Summary

Minimax

- Opponent acts rationally. Outcome only depends upon skills of minimizing max's advantage

Expectimax

- Opponent acts randomly (probabilistically). Outcome only depends on chance for opponent

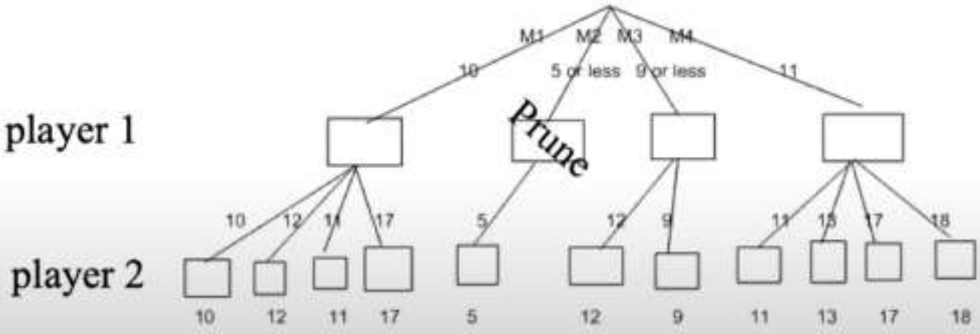
Expectiminimax

- Opponent acts rationally as well as randomness is also present. So opponent behaves as minimizing agent, however choice of moves is also done based on some random process (skills + chance)

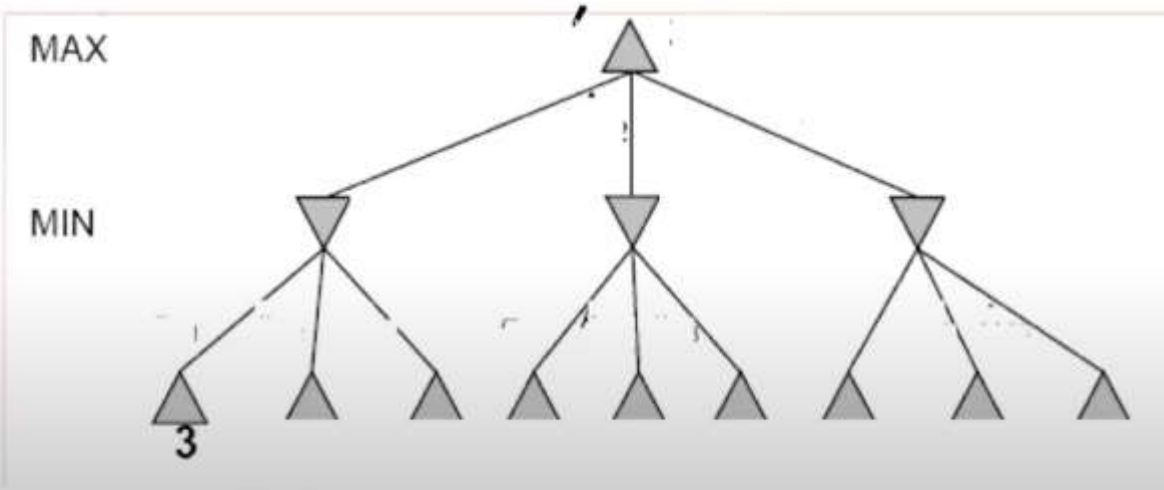
Pruning – MINIMAX with ALPHA BETA

- Want to visit as many board states as possible
  - Want to avoid whole branches (prune them)
    - Because they can't possibly lead to a good score
  - Example: having your queen taken in chess
    - (Queen sacrifices often very good tactic, though)
- Alpha-beta pruning
  - Can be used for entire search or cutoff search
  - Recognize that a branch cannot produce better score
    - Than a node you have already evaluated

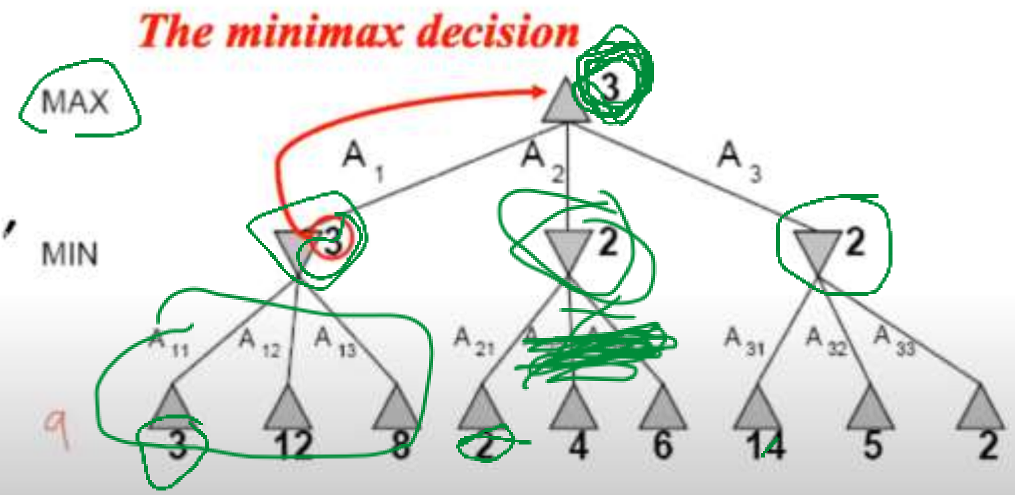
Example of Alpha-Beta Pruning



- Depth first search a good idea here

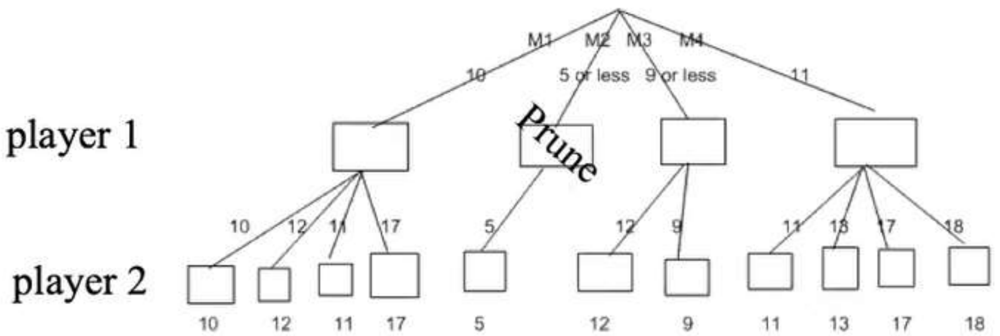


Two-Ply Game Tree



Minimax maximizes the worst-case outcome for max.

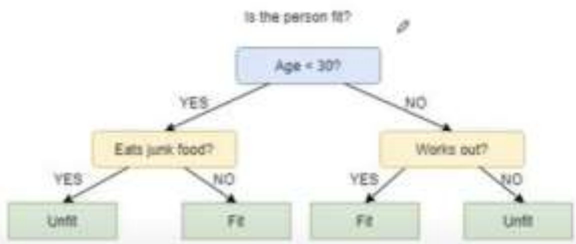
Example of Alpha-Beta Pruning



- Depth first search a good idea here

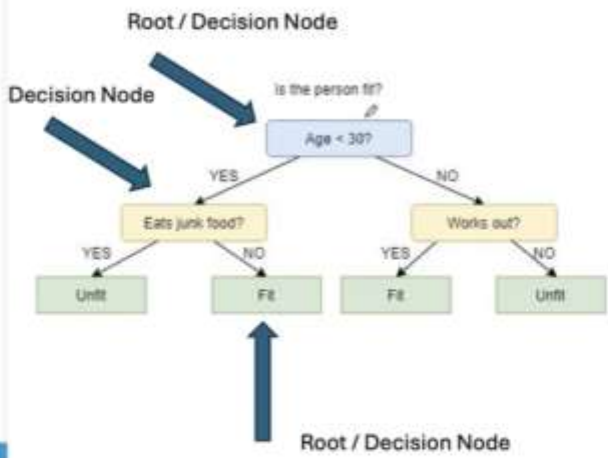
Decision Trees

- Conversion of data to TREE format
- Each Branch of a Tree represent a RULE
- Problem:
  - Which variable to choose to decide on the split
- Characteristics to look for:
  - Depth and Breadth of the Tree
  - Accuracy
- Concept:
  - Choose variable that reduces NOISE/ENTROPY ie., indecision



Decision Trees

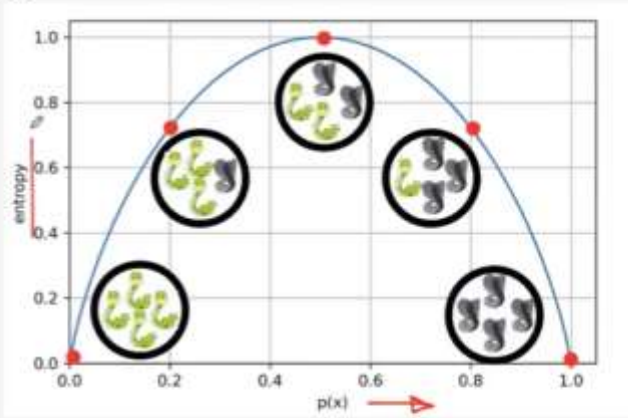
- Conversion of data to TREE format
- Each Branch of a Tree represent a RULE
- Problem:
  - Which variable to choose to decide on the split
- Characteristics to look for:
  - Depth and Breadth of the Tree
  - Accuracy
- Concept:
  - Choose variable that reduces NOISE/ENTROPY ie., indecision



How to decide Root/Subsequent Decision Nodes ?

Entropy Based Solution / Gini Index

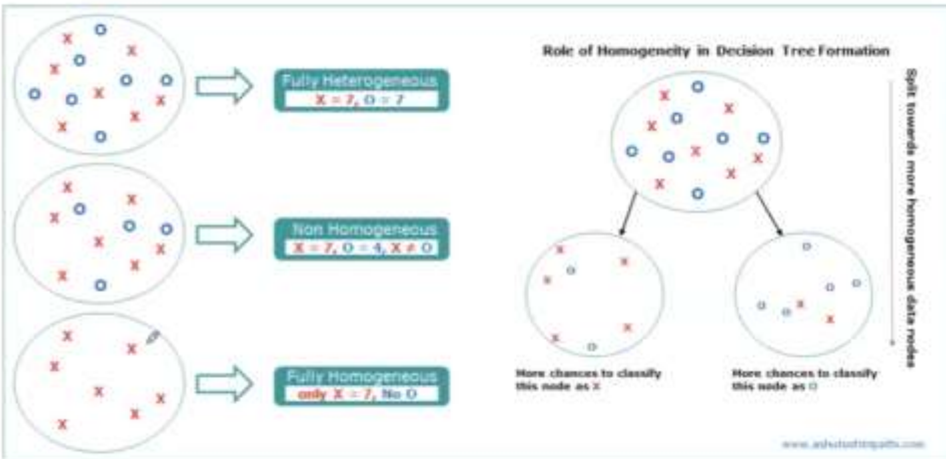
Choose a variable
that reduces entropy



Data Homogenity

$$H_x = -\sum_{i=1}^{c} p_i log_2 p_i$$

i denotes number of classes



Data

Outlook	Temperature	Humidity	Wind	Played football(yes/no)
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

Play Football

Yes	No
9	5

Entropy(Play Football) = ?

P(Yes) = count where play is yes/ total count of rows = 9/14

P(No) = 5/14

$$H_{Play} = -log_2 P(Yes) - log_2 P(No)$$

$$H_{Play} = -\frac{9}{14} log_2 \frac{9}{14} - \frac{5}{14} log_2 \frac{5}{14}$$

$$H_{Play} = 0.94$$

Outlook

$$H_{Play, outlook} = P(rainy)H(play, given\ rainy) + P(overcast)H(play, given\ overcast) + P(sunny)H(play, given\ sunny)$$

*here given means consider only those rows when calculating entropy

Outlook	Yes	No	Total
Rainy	3	2	5
Overcast	4	0	4
Sunny	2	3	5
			14

$$H_{Play, outlook} = \left(\frac{5}{14}\right)\left[-\frac{3}{5} log_2 \frac{3}{5} - \frac{2}{5} log_2 \frac{2}{5}\right] + \left(\frac{4}{14}\right)\left[-\frac{4}{4} log_2 \frac{4}{4} - \frac{0}{4} log_2 \frac{0}{4}\right] + \left(\frac{5}{14}\right)\left[-\frac{2}{5} log_2 \frac{2}{5} - \frac{3}{5} log_2 \frac{3}{5}\right]$$

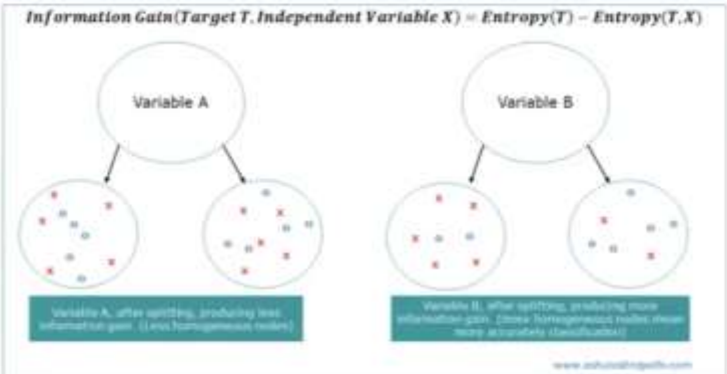
Entropy of homogeneous data (only one class yes) is coming as 0 and it is proved as we discussed in above lines.

$$H_{Play, outlook} = \left(\frac{5}{14}\right)[0.971] + \left(\frac{4}{14}\right) 0 + \left(\frac{5}{14}\right)[0.971] = 0.693$$

Information Gain

$$\boxed{Information\,Gain(Target\,T, Independent\,Variable\,X) = Entropy(T) - Entropy(T,X)}$$

$$\boldsymbol{Information\,Gain(Play, Outlook) = Entropy(Play) - Entropy(Play, Outlook)}$$
$$\boldsymbol{= 0.940 - 0.693 = 0.247}$$



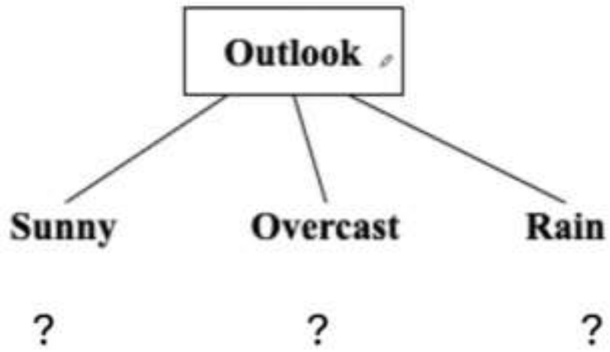
Information Gain - IG

- Repeat the steps for calculating IG for all other variables
- Choose the one with highest IG

- **IG(S, outlook) = 0.94 - 0.693 = 0.247**
- IG(S, Temperature) = 0.940 - 0.911 = 0.029
- IG(S, Humidity) = 0.940 - 0.788 = 0.152
- IG(S, Windy) = 0.940 - 0.8932 = 0.048

So outlook is taken as Root Node

Next Steps ...



Repeat the Above process for each sub table

Outlook	Temperature	Humidity	Wind	Played football(yes/no)
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

Outlook	Temperature	Humidity	Wind	Played football(yes/no)
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes

Outlook	Temperature	Humidity	Wind	Played football(yes/no)
Overcast	Hot	High	Weak	Yes
Overcast	Cool	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes

Outlook	Temperature	Humidity	Wind	Played football(yes/no)
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Rain	Mild	Normal	Weak	Yes
Rain	Mild	High	Strong	No

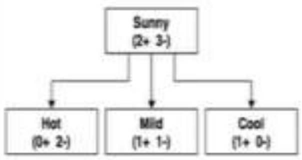
Outlook	Temperature	Humidity	Wind	Played football(yes/no)
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes

- Find IG for Temperature given Sunny
- Repeat it for humidity and wind
- Choose one with highest IG.

Gain(outlook = sunny|Temp.) = 0.570
Gain(outlook = sunny|Humidity) = 0.970
Gain(outlook = sunny|Wind) = 0.019

$$Gain(outlook = sunny|Temp.) = H(PlayTennis|outlook = sunny) - \sum_{v \in values(A)} \frac{|S_v|}{|S|} H(S_v)$$

Where, v ∈ {hot, mild, cold} and S -> (PlayTennis|outlook=sunny)



$$H(hot) = -\left[\frac{0}{2}\lg\frac{0}{2} + \frac{2}{2}\lg\frac{2}{2}\right] = 0$$

$$H(mild) = -\left[\frac{1}{2}\lg\frac{1}{2} + \frac{1}{2}\lg\frac{1}{2}\right] = 0$$

$$H(cool) = -\left[\frac{1}{1}\lg\frac{1}{1} + \frac{0}{1}\lg\frac{0}{1}\right] = 0$$

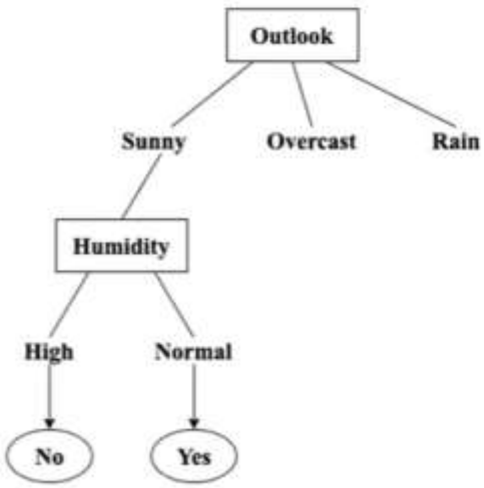
$$Gain(outlook = sunny|Temp.) = 0.970 - [\tfrac{2}{5} * 0 + \tfrac{2}{5} * 1 + \tfrac{1}{5} * 0]$$

$$= 0.970 - \frac{2}{5}$$

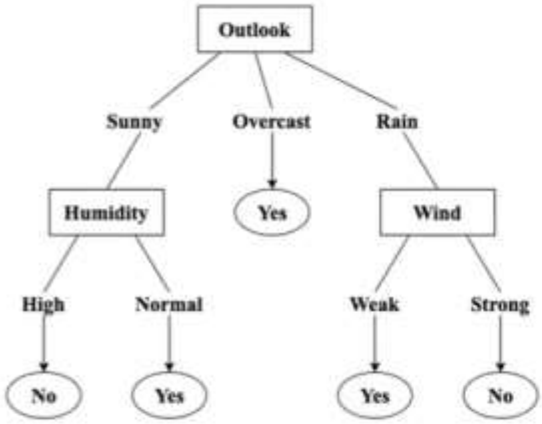
$$= 0.570$$

DT under Outlook=Sunny

- Similarily find important variable under Overcast and Rain



Final Decision Tree



Decision Tree for Regression Output: Numeric Value

Variance = $\sum_{i=1}^{n}(x_i - \mu)^2/n$

Standard Deviation = $\sqrt{\sum_{i=1}^{n}(x_i - \mu)^2/n}$, μ *is mean*

Mean = $\sum_{i=1}^{n} x_i / \sum_{i=1}^{n} f_i$ (Sum of all scores / sum of frequencies)

Outlook	Temperature	Humidity	Windy	Hours Played
Sunny	Hot	High	FALSE	25
Sunny	Hot	High	TRUE	30
Overcast	Hot	High	FALSE	46
Rainy	Mild	High	FALSE	45
Rainy	Cool	Normal	FALSE	52
Rainy	Cool	Normal	TRUE	23
Overcast	Cool	Normal	TRUE	43
Sunny	Mild	High	FALSE	35
Sunny	Cool	Normal	FALSE	38
Rainy	Mild	Normal	FALSE	46
Sunny	Mild	Normal	TRUE	48
Overcast	Mild	High	TRUE	52
Rainy	Mild	High	TRUE	30

Output Variable

Hours Played
25
30
46
45
52
23
43
35
38
46
48
52
44
30

Count $= n = 14$

Average $= \bar{x} = \frac{\sum x}{n} = 39.8$

Standard Deviation $= S = \sqrt{\frac{\sum(x - \bar{x})^2}{n}} = 9.32$

Coeffeicient of Variation $= CV = \frac{S}{\bar{x}} * 100\% = 23\%$

Standard Deviation Reduction if we make Outlook as Root Node = 1.66

Outlook column values	Hours Played	Count
	Standard Deviation for respective outlook condition	
rainy	10.87	5
overcast	3.49	4
sunny	7.78	5

Add caption

$$Weighted\ SD_{Hours,Outlook} = P(rainy) \cdot SD_{Hours,Rainy} + P(Overcast) \cdot SD_{Hours,Overcast} + P(Sunny) \cdot SD_{Hours,Sunny}$$

$$SD{:}\ Standard\ Deviation$$

$$Weighted\ SD_{Hours,Outlook} = \left(\frac{5}{14}\right) \cdot 10.87 + \left(\frac{4}{14}\right) \cdot 3.49 + \left(\frac{5}{14}\right) \cdot 7.78 = 7.66$$

$$SDR = SD(Hours) - SD(Hours, Outlook)$$

SDR for All Input Variables

Root node

SDR (Hours, Outlook) = 1.66

SDR (Hours, Temperature) = 0.39

SDR (Hours, Humidity) = 0.09

SDR (Hours, Windy) = 0.39

K- Nearest Neigbour