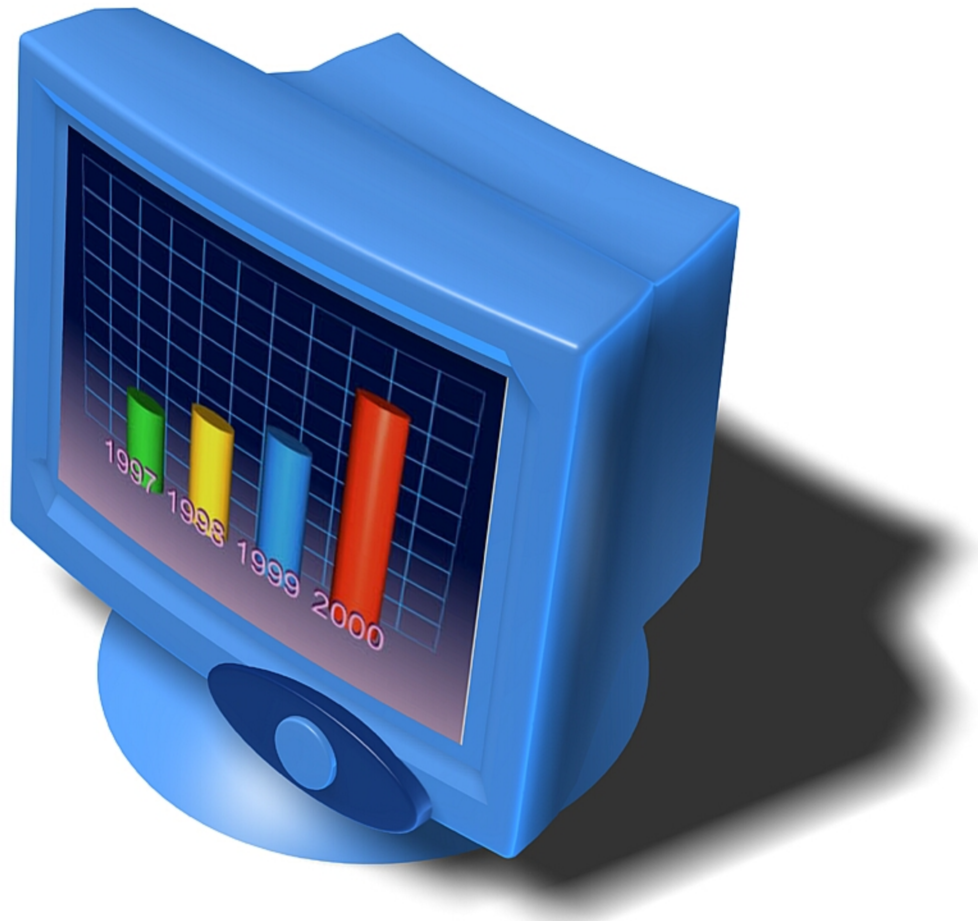
[4] J. Han and M. Kamber, *Data mining: concepts and techniques*. Morgan Kaufmann, 2006.

[5] I. H. Witten and E. Frank, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2005.

[6] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, vol. 1. Chapman & Hall/CRC, 1984.

[7] "Association rule learning," *Wikipedia, the free encyclopedia*. 21-Oct-2012.

Examples: Weka



Concluding remarks

- Conclusions