FIND-S: Step-1

Example	*Sky*	*AirTemp*	*Humidity*	*Wind*	*Water*	*Forecast*	*EnjoySport*
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

1. Initialize h to the most specific hypothesis in H

vtupulse.com

h0 = <Ø, Ø, Ø, Ø, Ø, Ø>

Mahesh Huddar

FIND-S: Step-2

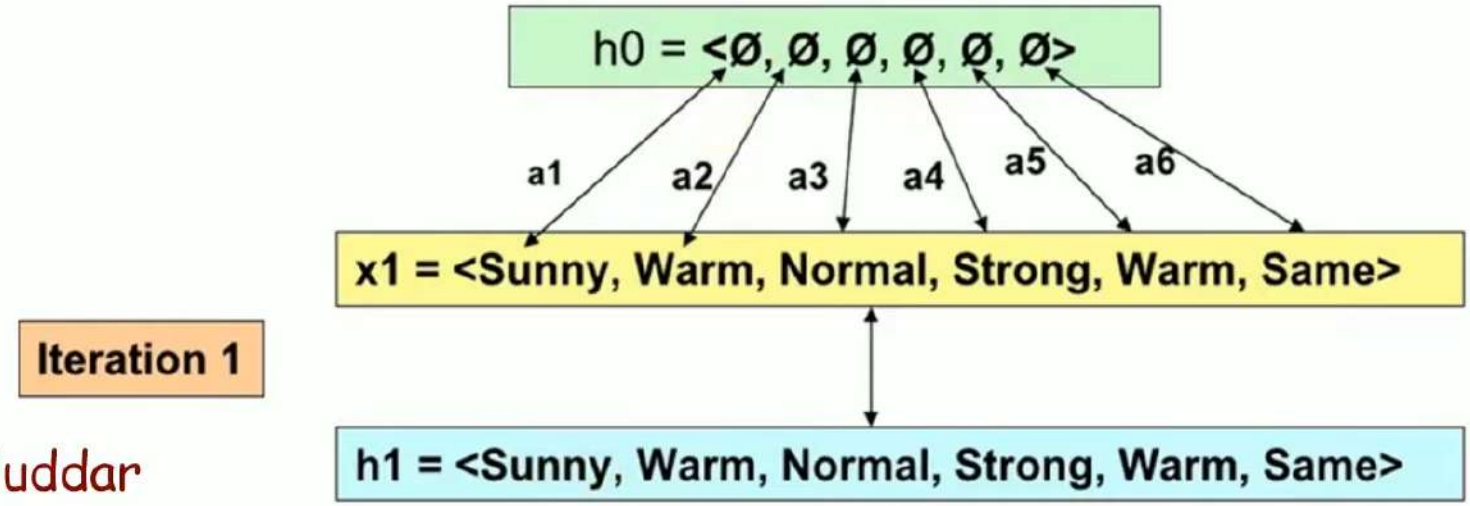
Example	*Sky*	*AirTemp*	*Humidity*	*Wind*	*Water*	*Forecast*	*EnjoySport*
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

2. For each positive training instance x
   - For each attribute constraint a_i in h

     If the constraint a_i is satisfied by x

     Then do nothing

     Else replace a_i in h by the next more general constraint that is satisfied by x



Iteration 1

vtupulse.com

Mahesh Huddar

FIND-S: Step-2

Example	*Sky*	*AirTemp*	*Humidity*	*Wind*	*Water*	*Forecast*	*EnjoySport*
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

Iteration 2

h1 = <Sunny, Warm, Normal, Strong, Warm, Same>

x2 = <Sunny, Warm, High, Strong, Warm, Same>

h2 = <Sunny, Warm, ?, Strong, Warm, Same>

vtupulse.com

Mahesh Huddar

FIND-S: Step-2

Example	*Sky*	*AirTemp*	*Humidity*	*Wind*	*Water*	*Forecast*	*EnjoySport*
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

Iteration 3

Ignore

h3 = <Sunny, Warm, ?, Strong, Warm, Same>

vtupulse.com

Mahesh Huddar

FIND-S: Step-2

Example	*Sky*	*AirTemp*	*Humidity*	*Wind*	*Water*	*Forecast*	*EnjoySport*
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

h3 = < Sunny, Warm, ?, Strong, Warm, Same >

Iteration 4

x4 = < Sunny, Warm, High, Strong, Cool, Change >

Step 3

Output

h4 = <Sunny, Warm, ?, Strong, ?, ?>

vtupulse.com

Mahesh Huddar

FIND-S Algorithm Solved Example - 2

example	*citations*	*size*	*inLibrary*	*price*	*editions*	*buy*
1	some	small	no	affordable	many	no
2	many	big	no	expensive	one	yes
3	some	big	always	expensive	few	no
4	many	medium	no	expensive	many	yes
5	many	small	no	affordable	many	yes

1. How many concepts are possible for this instance space?

2. How many hypotheses can be expressed by the hypothesis language?

3. Apply the FIND-S algorithm by hand on the given training set. Consider the examples in the specified order and write down your hypothesis each time after observing an example.

Mahesh Huddar

vtupulse.com

example	*citations*	*size*	*inLibrary*	*price*	*editions*	*buy*
1	some	small	no	affordable	many	no
2	many	big	no	expensive	one	yes
3	some	big	always	expensive	few	no
4	many	medium	no	expensive	many	yes
5	many	small	no	affordable	many	yes

1. How many concepts are possible for this instance space?

Solution: 2 * 3 * 2 * 2 * 3 = 72

2. How many hypotheses can be expressed by the hypothesis language?

Solution: 4 * 5 * 4 * 4 * 5 = 1600

Semantically Distinct Hypothesis = (3 * 4 * 3 * 3 * 4) + 1 = 433

Mahesh Huddar

vtupulse.com

example	*citations*	*size*	*inLibrary*	*price*	*editions*	*buy*
1	some	small	no	affordable	many	no
2	many	big	no	expensive	one	yes
3	some	big	always	expensive	few	no
4	many	medium	no	expensive	many	yes
5	many	small	no	affordable	many	yes

Step 1:

h0 = (ø, ø, ø, ø, ø)

Step 2:

X1 = (some, small, no, expensive, many) – No

Negative Example Hence Ignore

h1 = (ø, ø, ø, ø, ø)

X2 = (many, big, no, expensive, one) – Yes

h2 = (many, big, no, expensive, one)

Mahesh Huddar

vtupulse.com

example	*citations*	*size*	*inLibrary*	*price*	*editions*	*buy*
1	some	small	no	affordable	many	no
2	many	big	no	expensive	one	yes
3	some	big	always	expensive	few	no
4	many	medium	no	expensive	many	yes
5	many	small	no	affordable	many	yes

Step 2:

h2 = (many, big, no, expensive, one)

X3 = (some, big, always, expensive, few) – No

Negative example hence Ignore

h3 = (many, big, no, expensive, one)

X4 = (many, medium, no, expensive, many) – Yes

h4 = (many, ?, no, expensive, ?)

example	*citations*	*size*	*inLibrary*	*price*	*editions*	*buy*
1	some	small	no	affordable	many	no
2	many	big	no	expensive	one	yes
3	some	big	always	expensive	few	no
4	many	medium	no	expensive	many	yes
5	many	small	no	affordable	many	yes

Step 2:

h4 = (many, ?, no, expensive, ?)

X5 = (many, small, no, affordable, many) – Yes

h5 = (many, ?, no, ?, ?)

Step 3:

Final Hypothesis is:

h5 = (many, ?, no, ?, ?)

Find S Algorithm – Solved Example – 3

Ex	Eyes	Nose	Head	Fcolor	Hair	Smile
1	Round	Triangle	Round	Purple	Yes	Yes
2	Square	Square	Square	Green	Yes	No
3	Square	Triangle	Round	Yellow	Yes	Yes
4	Round	Triangle	Round	Green	No	No
5	Square	Square	Round	Yellow	Yes	Yes

H0: (0, 0, 0, 0, 0)

Ex1: (Round, Triangle, Round, Purple, Yes) - POS

H1: (Round, Triangle, Round, Purple, Yes)

Ex2: (Square, Square, Square, Green, Yes) – NEG - IGNORE

H2: (Round, Triangle, Round, Purple, Yes)

Ex3: (Square, Triangle, Round, Yellow, Yes) - POS

H3: (?, Triangle, Round, ?, Yes)

Ex4: (Round, Triangle, Round, Green, No) – NEG - IGNORE

H4: (?, Triangle, Round, ?, Yes)

Ex5: (Square, Square, Round, Yellow, Yes) - POS

H5: (?, ?, Round, ?, Yes)

Hence, the Maximally Specific Hypothesis is:

(?, ?, Round, ?, Yes)

Example	*Sky*	*AirTemp*	*Humidity*	*Wind*	*Water*	*Forecast*	*EnjoySport*
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

S_0: ⟨∅, ∅, ∅, ∅, ∅. ∅⟩

S_1: ⟨*Sunny, Warm, Normal, Strong, Warm, Same*⟩

S_2: S_3: ⟨*Sunny, Warm, ?, Strong, Warm, Same*⟩

S_4 ⟨*Sunny, Warm, ?, Strong, ?, ?*⟩

vtupulse.com

G_4: ⟨*Sunny, ?, ?, ?, ?, ?*⟩ ⟨*?, Warm, ?, ?, ?, ?*⟩

G_3: ⟨*Sunny, ?, ?, ?, ?, ?*⟩ ⟨*?, Warm, ?, ?, ?, ?*⟩ ⟨*?, ?, Normal, ?, ?, ?*⟩ ⟨*?, ?, ?, ?, Cool, ?*⟩ ⟨*?, ?, ?, ?, ?, Same*⟩

G_0: G_1: G_2: ⟨?, ?, ?, ?, ?, ?⟩

Mahesh Huddar

Learned Version Space by Candidate Elimination Algorithm

S

⟨Sunny, Warm, ?, Strong, ?, ?⟩

⟨Sunny, ?, ?, Strong, ?, ?⟩

⟨Sunny, Warm, ?, ?, ?, ?⟩

⟨?, Warm, ?, Strong, ?, ?⟩

vtupulse.com

G

⟨Sunny, ?, ?, ?, ?, ?⟩

⟨?, Warm, ?, ?, ?, ?⟩

Mahesh Huddar

Candidate Elimination Algorithm Solved Example - 2

S0: (0, 0, 0)

S1: (0, 0, 0)

S2: (0, 0, 0)

S3: (Small, Red, Circle)

S4: (Small, Red, Circle)

S5: (Small, ?, Circle)

S: G: (Small, ?, Circle)

G5: (Small, ?, Circle)

G4: (Small, ?, Circle)

G3: (Small, ?, Circle)

G2: (Small, Blue, ?) (Small, ?, Circle) (?, Blue, ?) (Big, ?, Triangle) (?, Blue, Triangle)

G1: (Small, ?, ?) (?, Blue, ?) (?, ?, Triangle)

G0: (?, ?, ?)

Size	Color	Shape	Class / Label
Big	Red	Circle	No
Small	Red	Triangle	No
Small	Red	Circle	Yes
Big	Blue	Circle	No
Small	Blue	Circle	Yes

vtupulse.com

Mahesh Huddar

Candidate Elimination Algorithm Solved Example - 3

- S0: (0, 0, 0, 0, 0)
- S1: (0, 0, 0, 0, 0)
- S2: (Many, Big, No, Exp, Many)
- S3: (Many, ?, No, Exp, ?)
- S4: (Many, ?, No, ?, ?)

Example	Citations	Size	InLibrary	Price	Editions	Buy
1	Some	Small	No	Affordable	One	No
2	Many	Big	No	Expensive	Many	Yes
3	Many	Medium	No	Expensive	Few	Yes
4	Many	Small	No	Affordable	Many	Yes

Final Hypothesis Set: (Many, ?, No, ?, ?) (Many, ?, ?, ?, ?)

- G4: (Many,?,?,?, ?)
- G3: (Many,?,?,?, ?) (?,?,?,Exp,?)
- G2: (Many,?,?,?, ?) (?, Big,?,?,?) (?,?,?,Exp,?) (?,?,?,?,Many)
- G1: (Many,?,?,?, ?) (?, Big,?,?,?) (?,Medium,?,?,?) (?,?,?,Exp,?) (?,?,?,?,Many) (?,?,?,?,Few)
- G0: (?, ?, ?, ?, ?)

vtupulse.com

Subscribe to Mahesh Huddar

Candidate Elimination Algorithm - Solved Example - 4

Example	Shape	Size	Color	Surface	Thickness	Target Concept
1	Circular	Large	Light	Smooth	Thick	Malignant (+)
2	Circular	Large	Light	Irregular	Thick	Malignant (+)
3	Oval	Large	Dark	Smooth	Thin	Benign (-)
4	Oval	Large	Light	Irregular	Thick	Malignant (+)

S0: (0, 0, 0, 0, 0)

S1: (Circular, Large, Light, Smooth, Thick)

S2: (Circular, Large, Light, ?, Thick)

S3: (Circular, Large, Light, ?, Thick)

S4: (?, Large, Light, ?, Thick)

G4: (?, ?, Light, ?, ?) (?, ?, ?, Irregular, ?) (?, ?, ?, ?, Thick)

G3: ~~(Circular, ?, ?, ?, ?)~~ (?, ?, Light, ?, ?) (?, ?, ?, Irregular, ?) (?, ?, ?, ?, Thick)

~~(?, Large, ?, ?, ?)~~

G2: (?, ?, ?, ?, ?)

G1: (?, ?, ?, ?, ?)

G0: (?, ?, ?, ?, ?)

Candidate Elimination Algorithm - Solved Example - 5

Ex	Eyes	Nose	Head	Fcolor	Hair	Smile
1	Round	Triangle	Round	Purple	Yes	Yes
2	Square	Square	Square	Green	Yes	No
3	Square	Triangle	Round	Yellow	Yes	Yes
4	Round	Triangle	Round	Green	No	No
5	Square	Square	Round	Yellow	Yes	Yes

S0: (0, 0, 0, 0, 0)

S1: (Round, Triangle, Round, Purple, Yes)

S2: (Round, Triangle, Round, Purple, Yes)

S3: (?, Triangle, Round, ?, Yes)

S4: (?, Triangle, Round, ?, Yes)

S5: (?, ?, Round, ?, Yes)

G5: (?, ?, Round, ?, Yes)

G4: (Square, Triangle, ?, ?, ?) (?, Triangle, Square, ?, ?) (?, Triangle, ?, Yellow, ?) (?, Triangle, ?, Purple, ?) (?, Triangle, ?, ?, yes)
(Square, ?, Round, ?, ?) (?, Square, Round, ?, ?) (?, ?, Round, Yellow, ?) (?, ?, Round, Purple, ?) (?, ?, Round, ?, Yes)

G3: (?, Triangle, ?, ?, ?) (?, ?, Round, ?, ?)

G2: (Round, ?, ?, ?, ?) (?, Triangle, ?, ?, ?) (?, ?, Round, ?, ?) (?, ?, ?, Purple, ?)

G1: (?, ?, ?, ?, ?)

G0: (?, ?, ?, ?, ?)

Day	Outlook	Temp	Humidity	Wind	Play Tennis
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

Attribute: Outlook

$Values\ (Outlook) = Sunny, Overcast, Rain$

$S = [9+, 5-]$ $\qquad Entropy(S) = -\frac{9}{14}log_2\frac{9}{14} - \frac{5}{14}log_2\frac{5}{14} = 0.94$

$S_{Sunny} \leftarrow [2+, 3-]$ $\qquad Entropy(S_{Sunny}) = -\frac{2}{5}log_2\frac{2}{5} - \frac{3}{5}log_2\frac{3}{5} = 0.971$

$S_{Overcast} \leftarrow [4+, 0-]$ $\qquad Entropy(S_{Overcast}) = -\frac{4}{4}log_2\frac{4}{4} - \frac{0}{4}log_2\frac{0}{4} = 0$

$S_{Rain} \leftarrow [3+, 2-]$ $\qquad Entropy(S_{Rain}) = -\frac{3}{5}log_2\frac{3}{5} - \frac{2}{5}log_2\frac{2}{5} = 0.971$

$$Gain\ (S, Outlook) = Entropy(S) - \sum_{v \in \{Sunny, Overcast, Rain\}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$Gain(S, Outlook)$$

$$= Entropy(S) - \frac{5}{14}Entropy(S_{Sunny}) - \frac{4}{14}Entropy(S_{Overcast})$$

$$-\frac{5}{14}Entropy(S_{Rain})$$

$$Gain(S, Outlook) = 0.94 - \frac{5}{14}0.971 - \frac{4}{14}\ 0 - \frac{5}{14}0.971 = 0.2464$$

Day	Outlook	Temp	Humidity	Wind	Play Tennis
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

Attribute: Temp

$Values\ (Temp) = Hot, Mild, Cool$

$S = [9+, 5-]$ $\quad Entropy(S) = -\frac{9}{14}log_2\frac{9}{14} - \frac{5}{14}log_2\frac{5}{14} = 0.94$

$S_{Hot} \leftarrow [2+, 2-]$ $\quad Entropy(S_{Hot}) = -\frac{2}{4}log_2\frac{2}{4} - \frac{2}{4}log_2\frac{2}{4} = 1.0$

$S_{Mild} \leftarrow [4+, 2-]$ $\quad Entropy(S_{Mild}) = -\frac{4}{6}log_2\frac{4}{6} - \frac{2}{6}log_2\frac{2}{6} = 0.9183$

$S_{Cool} \leftarrow [3+, 1-]$ $\quad Entropy(S_{Cool}) = -\frac{3}{4}log_2\frac{3}{4} - \frac{1}{4}log_2\frac{1}{4} = 0.8113$

$$Gain\ (S, Temp) = Entropy(S) - \sum_{v \in \{Hot, Mild, Cool\}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$Gain(S, Temp)$$

$$= Entropy(S) - \frac{4}{14}Entropy(S_{Hot}) - \frac{6}{14}Entropy(S_{Mild})$$

$$- \frac{4}{14}Entropy(S_{Cool})$$

$$Gain(S, Temp) = 0.94 - \frac{4}{14}1.0 - \frac{6}{14}\ 0.9183 - \frac{4}{14}0.8113 = 0.0289$$

Mahesh Huddar

vtupulse.com

Day	Outlook	Temp	Humidity	Wind	Play Tennis
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

Attribute: Humidity

$Values\ (Humidity) = High, Normal$

$S = [9+, 5-]$ $\quad Entropy(S) = -\frac{9}{14} log_2 \frac{9}{14} - \frac{5}{14} log_2 \frac{5}{14} = 0.94$

$S_{High} \leftarrow [3+, 4-]$ $\quad Entropy(S_{High}) = -\frac{3}{7} log_2 \frac{3}{7} - \frac{4}{7} log_2 \frac{4}{7} = 0.9852$

$S_{Normal} \leftarrow [6+, 1-]$ $\quad Entropy(S_{Normal}) = -\frac{6}{7} log_2 \frac{6}{7} - \frac{1}{7} log_2 \frac{1}{7} = 0.5916$

$$Gain\ (S, Humidity) = Entropy(S) - \sum_{v \in \{High, Normal\}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$Gain(S, Humidity)$$

$$= Entropy(S) - \frac{7}{14} Entropy(S_{High}) - \frac{7}{14} Entropy(S_{Normal})$$

$$Gain(S, Humidity) = 0.94 - \frac{7}{14} 0.9852 - \frac{7}{14}\ 0.5916 = 0.1516$$

Day	Outlook	Temp	Humidity	Wind	Play Tennis
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

Attribute: Wind

$Values\ (Wind) = Strong, Weak$

$S = [9+, 5-]$ $\qquad Entropy(S) = -\frac{9}{14} log_2 \frac{9}{14} - \frac{5}{14} log_2 \frac{5}{14} = 0.94$

$S_{Strong} \leftarrow [3+, 3-]$ $\qquad Entropy(S_{Strong}) = 1.0$

$S_{Weak} \leftarrow [6+, 2-]$ $\qquad Entropy(S_{Weak}) = -\frac{6}{8} log_2 \frac{6}{8} - \frac{2}{8} log_2 \frac{2}{8} = 0.8113$

$$Gain\ (S, Wind) = Entropy(S) - \sum_{v \in \{Strong, Weak\}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$Gain(S, Wind) = Entropy(S) - \frac{6}{14} Entropy(S_{Strong}) - \frac{8}{14} Entropy(S_{Weak})$$

$$Gain(S, Wind) = 0.94 - \frac{6}{14}\ 1.0 - \frac{8}{14}\ 0.8113 = 0.0478$$

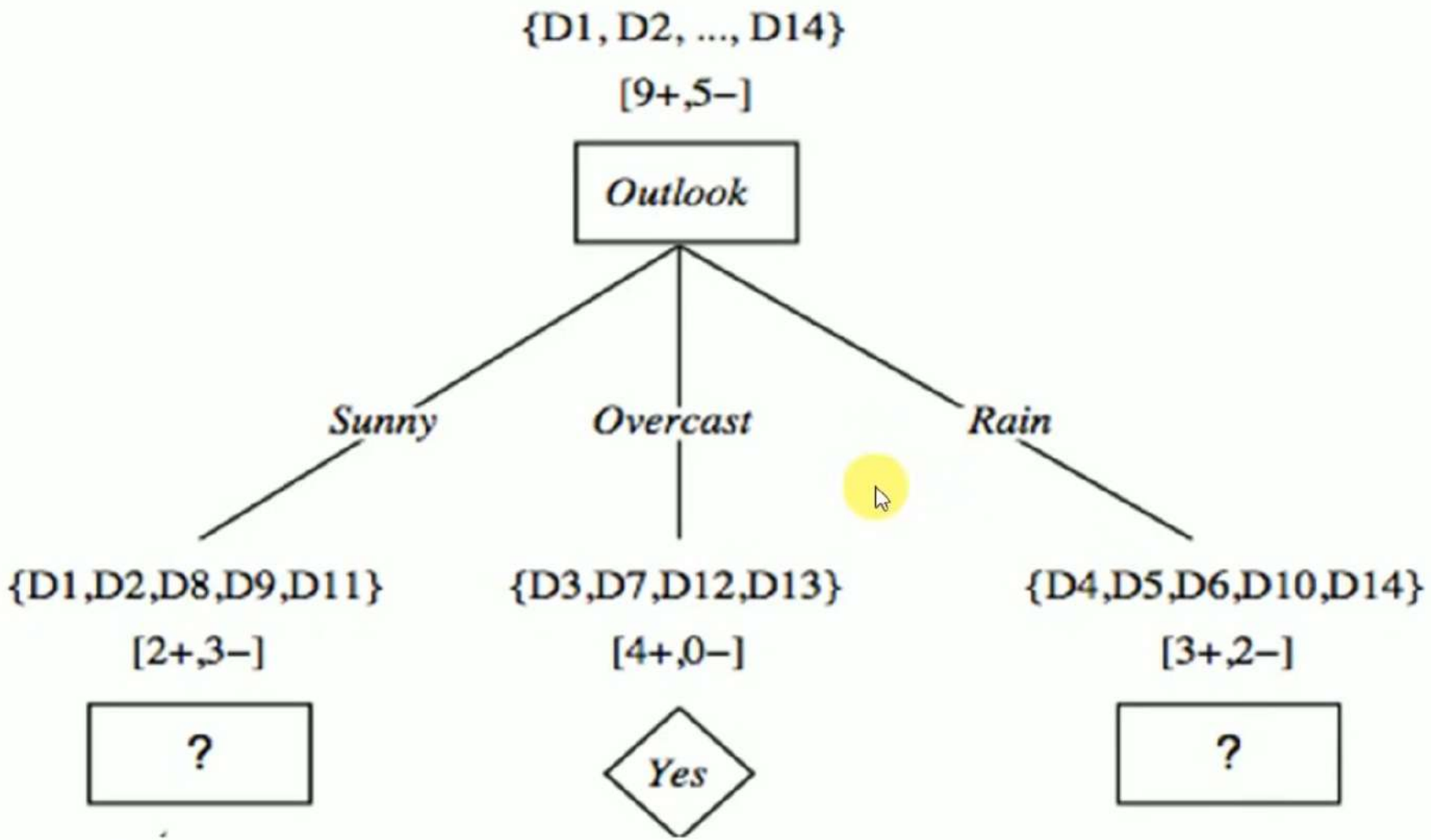
Day	Outlook	Temp	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	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

$$Gain(S, Outlook) = 0.2464$$

$$Gain(S, Temp) = 0.0289$$

$$Gain(S, Humidity) = 0.1516$$

$$Gain(S, Wind) = 0.0478$$


{D1, D2, ..., D14}
[9+,5−]
Outlook
Sunny
Overcast
Rain
{D1,D2,D8,D9,D11}
[2+,3−]
?
{D3,D7,D12,D13}
[4+,0−]
Yes
{D4,D5,D6,D10,D14}
[3+,2−]
?

Day	Temp	Humidity	Wind	Play Tennis
D1	Hot	High	Weak	No
D2	Hot	High	Strong	No
D8	Mild	High	Weak	No
D9	Cool	Normal	Weak	Yes
D11	Mild	Normal	Strong	Yes

Attribute: Temp

$Values\ (Temp) = Hot, Mild, Cool$

$S_{Sunny} = [2+, 3-]$ $\quad Entropy(S_{Sunny}) = -\frac{2}{5}log_2\frac{2}{5} - \frac{3}{5}log_2\frac{3}{5} = 0.97$

$S_{Hot} \leftarrow [0+, 2-]$ $\quad Entropy(S_{Hot}) = 0.0$

$S_{Mild} \leftarrow [1+, 1-]$ $\quad Entropy(S_{Mild}) = 1.0$

$S_{Cool} \leftarrow [1+, 0-]$ $\quad Entropy(S_{Cool}) = 0.0$

$$Gain\ (S_{Sunny}, Temp) = Entropy(S) - \sum_{v \in \{Hot, Mild, Cool\}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$Gain(S_{Sunny}, Temp)$$

$$= Entropy(S) - \frac{2}{5}Entropy(S_{Hot}) - \frac{2}{5}Entropy(S_{Mild})$$

$$- \frac{1}{5}Entropy(S_{Cool})$$

$$Gain(S_{sunny}, Temp) = 0.97 - \frac{2}{5}0.0 - \frac{2}{5}1 - \frac{1}{5}0.0 = 0.570$$

Day	Temp	Humidity	Wind	Play Tennis
Dl	Hot	High	Weak	No
D2	Hot	High	Strong	No
D8	Mild	High	Weak	No
D9	Cool	Normal	Weak	Yes
Dl1	Mild	Normal	Strong	Yes

Attribute: Humidity

$Values\ (Humidity) = High, Normal$

$S_{Sunny} = [2+, 3-]$

$S_{high} \leftarrow [0+, 3-]$

$S_{Normal} \leftarrow [2+, 0-]$

$Entropy(S) = -\frac{2}{5}log_2\frac{2}{5} - \frac{3}{5}log_2\frac{3}{5} = 0.97$

$Entropy(S_{High}) = 0.0$

$Entropy(S_{Normal}) = 0.0$

$$Gain\ (S_{Sunny}, Humidity) = Entropy(S) - \sum_{v \in \{High, Normal\}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$Gain(S_{Sunny}, Humidity) = Entropy(S) - \frac{3}{5}Entropy(S_{High}) - \frac{2}{5}Entropy(S_{Normal})$$

$$Gain(S_{sunny}, Humidity) = 0.97 - \frac{3}{5}\ 0.0 - \frac{2}{5}0.0 = 0.97$$

Day	Temp	Humidity	Wind	Play Tennis
Dl	Hot	High	Weak	No
D2	Hot	High	Strong	No
D8	Mild	High	Weak	No
D9	Cool	Normal	Weak	Yes
Dl1	Mild	Normal	Strong	Yes

Attribute: Wind

$Values\ (Wind) = Strong, Weak$

$S_{Sunny} = [2+, 3-]$

$S_{Strong} \leftarrow [1+, 1-]$

$S_{Weak} \leftarrow [1+, 2-]$

$Entropy(S) = -\frac{2}{5}log_2\frac{2}{5} - \frac{3}{5}log_2\frac{3}{5} = 0.97$

$Entropy(S_{Strong}) = 1.0$

$Entropy(S_{Weak}) = -\frac{1}{3}log_2\frac{1}{3} - \frac{2}{3}log_2\frac{2}{3} = 0.9183$

$$Gain\ (S_{Sunny}, Wind) = Entropy(S) - \sum_{v \in \{Strong, Weak\}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$Gain(S_{Sunny}, Wind) = Entropy(S) - \frac{2}{5}Entropy(S_{Strong}) - \frac{3}{5}Entropy(S_{Weak})$$

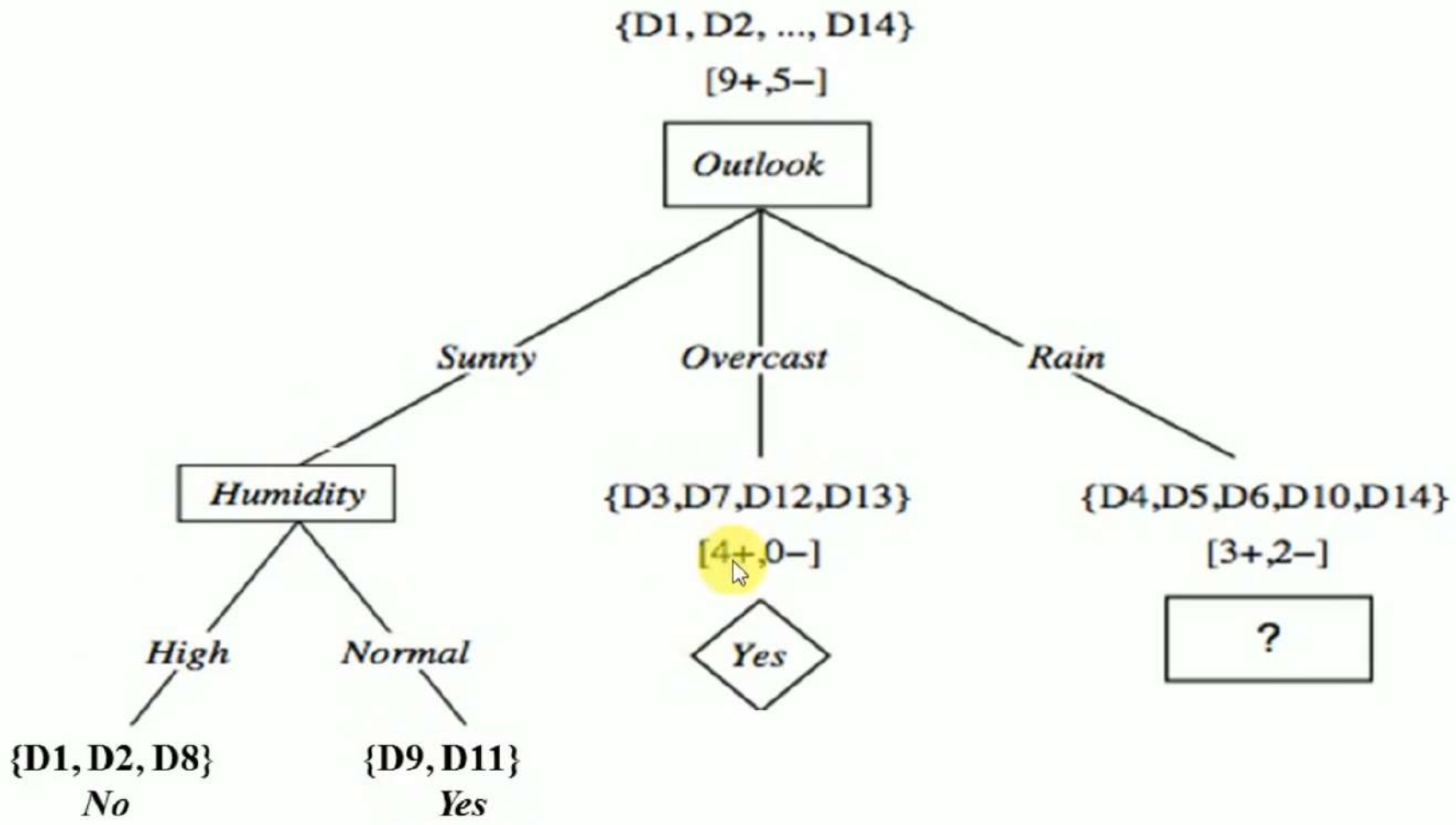
$$Gain(S_{sunny}, Wind) = 0.97 - \frac{2}{5}1.0 - \frac{3}{5}\ 0.918 = 0.0192$$

Day	Temp	Humidity	Wind	Play Tennis
D1	Hot	High	Weak	No
D2	Hot	High	Strong	No
D8	Mild	High	Weak	No
D9	Cool	Normal	Weak	Yes
D11	Mild	Normal	Strong	Yes

$$Gain(S_{sunny}, Temp) = 0.570$$

$$Gain(S_{sunny}, Humidity) = 0.97$$

$$Gain(S_{sunny}, Wind) = 0.0192$$


{D1, D2, ..., D14}
[9+,5−]
Outlook
Sunny
Overcast
Rain
Humidity
{D3,D7,D12,D13}
[4+,0−]
{D4,D5,D6,D10,D14}
[3+,2−]
High
Normal
Yes
?
{D1, D2, D8}
No
{D9, D11}
Yes

Day	Temp	Humidity	Wind	Play Tennis
D4	Mild	High	Weak	Yes
D5	Cool	Normal	Weak	Yes
D6	Cool	Normal	Strong	No
D10	Mild	Normal	Weak	Yes
D14	Mild	High	Strong	No

Attribute: Temp

$Values\ (Temp) = Hot, Mild, Cool$

$S_{Rain} = [3+, 2-]$ $\quad Entropy(S_{Sunny}) = -\frac{3}{5}log_2\frac{3}{5} - \frac{2}{5}log_2\frac{2}{5} = 0.97$

$S_{Hot} \leftarrow [0+, 0-]$ $\quad Entropy(S_{Hot}) = 0.0$

$S_{Mild} \leftarrow [2+, 1-]$ $\quad Entropy(S_{Mild}) = -\frac{2}{3}log_2\frac{2}{3} - \frac{1}{3}log_2\frac{1}{3} = 0.9183$

$S_{Cool} \leftarrow [1+, 1-]$ $\quad Entropy(S_{Cool}) = 1.0$

$$Gain\ (S_{Rain}, Temp) = Entropy(S) - \sum_{v \in \{Hot, Mild, Cool\}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$Gain(S_{Rain}, Temp)$$

$$= Entropy(S) - \frac{0}{5}Entropy(S_{Hot}) - \frac{3}{5}Entropy(S_{Mild})$$

$$- \frac{2}{5}Entropy(S_{Cool})$$

$$Gain(S_{Rain}, Temp) = 0.97 - \frac{0}{5}0.0 - \frac{3}{5}\ 0.918 - \frac{2}{5}\ 1.0 = 0.0192$$

Day	Temp	Humidity	Wind	Play Tennis
D4	Mild	High	Weak	Yes
D5	Cool	Normal	Weak	Yes
D6	Cool	Normal	Strong	No
DI0	Mild	Normal	Weak	Yes
DI4	Mild	High	Strong	No

Attribute: Humidity

$Values\ (Humidity) = High, Normal$

$S_{Rain} = [3+, 2-]$

$S_{High} \leftarrow [1+, 1-]$

$S_{Normal} \leftarrow [2+, 1-]$

$Entropy(S_{Sunny}) = -\frac{3}{5}log_2\frac{3}{5} - \frac{2}{5}log_2\frac{2}{5} = 0.97$

$Entropy(S_{High}) = 1.0$

$Entropy(S_{Normal}) = -\frac{2}{3}log_2\frac{2}{3} - \frac{1}{3}log_2\frac{1}{3} = 0.9183$

$$Gain\ (S_{Rain}, Humidity) = Entropy(S) - \sum_{v \in \{High, Normal\}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$Gain(S_{Rain}, Humidity) = Entropy(S) - \frac{2}{5}Entropy(S_{High}) - \frac{3}{5}Entropy(S_{Normal})$$

$$Gain(S_{Rain}, Humidity) = 0.97 - \frac{2}{5}\ 1.0 - \frac{3}{5}0.918 = 0.0192$$

Day	Temp	Humidity	Wind	Play Tennis
D4	Mild	High	Weak	Yes
D5	Cool	Normal	Weak	Yes
D6	Cool	Normal	Strong	No
DI0	Mild	Normal	Weak	Yes
DI4	Mild	High	Strong	No

Attribute: Wind

$Values\ (wind) = Strong, Weak$

$S_{Rain} = [3+, 2-]$

$S_{Strong} \leftarrow [0+, 2-]$

$S_{Weak} \leftarrow [3+, 0-]$

$Entropy(S_{Sunny}) = -\frac{3}{5}log_2\frac{3}{5} - \frac{2}{5}log_2\frac{2}{5} = 0.97$

$Entropy(S_{Strong}) = 0.0$

$Entropy(S_{weak}) = 0.0$

$$Gain\ (S_{Rain}, Wind) = Entropy(S) - \sum_{v \in \{Strong, Weak\}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$Gain(S_{Rain}, Wind) = Entropy(S) - \frac{2}{5}Entropy(S_{Strong}) - \frac{3}{5}Entropy(S_{Weak})$$

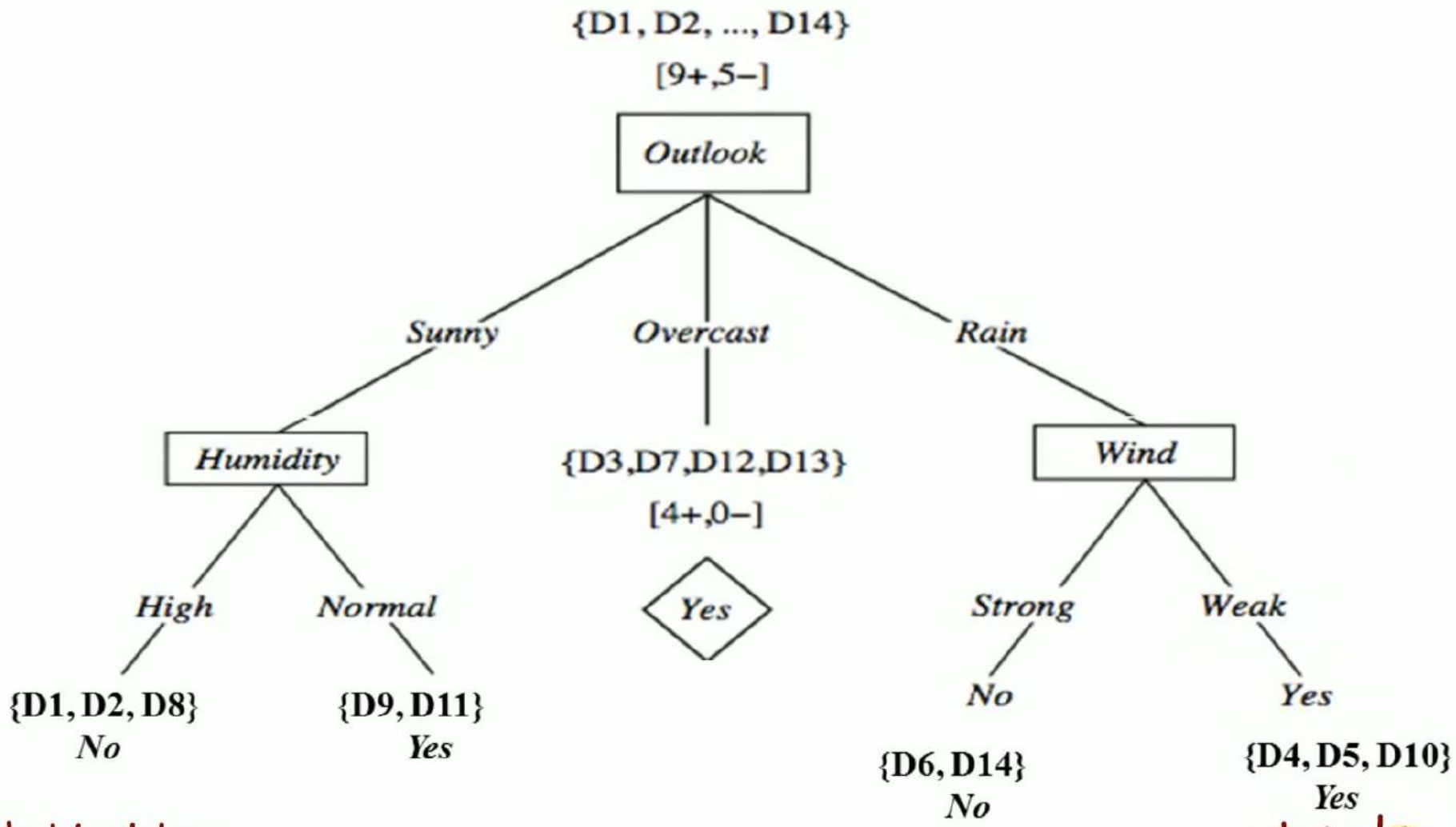
$$Gain(S_{Rain}, Wind) = 0.97 - \frac{2}{5}\ 0.0 - \frac{3}{5}0.0 = 0.97$$

Day	Temp	Humidity	Wind	Play Tennis
D4	Mild	High	Weak	Yes
D5	Cool	Normal	Weak	Yes
D6	Cool	Normal	Strong	No
Dl0	Mild	Normal	Weak	Yes
Dl4	Mild	High	Strong	No

$$Gain(S_{Rain}, Temp) = 0.0192$$

$$Gain(S_{Rain}, Humidity) = 0.0192$$

$$Gain(S_{Rain}, Wind) = 0.97$$



Mahesh Huddar
vtupulse.com

Instance	Classification	a1	a2
1	+	T	T
2	+	T	T
3	-	T	F
4	+	F	F
5	-	F	T
6	-	F	T

Attribute: a1

$Values\,(a1) = T, F$

$S = [3+,\ 3-]$ $\qquad Entropy(S) = 1.0$

$S_T = [2+,\ 1-]$ $\qquad Entropy(S_T) = -\frac{2}{3}log_2\frac{2}{3} - \frac{1}{3}log_2\frac{1}{3} = 0.9183$

$S_F \leftarrow [1+,\ 2-]$ $\qquad Entropy(S_F) = -\frac{1}{3}log_2\frac{1}{3} - \frac{2}{3}log_2\frac{2}{3} = 0.9183$

$$Gain\,(S, a1) = Entropy(S) - \sum_{v \in \{T,F\}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$Gain(S, a1) = Entropy(S) - \frac{3}{6}Entropy(S_T) - \frac{3}{6}Entropy(S_F)$$

$$Gain(S, a1) = 1.0 - \frac{3}{6} * 0.9183 - \frac{3}{6} * 0.9183 = 0.0817$$

Example - 2
Decision Tree Algorithm – ID3
Solved Example

Instance	Classification	a1	a2
1	+	T	T
2	+	T	T
3	-	T	F
4	+	F	F
5	-	F	T
6	-	F	T

Attribute: a2

$Values\ (a2) = T, F$

$S = [3+,\ 3-]$ $\quad$ $Entropy(S) = 1.0$

$S_T = [2+,\ 2-]$ $\quad$ $Entropy(S_T) = 1.0$

$S_F \leftarrow [1+,\ 1-]$ $\quad$ $Entropy(S_F) = 1.0$

$$Gain\ (S, a2) = Entropy(S) - \sum_{v \in \{T,F\}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$Gain(S, a2) = Entropy(S) - \frac{4}{6} Entropy(S_T) - \frac{2}{6} Entropy(S_F)$$

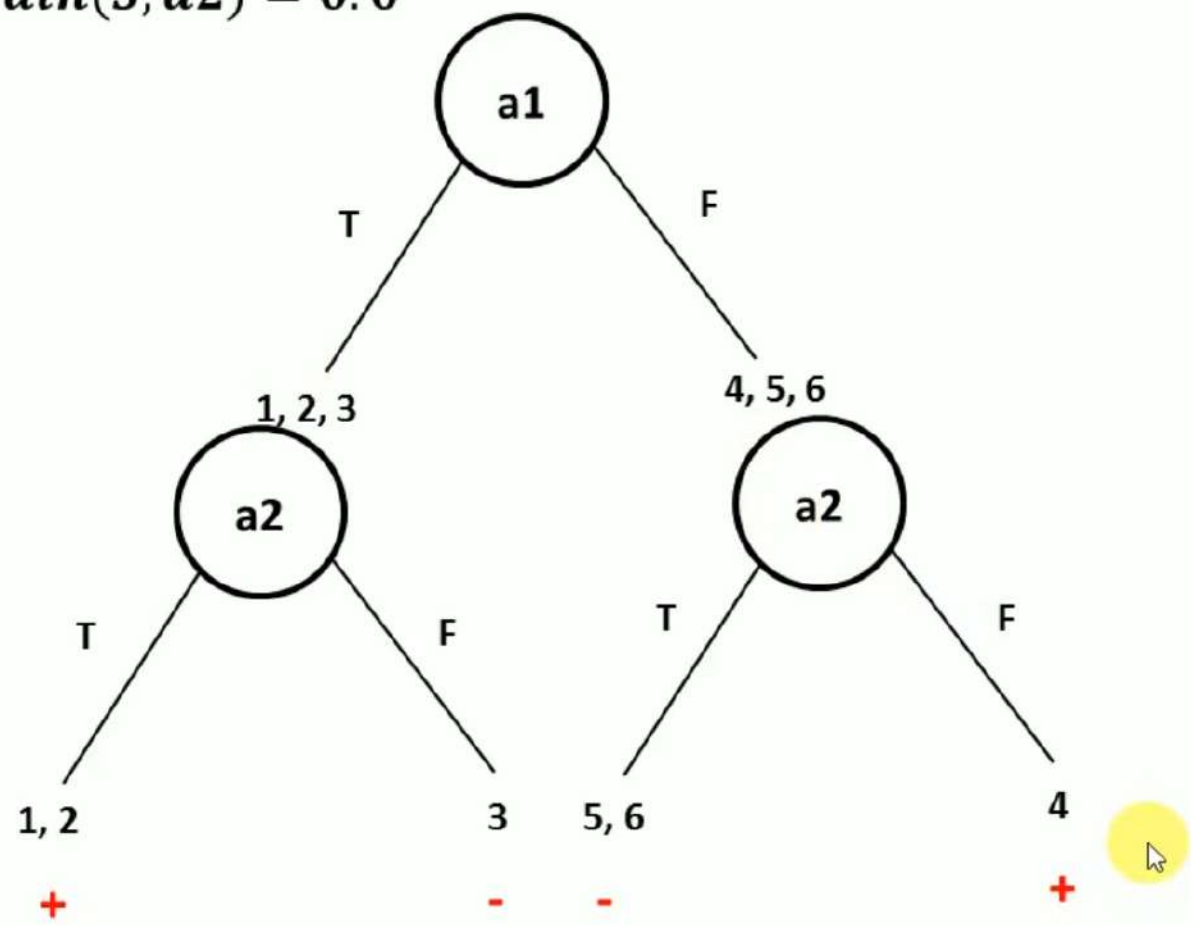
$$Gain(S, a2) = 1.0 - \frac{4}{6} * 1.0 - \frac{2}{6} * 1.0 = 0.0$$

Example - 2
Decision Tree Algorithm – ID3
Solved Example

Instance	Classification	a1	a2
1	+	T	T
2	+	T	T
3	-	T	F
4	+	F	F
5	-	F	T
6	-	F	T

$$Gain(S, a1) = 0.0817 \quad - \; Maximum\ Gain$$

$$Gain(S, a2) = 0.0$$



Example - 2
Decision Tree Algorithm – ID3
Solved Example

Instance	a1	a2	a3	Classification
1	True	Hot	High	No
2	True	Hot	High	No
3	False	Hot	High	Yes
4	False	Cool	Normal	Yes
5	False	Cool	Normal	Yes
6	True	Cool	High	No
7	True	Hot	High	No
8	True	Hot	Normal	Yes
9	False	Cool	Normal	Yes
10	False	Cool	High	Yes

Attribute: a1

$$Values\,(a1) = True,\ False$$

$$S = [6+,\ 4-] \qquad Entropy(S) = -\frac{6}{10}log_2\frac{6}{10} - \frac{4}{10}log_2\frac{4}{10} = 0.9709$$

$$S_{True} = [1+, 4-] \qquad Entropy(S_{True}) = -\frac{1}{5}log_2\frac{1}{5} - \frac{4}{5}log_2\frac{4}{5} = 0.7219$$

$$S_{Flase} \leftarrow [5+,\ 0-] \qquad Entropy(S_{False}) = 0.0$$

$$Gain\,(S, a1) = Entropy(S) - \sum_{v \in \{True, False\}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$Gain(S, a1) = Entropy(S) - \frac{5}{10}Entropy(S_{True}) - \frac{5}{10}Entropy(S_{False})$$

$$Gain(S, a1) = 0.9709 - \frac{5}{10} * 0.7219 - \frac{5}{10} * 1 = 0.6099$$

Example - 3
Decision Tree Algorithm – ID3
Solved Example

Instance	a1	a2	a3	Classification
1	True	Hot	High	No
2	True	Hot	High	No
3	False	Hot	High	Yes
4	False	Cool	Normal	Yes
5	False	Cool	Normal	Yes
6	True	Cool	High	No
7	True	Hot	High	No
8	True	Hot	Normal	Yes
9	False	Cool	Normal	Yes
10	False	Cool	High	Yes

Attribute: a2

$Values\ (a2) = Hot,\ Cool$

$S = [6+,\ 4-]$ $\quad Entropy(S) = -\frac{6}{10}log_2\frac{6}{10} - \frac{4}{10}log_2\frac{4}{10} = 0.9709$

$S_{Hot} = [2+, 3-]$ $\quad Entropy(S_{Hot}) = -\frac{2}{5}log_2\frac{2}{5} - \frac{3}{5}log_2\frac{3}{5} = 0.9709$

$S_{Cool} \leftarrow [4+,\ 1-]$ $\quad Entropy(S_{Cool}) = -\frac{4}{5}log_2\frac{4}{5} - \frac{1}{5}log_2\frac{1}{5} = 0.7219$

$$Gain\ (S, a2) = Entropy(S) - \sum_{v \in \{Hot, Cool\}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$Gain(S, a2) = Entropy(S) - \frac{5}{10} Entropy(S_{Hot}) - \frac{5}{10} Entropy(S_{Cool})$$

$$Gain(S, a2) = 0.9709 - \frac{5}{10} * 0.9709 - \frac{5}{10} * 0.7219 = 0.1245$$

Example - 3
Decision Tree Algorithm – ID3
Solved Example

Instance	a1	a2	a3	Classification
1	True	Hot	High	No
2	True	Hot	High	No
3	False	Hot	High	Yes
4	False	Cool	Normal	Yes
5	False	Cool	Normal	Yes
6	True	Cool	High	No
7	True	Hot	High	No
8	True	Hot	Normal	Yes
9	False	Cool	Normal	Yes
10	False	Cool	High	Yes

Attribute: a3

$Values\ (a3) = High,\ Normal$

$S = [6+,\ 4-]$ $\quad Entropy(S) = -\frac{6}{10}log_2\frac{6}{10} - \frac{4}{10}log_2\frac{4}{10} = 0.9709$

$S_{High} = [2+, 4-]$ $\quad Entropy(S_{High}) = -\frac{2}{6}log_2\frac{2}{6} - \frac{4}{6}log_2\frac{4}{6} = 0.9183$

$S_{Normal} \leftarrow [4+,\ 0-]$ $\quad Entropy(S_{Normal}) = 0.0$

$$Gain\ (S, a3) = Entropy(S) - \sum_{v \in \{High, Normal\}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$Gain(S, a3) = Entropy(S) - \frac{6}{10}Entropy(S_{High}) - \frac{4}{10}Entropy(S_{Normal})$$

$$Gain(S, a3) = 0.9709 - \frac{6}{10} * 0.9183 - \frac{4}{10} * 0.0 = 0.4199$$

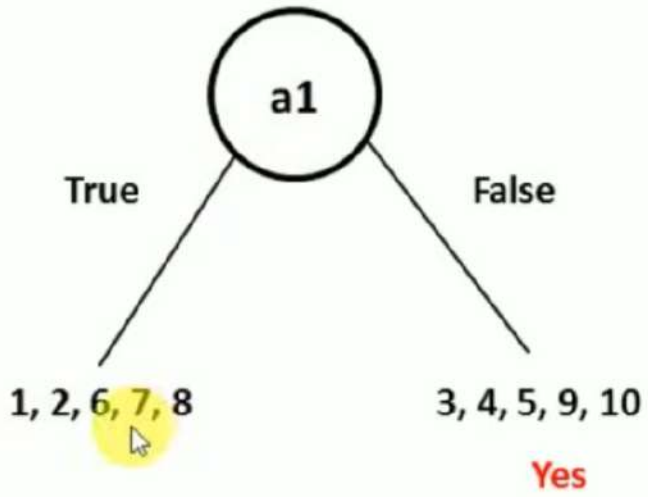
Example - 3
Decision Tree Algorithm – ID3
Solved Example

Instance	a1	a2	a3	Classification
1	True	Hot	High	No
2	True	Hot	High	No
3	False	Hot	High	Yes
4	False	Cool	Normal	Yes
5	False	Cool	Normal	Yes
6	True	Cool	High	No
7	True	Hot	High	No
8	True	Hot	Normal	Yes
9	False	Cool	Normal	Yes
10	False	Cool	High	Yes

$$Gain(S, a1) = 0.6099 \quad - \; Maximum\ Gain$$

$$Gain(S, a2) = 0.1245$$

$$Gain(S, a3) = 0.4199$$



Example - 3
Decision Tree Algorithm – ID3
Solved Example

Instance	a2	a3	Classification
1	Hot	High	No
2	Hot	High	No
6	Cool	High	No
7	Hot	High	No
8	Hot	Normal	Yes

Attribute: a2

$Values\,(a2) = Hot,\ Cool$

$S_{a1} = [1+,\ 4-]$ $\quad Entropy(S_{a1}) = -\frac{1}{5}log_2\frac{1}{5} - \frac{4}{5}log_2\frac{4}{5} = 0.7219$

$S_{Hot} = [1+, 3-]$ $\quad Entropy(S_{Hot}) = -\frac{1}{4}log_2\frac{1}{4} - \frac{3}{4}log_2\frac{3}{4} = 0.8112$

$S_{Cool} \leftarrow [0+,\ 1-]$ $\quad Entropy(S_{Cool}) = 0.0$

$$Gain\,(S, a2) = Entropy(S) - \sum_{v \in \{Hot, Cool\}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$Gain(S, a2) = Entropy(S) - \frac{4}{5}Entropy(S_{Hot}) - \frac{1}{5}Entropy(S_{Cool})$$

$$Gain(S, a2) = 0.9709 - \frac{4}{5} * 0.8112 - \frac{1}{5} * 0.0 = 0.3219$$

Example - 3
Decision Tree Algorithm – ID3
Solved Example

Instance	a2	a3	Classification
1	Hot	High	No
2	Hot	High	No
6	Cool	High	No
7	Hot	High	No
8	Hot	Normal	Yes

Attribute: a3

$Values\ (a3) = High,\ Normal$

$S_{a1} = [1+,\ 4-]$ $\quad Entropy(S_{a1}) = -\frac{1}{5}log_2\frac{1}{5} - \frac{4}{5}log_2\frac{4}{5} = 0.7219$

$S_{High} = [0+, 4-]$ $\quad Entropy(S_{High}) = 0.0$

$S_{Normal} \leftarrow [1+,\ 0-]$ $\quad Entropy(S_{Normal}) = 0.0$

$$Gain\ (S, a3) = Entropy(S) - \sum_{v \in \{High, Normal\}} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$Gain(S, a3) = Entropy(S) - \frac{4}{5} Entropy(S_{High}) - \frac{1}{5} Entropy(S_{Normal})$$

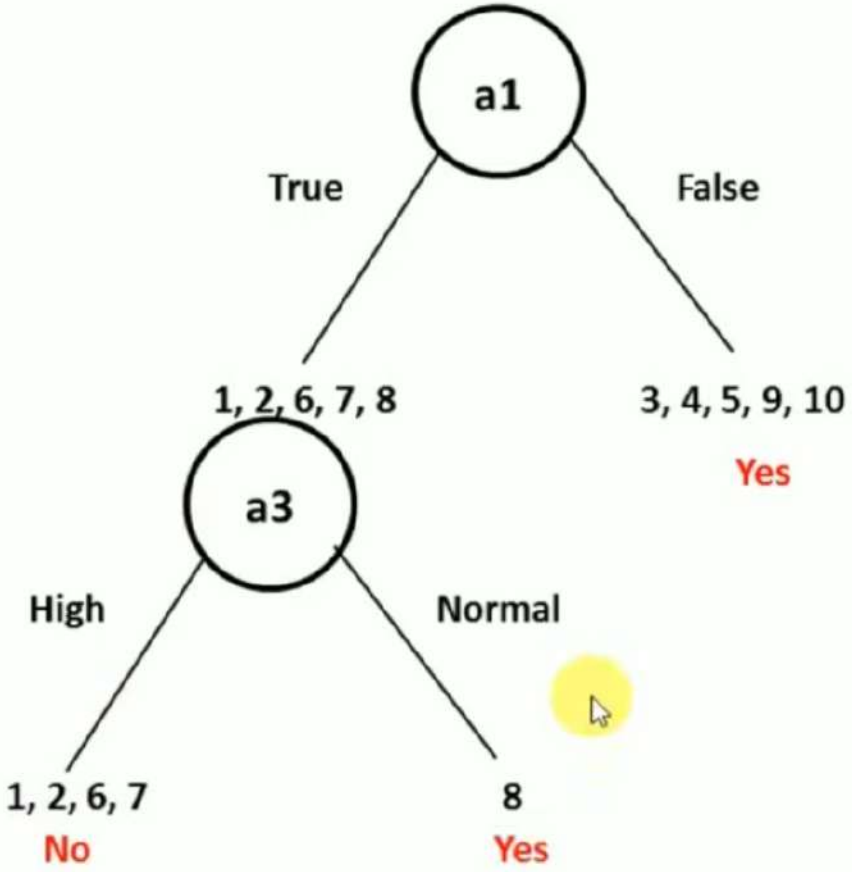
$$Gain(S, a3) = 0.9709 - \frac{4}{5} * 0.0 - \frac{1}{5} * 0.0 = 0.7219$$

Example - 3
Decision Tree Algorithm – ID3
Solved Example

Instance	a2	a3	Classification
1	Hot	High	No
2	Hot	High	No
6	Cool	High	No
7	Hot	High	No
8	Hot	Normal	Yes

$$Gain(S_{a1}, a2) = 0.3219$$

$$Gain(S_{a1}, a3) = 0.7219 - \textit{Maximum Gain}$$



Example - 3
Decision Tree Algorithm – ID3
Solved Example

Day	*Outlook*	*Temperature*	*Humidity*	*Wind*	*PlayTennis*
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

$$P(PlayTennis = yes) = 9/14 = .64$$

$$P(PlayTennis = no) = 5/14 = .36$$

NAIVE BAYES CLASSIFIER

Example - 1

Outlook	Y	N
sunny	2/9	3/5
overcast	4/9	0
rain	3/9	2/5
Tempreature		
hot	2/9	2/5
mild	4/9	2/5
cool	3/9	1/5

Humidity	Y	N
high	3/9	4/5
normal	6/9	1/5
Windy		
Strong	3/9	3/5
Weak	6/9	2/5

NAIVE BAYES CLASSIFIER – Example -1

$$\langle Outlook = sunny, Temperature = cool, Humidity = high, Wind = strong \rangle$$

$$v_{NB} = \underset{v_j \in \{yes, no\}}{\operatorname{argmax}} P(v_j) \prod_i P(a_i | v_j)$$

$$= \underset{v_j \in \{yes, no\}}{\operatorname{argmax}} P(v_j) \quad P(Outlook = sunny | v_j) P(Temperature = cool | v_j)$$

$$\cdot P(Humidity = high | v_j) P(Wind = strong | v_j)$$

$$v_{NB}(yes) = P(yes)\ P(sunny|yes)\ P(cool|yes)\ P(high|yes)\ P(strong|yes) = .0053$$

$$v_{NB}(no) = P(no)\ P(sunny|no)\ P(cool|no)\ P(high|no)\ P(strong|no) = .0206$$

$$v_{NB}(yes) = \frac{v_{NB}(yes)}{v_{NB}(yes) + v_{NB}(no)} = 0.205 \qquad v_{NB}(no) = \frac{v_{NB}(no)}{v_{NB}(yes) + v_{NB}(no)} = 0.795$$

Suggested: 5. Solved Example Naive Bayes Classifier to classify ...

NAIVE BAYES CLASSIFIER Example – 2

- Estimate conditional probabilities of each attributes {color, legs, height, smelly} for the species classes: {M, H} using the data given in the table.
- Using these probabilities estimate the probability values for the new instance – (Color=Green, legs=2, Height=Tall, and Smelly=No).

No	Color	Legs	Height	Smelly	Species
1	White	3	Short	Yes	M
2	Green	2	Tall	No	M
3	Green	3	Short	Yes	M
4	White	3	Short	Yes	M
5	Green	2	Short	No	H
6	White	2	Tall	No	H
7	White	2	Tall	No	H
8	White	2	Short	Yes	H

Subscribe to Mahesh Huddar

Visit: vtupulse.com

NAIVE BAYES CLASSIFIER

EXAMPLE - 2

No	Color	Legs	Height	Smelly	Species
1	White	3	Short	Yes	M
2	Green	2	Tall	No	M
3	Green	3	Short	Yes	M
4	White	3	Short	Yes	M
5	Green	2	Short	No	H
6	White	2	Tall	No	H
7	White	2	Tall	No	H
8	White	2	Short	Yes	H

New Instance

(Color=Green, legs=2, Height=Tall, and Smelly=No)

$$P(M) = \frac{4}{8} = 0.5 \quad P(H) = \frac{4}{8} = 0.5$$

Color	M	H
White	2/4	3/4
Green	2/4	1/4

Legs	M	H
2	1/4	4/4
3	3/4	0/4

Height	M	H
Tall	3/4	2/4
Short	1/4	2/4

Smelly	M	H
Yes	3/4	1/4
No	1/4	3/4

NAIVE BAYES CLASSIFIER - EXAMPLE - 2

$$P(M) = \frac{4}{8} = 0.5 \quad P(H) = \frac{4}{8} = 0.5$$

Color	M	H
White	2/4	3/4
Green	2/4	1/4

Legs	M	H
2	1/4	4/4
3	3/4	0/4

Height	M	H
Tall	3/4	2/4
Short	1/4	2/4

Smelly	M	H
Yes	3/4	1/4
No	1/4	3/4

$$p(M|New\ Instance) = p(M) * p(Color = Green|M) * p(Legs = 2|M) * p(Height = tall|M) * p(Smelly = no\ |M)$$

$$p(M|New\ Instance) = 0.5 * \frac{2}{4} * \frac{1}{4} * \frac{3}{4} * \frac{1}{4} = 0.0117$$

$$p(H|New\ Instance) = p(H) * p(Color = Green|H) * p(Legs = 2|H) * p(Height = tall|H) * p(Smelly = no\ |H)$$

$$p(H|New\ Instance) = 0.5 * \frac{1}{4} * \frac{4}{4} * \frac{2}{4} * \frac{3}{4} = 0.047$$

$$p(H|New\ Instance) > p(M|New\ Instance)$$

Hence the new instance belongs to Speices H

NAIVE BAYES CLASSIFIER

EXAMPLE - 3

Example No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

$$p(Yes) = \frac{5}{10} = 0.5$$

$$p(No) = \frac{5}{10} = 0.5$$

Color	Yes	No
Red	3/5	2/5
Yellow	2/5	3/5

Type	Yes	No
Sports	4/5	2/5
SUV	1/5	3/5

Origin	Yes	No
Domestic	2/5	3/5
Imported	3/5	2/5

$$New\ Instance = (Red, SUV, Domestic)$$ **No**

$$P(Yes|New\ Instance) = p(Yes) * P(Color = Red|Yes) * P(Type = SUV|Yes) * P(Origin = Domestic|Yes)$$

$$P(Yes|New\ Instance) = \frac{5}{10} * \frac{3}{5} * \frac{1}{5} * \frac{2}{5} = \frac{3}{125} = 0.024$$

$$P(No|New\ Instance) = p(No) * P(Color = Red|No) * P(Type = SUV|No) * P(Origin = Domestic|No)$$

$$P(No|New\ Instance) = \frac{5}{10} * \frac{2}{5} * \frac{3}{5} * \frac{3}{5} = \frac{9}{125} = 0.072$$

$$P(No|New\ Instance) > P(Yes|New\ Instance)$$

NAIVE BAYES CLASSIFIER EXAMPLE – 4

Consider a football game between two rival teams, say team A and team B. Suppose team A wins 65% of the time and team B wins the remaining matches. Among the games won by team A, only 35% of them comes from playing at team B's football field. On the other hand, 75% of the victories for team B are obtained while playing at home.

1. If team B is to host the next match between the two teams, what is the probability that it will emerge as the winner?
2. If team B is to host the next match between the two teams, who will emerge as the winner?

Solution:

Y – Winning football match
X – Hosting football match

Probability that team A wins is $P(Y_A) = 0.65$.

Probability that team B wins is $P(Y_B) = 1 - P(Y_A) = 0.35$

Probability that team B hosted the match it had won is $P(X_B | Y_B) = 0.75$.

Probability that team B hosted the match won by team A is $P(X_B | Y_A) = 0.35$.

NAIVE BAYES CLASSIFIER EXAMPLE – 4

Suggested: Big Data Analytics

1. If team B is to host the next match between the two teams, what is the probability that it will emerge as the winner?

Solution:

$$P(Y_B|X_B) = \frac{P(X_B|Y_B) \times P(Y_B)}{P(X_B)}$$

$$= \frac{P(X_B|Y_B) \times P(Y_B)}{P(X_B|Y_A)P(Y_A) + P(X_B|Y_B)P(Y_B)}$$

$$= \frac{0.75 \times 0.35}{(0.35 \times 0.65 + 0.75 \times 0.35)}$$

$$= 0.5357$$

Probability that team A wins is $P(Y_A) = 0.65$.

Probability that team B wins is $P(Y_B) = 1 - P(Y_A) = 0.35$

Probability that team B hosted the match it had won is $P(X_B|Y_B) = 0.75$.

Probability that team B hosted the match won by team A is $P(X_B|Y_A) = 0.35$.

NAIVE BAYES CLASSIFIER EXAMPLE – 4

2. If team B is to host the next match between the two teams, who will emerge as the winner?

Solution:

$$P(Y_A|X_B) = \frac{P(X_B|Y_A) \times P(Y_A)}{P(X_B)}$$

$$= \frac{P(X_B|Y_A) \times P(Y_A)}{P(X_B|Y_A)P(Y_A) + P(X_B|Y_B)P(Y_B)}$$

$$= \frac{0.35 \times 0.65}{(0.35 \times 0.65 + 0.75 \times 0.35)}$$

$$= 0.4642$$

Probability that team A wins is $P(Y_A) = 0.65$.

Probability that team B wins is $P(Y_B) = 1 - P(Y_A) = 0.35$

Probability that team B hosted the match it had won is $P(X_B|Y_B) = 0.75$.

Probability that team B hosted the match won by team A is $P(X_B|Y_A) = 0.35$.

$$P(Y_B|X_B) = 0.5357$$

Naïve Bayes Classifier – Solved Example 5

Record	A	B	C	Class
1	0	0	0	+
2	0	0	1	−
3	0	1	1	−
4	0	1	1	−
5	0	0	1	+
6	1	0	1	+
7	1	0	1	−
8	1	0	1	−
9	1	1	1	+
10	1	0	1	+

How to Compute

- Conditional Probabilities
- Predict the Label for the new instance

Naïve Bayes Classifier – Solved Example 5

Consider the dataset given in the table and

1. Estimate the conditional probabilities of P(A|+), P(B|+), P(C|+), P(A|-), P(B|-) and P(C|-)

2. Use the conditional probability estimates and predict the class label for the test sample P(A=0, B=1, C=0)

Naïve Bayes Classifier – Solved Example 5

Record	A	B	C	Class
1	0	0	0	+
2	0	0	1	−
3	0	1	1	−
4	0	1	1	−
5	0	0	1	+
6	1	0	1	+
7	1	0	1	−
8	1	0	1	−
9	1	1	1	+
10	1	0	1	+

A	0	1
+	2/5	3/5
-	3/5	2/5

B	0	1
+	4/5	1/5
-	3/5	2/5

C	0	1
+	1/5	4/5
-	0/5	5/5

$$New \Rightarrow P(A=0, B=1, C=0)$$

$$P(+|New) = \frac{P(+) * P(A=0|+) * P(B=1|+) * P(C=0|+)}{\cancel{P(A=0, B=1, C=0)}}$$

$$= 0.5 * \frac{2}{5} * \frac{1}{5} * \frac{1}{5} = 0.008$$

+

$$P(-|New) = \frac{P(-) * P(A=0|-) * P(B=1|-) * P(C=0|-)}{\cancel{P(A=0, B=1, C=0)}}$$

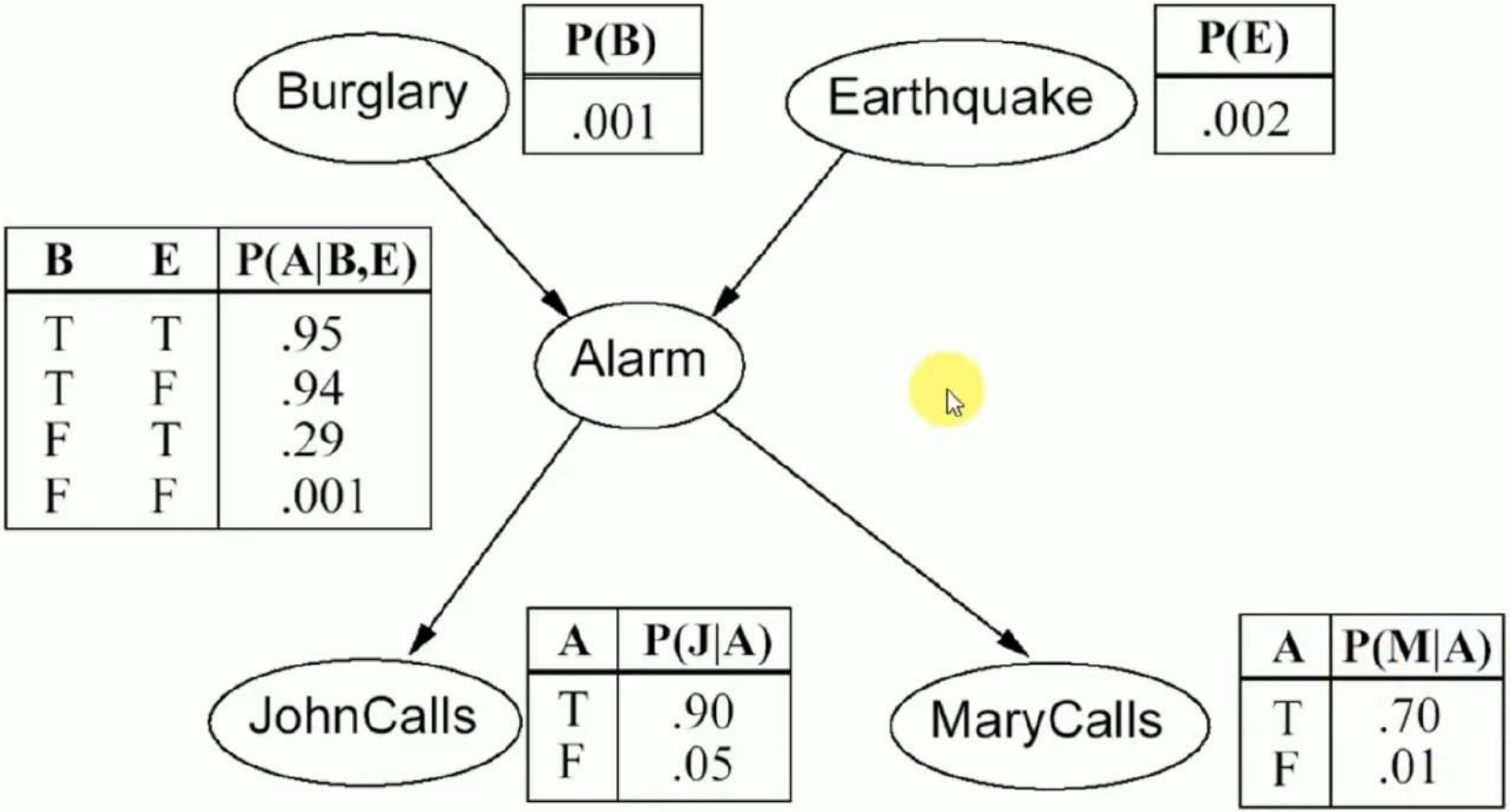
$$= 0.5 * \frac{3}{5} * \frac{2}{5} * \frac{0}{5} = 0.0$$

$$\boxed{P(+|New) > P(-|New)}$$

BAYESIAN BELIEF NETWORKS – EXAMPLE – 1

- You have a new burglar alarm installed at home.
- It is fairly reliable at detecting burglary, but also sometimes responds to minor earthquakes.
- You have two neighbors, John and Merry , who promised to call you at work when they hear the alarm.
- John always calls when he hears the alarm, but sometimes confuses telephone ringing with the alarm and calls too.
- Merry likes loud music and sometimes misses the alarm.
- Given the evidence of who has or has not called, we would like to estimate the probability of a burglary.

BAYESIAN BELIEF NETWORKS – EXAMPLE – 1



P(B)
.001

P(E)
.002

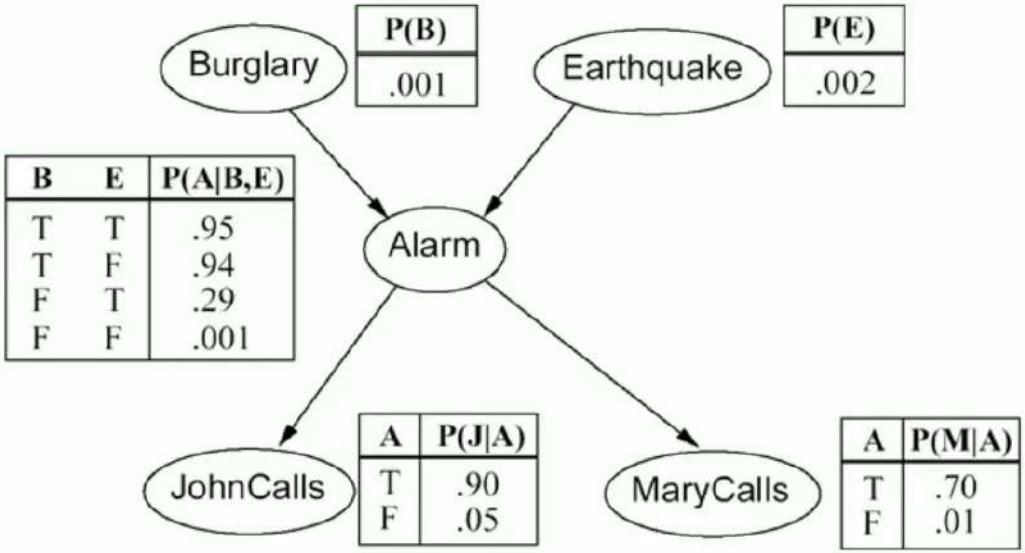
B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

BAYESIAN BELIEF NETWORKS – EXAMPLE – 1

1. What is the probability that the alarm has sounded but neither a burglary nor an earthquake has occurred, and both John and Merry call?



Solution:

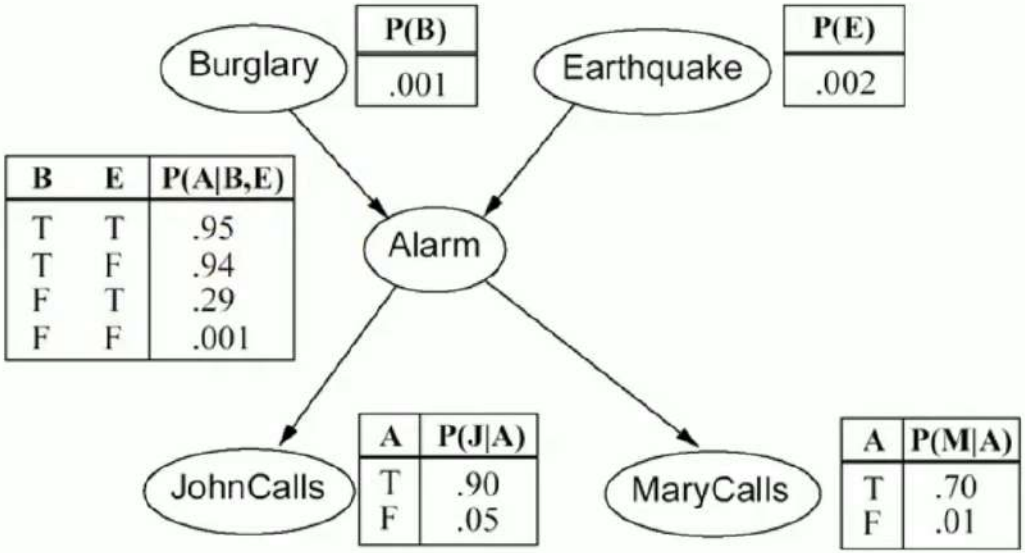
$$P(j \wedge m \wedge a \wedge \neg b \wedge \neg e) = P(j \mid a)\, P(m \mid a)\, P(a \mid \neg b, \neg e)\, P(\neg b)\, P(\neg e)$$

$$= 0.90 \times 0.70 \times 0.001 \times 0.999 \times 0.998$$

$$= 0.00062$$

BAYESIAN BELIEF NETWORKS – EXAMPLE – 1

2. What is the probability that John call?



Solution:

$P(j) = P(j \mid a)\, P(a) + P(j \mid \neg a)\, P(\neg a)$

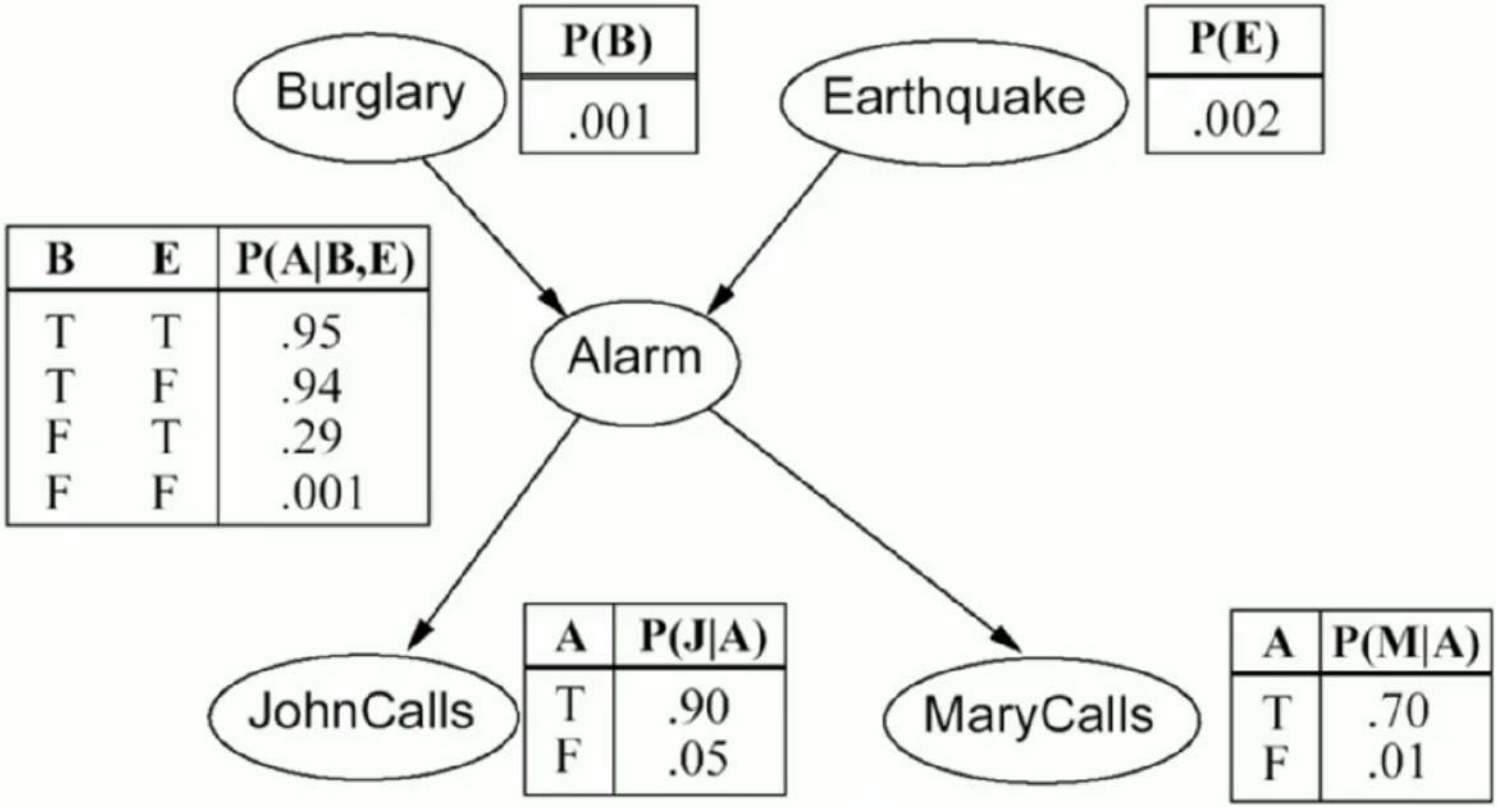
$= P(j|a)\{P(a|b,e)*P(b,e)+P(a|\neg b,e)*P(\neg b,e)+P(a|b,\neg e)*P(b,\neg e)+P(a|\neg b,\neg e)*P(\neg b,\neg e)\}$

$+ P(j|\neg a)\{P(\neg a|b,e)*P(b,e)+P(\neg a|\neg b,e)* P(\neg b,e)+P(\neg a|b,\neg e)* P(b,\neg e)+P(\neg a|\neg b, \neg e)*$

$P(\neg b, \neg e)\}$

$= 0.90 * 0.00252 + 0.05 * 0.9974 = 0.0521$

BAYESIAN BELIEF NETWORKS – EXAMPLE – 2



3. What is the probability that there is a burglary given that John and Merry calls?

BAYESIAN BELIEF NETWORKS – EXAMPLE – 2

- Suppose, we are given for the evidence variables $E_1,\ldots,E_m$, their values $e_1,\ldots,e_m$, and we want to predict whether the query variable X has the value x or not.
- For this we compute and compare the following:

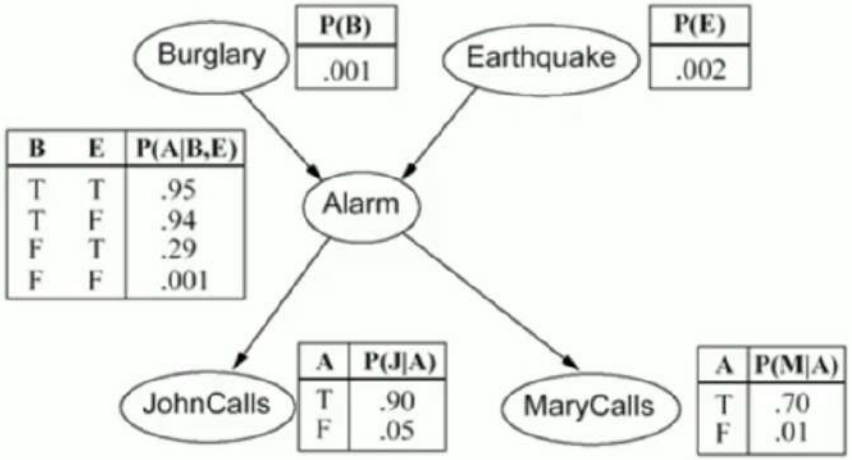
$$P(x \mid e_1,\ldots,e_m) = \frac{P(x,e_1,\ldots,e_m)}{P(e_1,\ldots,e_m)} = \alpha P(x,e_1,\ldots,e_m)$$

$$P(\neg x \mid e_1,\ldots,e_m) = \frac{P(\neg x,e_1,\ldots,e_m)}{P(e_1,\ldots,e_m)} = \alpha P(\neg x,e_1,\ldots,e_m)$$

$$\propto = \frac{1}{(P(x,e_1,\ldots em) + P(\neg x,e_1,\ldots em))}$$

BAYESIAN BELIEF NETWORKS – EXAMPLE – 2

3. What is the probability that there is a burglary given that John and Merry calls?



$$P(b \mid j,m) = \alpha\, P(b) \sum_a P(j|a)\, P(m|a) \sum_e P(a|b,e)P(e)$$

$$= \alpha\, P(b) \sum_a P(j|a)\, P(m|a)\, \{P(a|b,e)P(e) + P(a|b,\neg e)P(\neg e)\}$$

$$= \alpha\, P(b)\, [\, P(j|a)P(m|a)\, \{P(a|b,e)P(e) + P(a|b,\neg e)P(\neg e)\, \}$$

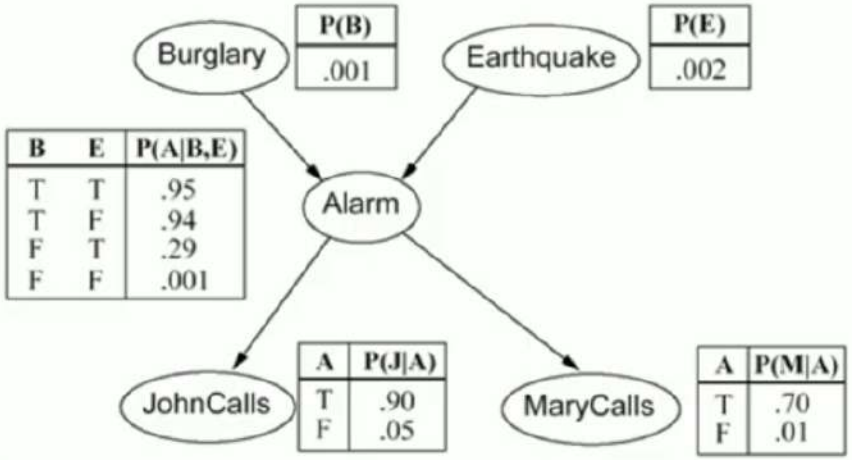
$$+ P(j|\neg a)P(m|\neg a)\, \{P(\neg a|b,e)P(e) + P(\neg a|b,\neg e)P(\neg e)\}]$$

$$= \alpha * .001*(.9*.7*(.95*.002 + .94*.998) + .05*.01*(.05*.002 + .71*.998)\,)$$

$$= \alpha * .00059$$

BAYESIAN BELIEF NETWORKS – EXAMPLE – 2

3. What is the probability that there is a burglary given that John and Merry calls?



$$P(\neg b \mid j,m) = \alpha\, P(\neg b) \sum_a P(j|a)P(m|a) \sum_e P(a|\neg b,e)P(e)$$

$$= \alpha\, P(\neg b) \sum_a P(j|a)P(m|a)\ \{ P(a|\neg b,e)P(e) + P(a|\neg b,\neg e)P(\neg e)\}$$

$$= \alpha\, P(\neg b)\ [\ P(j|a)P(m|a)\ \{ P(a|\neg b,e)P(e) + P(a|\neg b,\neg e)P(\neg e)\ \}$$

$$+ P(j|\neg a)P(m|\neg a)\ \{ P(\neg a|\neg b,e)P(e) + P(\neg a|\neg b,\neg e)P(\neg e)\ \}]$$

$$= \alpha * .999*(.9*.7*(.29*.002 + .001*.998) + .05*.01*(.71*.002 + .999*.998))$$

$$= \alpha * .0015$$

BAYESIAN BELIEF NETWORKS – EXAMPLE – 2

3. What is the probability that there is a burglary given that John and Merry calls?

$$\alpha = \frac{1}{(P(b,j,m) + P(\neg b,j,m))}$$

$$\alpha = \frac{1}{(.00059 + .0015)}$$

$$= 478.5$$

$$P(b \mid j, m) = \propto * P(b, j, m)$$

$$= 478.5 * .00059$$

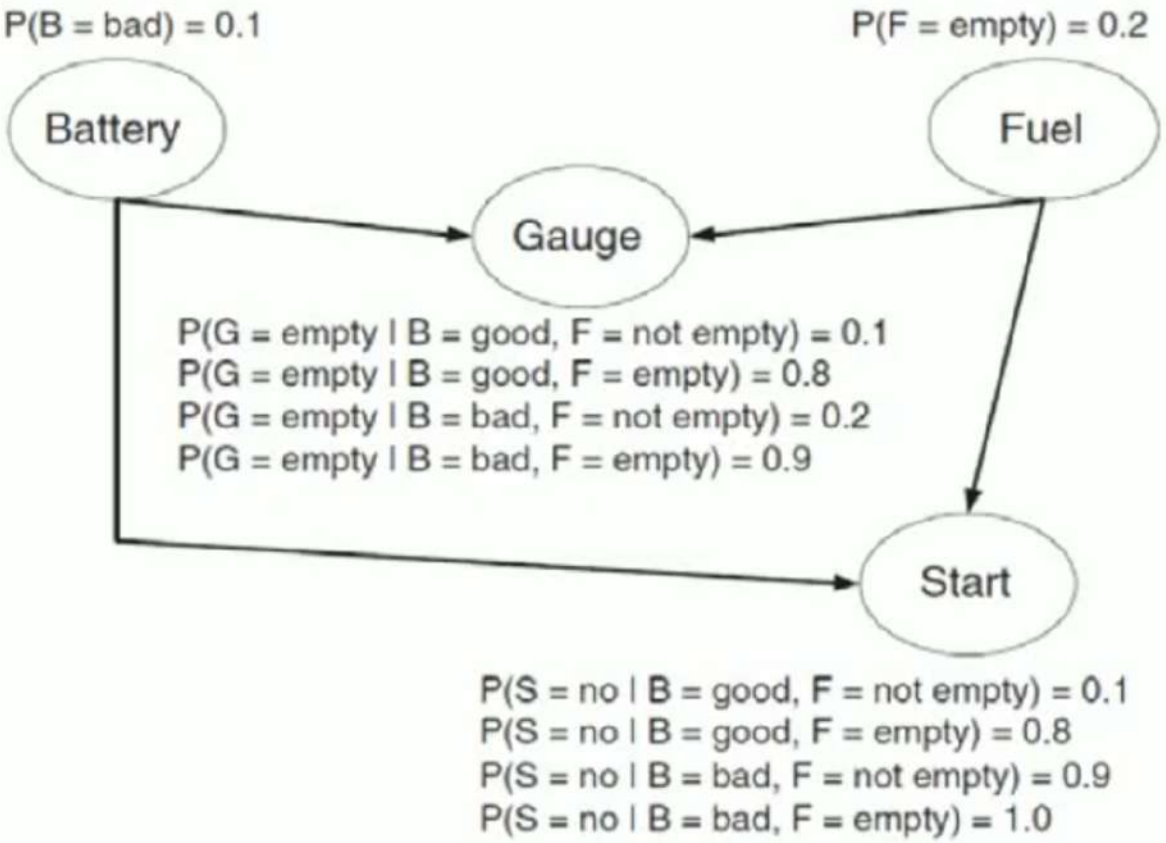
$$= 0.28$$

$$P(\neg b \mid j,m) = \propto * P(\neg b, j, m)$$

$$= 478.5 * .0015$$

$$= 0.72$$

Bayesian Belief Network – Solved Example 3



1. P(B=Good, F=Empty, G=Empty, S=Yes)
2. P(B=Bad, F=Empty, G=Not Empty, S=No)
3. Given the battery is bad, Compute the probability that the car will start

Bayesian Belief Network – Solved Example 3

1. **P(B=Good, F=Empty, G=Empty, S=Yes)**

$$
\begin{aligned}
& P(B = good, F = empty, G = empty, S = yes) \\
= \; & P(B = good) \times P(F = empty) \times P(G = empty | B = good, F = empty) \\
& \times P(S = yes | B = good, F = empty) \\
= \; & 0.9 \times 0.2 \times 0.8 \times 0.2 = 0.0288.
\end{aligned}
$$

Bayesian Belief Network – Solved Example 3

2. P(B=Bad, F=Empty, G=Not Empty, S=No)

$$
\begin{aligned}
& P(B = bad, F = empty, G = not\ empty, S = no) \\
= \quad & P(B = bad) \times P(F = empty) \times P(G = not\ empty | B = bad, F = empty) \\
& \times P(S = no | B = bad, F = empty) \\
= \quad & 0.1 \times 0.2 \times 0.1 \times 1.0 = 0.002.
\end{aligned}
$$

Bayesian Belief Network – Solved Example 3

3. **Given the battery is bad, Compute the probability that the car will start.**

$$p(S = yes \mid B = bad)$$

$$= p(S = yes, B = bad) \Big/ p(B = bad)$$

$$= \sum_{\alpha} p(S = yes, B = bad, F = \alpha) \Big/ p(B = bad)$$

$$= \sum_{\alpha} p(S = yes \mid B = bad, F = \alpha) \times p(B = bad, F = \alpha) \Big/ p(B = bad)$$

$$= \sum_{\alpha} p(S = yes \mid B = bad, F = \alpha) \times p(B = bad) \times p(\mid F = \alpha) \Big/ p(B = bad)$$

$$= \sum_{\alpha} p(S = yes \mid B = bad, F = \alpha) \times p(\mid F = \alpha)$$

$$= P(S = Yes \mid B = bad, F = Empty) * P(F = Empty) +$$

$$P(S = Yes \mid B = bad, F = Non - Empty) * P(F = No - Empty)$$

$$= (0 * 0.2) + (0.1 * 0.8)$$

$$= 0.08$$

KNN Classifier Solved Example - 1

Sepal Length	Sepal Width	Species
5.3	3.7	Setosa
5.1	3.8	Setosa
7.2	3.0	Virginica
5.4	3.4	Setosa
5.1	3.3	Setosa
5.4	3.9	Setosa
7.4	2.8	Virginica
6.1	2.8	Verscicolor
7.3	2.9	Virginica
6.0	2.7	Verscicolor
5.8	2.8	Virginica
6.3	2.3	Verscicolor
5.1	2.5	Verscicolor
6.3	2.5	Verscicolor
5.5	2.4	Verscicolor

Sepal Length	Sepal Width	Species
5.2	3.1	?

Step 1: Find Distance

$$\text{Distance (Sepal Length, Sepal Width)} = \sqrt{(x-a)^2 + (y-b)^2}$$

$$\text{Distance (Sepal Length, Sepal Width)} = \sqrt{(5.2-5.3)^2 + (3.1-3.7)^2}$$

$$\text{Distance (Sepal Length, Sepal Width)} = 0.608$$

Sepal Length	Sepal Width	Species	Distance
5.3	3.7	Setosa	0.608

KNN Classifier Solved Example - 1

Sepal Length	Sepal Width	Species	Distance	Rank
5.3	3.7	Setosa	0.608	3
5.1	3.8	Setosa	0.707	6
7.2	3.0	Virginica	2.002	13
5.4	3.4	Setosa	0.36	2
5.1	3.3	Setosa	0.22	1
5.4	3.9	Setosa	0.82	8
7.4	2.8	Virginica	2.22	15
6.1	2.8	Verscicolor	0.94	10
7.3	2.9	Virginica	2.1	14
6.0	2.7	Verscicolor	0.89	9
5.8	2.8	Virginica	0.67	5
6.3	2.3	Verscicolor	1.36	12
5.1	2.5	Verscicolor	0.60	4
6.3	2.5	Verscicolor	1.25	11
5.5	2.4	Verscicolor	0.75	7

Step 3: Find the Nearest Neighbor

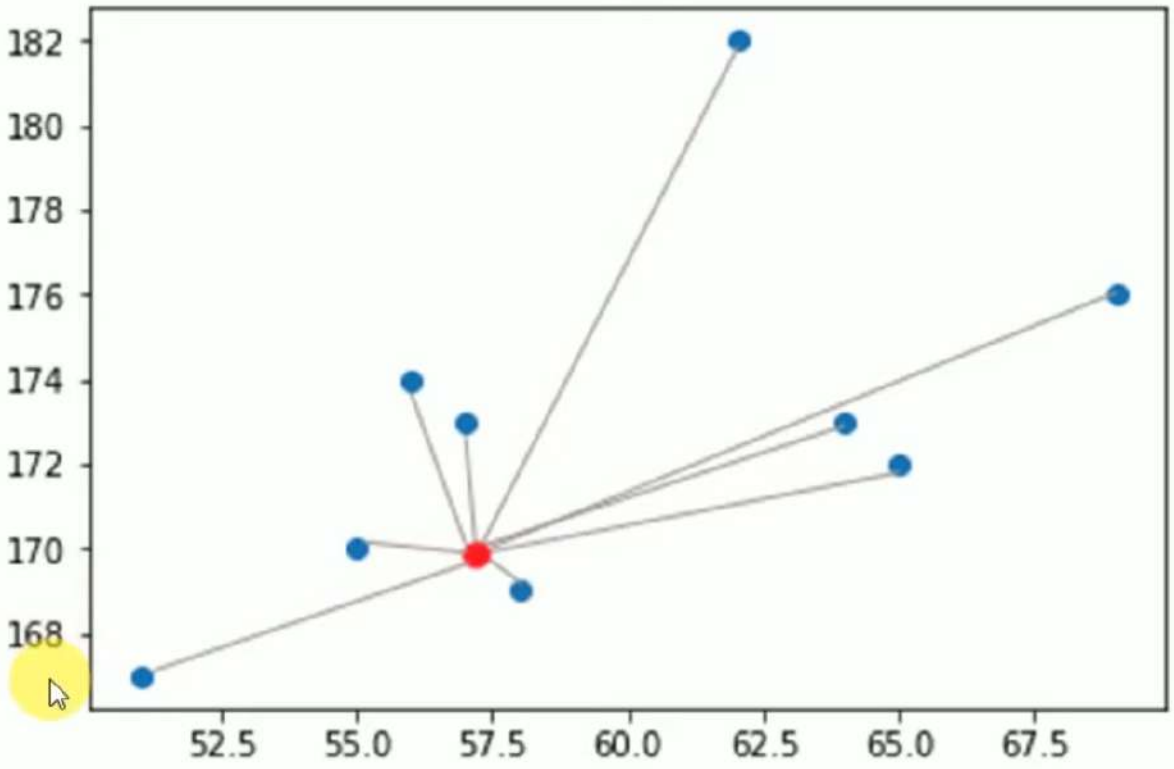
If k = 1 – Setosa

If k = 2 – Setosa

If k = 5 – Setosa

K Nearest Neighbor Algorithm - Solved Example - 2

Height (CM)	Weight (KG)	Class
167	51	Underweight
182	62	Normal
176	69	Normal
173	64	Normal
172	65	Normal
174	56	Underweight
169	58	Normal
173	57	Normal
170	55	Normal
170	57	?



K Nearest Neighbor Algorithm - Solved Example - 2

Height (CM)	Weight (KG)	Class
167	51	Underweight
182	62	Normal
176	69	Normal
173	64	Normal
172	65	Normal
174	56	Underweight
169	58	Normal
173	57	Normal
170	55	Normal
170	57	?

THE DISTANCE FORMULA

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

$$d_1 = \sqrt{(170 - 167)^2 + (57 - 51)^2}$$

$$d_2 = \sqrt{(170 - 180)^2 + (57 - 62)^2}$$

K Nearest Neighbor Algorithm - Solved Example - 2

Height (CM)	Weight (KG)	Class	Distance	Rank
169	58	Normal	1.4	1
170	55	Normal	2	2
173	57	Normal	3	3
174	56	Underweight	4.1	4
167	51	Underweight	6.7	5
173	64	Normal	7.6	6
172	65	Normal	8.2	7
182	62	Normal	13	8
176	69	Normal	13.4	9
170	57	?		

- If K=1, Normal
- If K=2, Normal
- If K=3, Normal
- If K=4, Normal
- If K=5, Normal

K-Nearest Neighbors Algorithm

	Pepper	Ginger	Chilly	Liked
A	True	True	True	False
B	True	False	False	True
C	False	True	True	False
D	False	True	False	True
E	True	False	False	True

Solved Example 3

Subscribe to Mahesh Huddar

Visit: vtupulse.com

K-Nearest Neighbors Algorithm Solved Example - 3

- The "Restaurant A" sells burger with optional flavors: Pepper, Ginger and Chilly.
- Every day this week you have tried a burger (A to E) and kept a record of which you liked.
- Using Hamming distance, show how the 3NN classifier with majority voting would classify

{ pepper: false, ginger: true, chilly : true}

K-Nearest Neighbors Algorithm Solved Example - 3

	Pepper	Ginger	Chilly	Liked
A	True	True	True	False
B	True	False	False	True
C	False	True	True	False
D	False	True	False	True
E	True	False	False	True

New Example - Q: pepper: false, ginger: true, chilly : true

K-Nearest Neighbors Algorithm Solved Example - 3

- But How to calculate the distance for attributes with nominal or categorical values.
- Here we can use Hamming distance to find the distance between the categorical values.
- Let x1 and x2 are the attribute values of two instances.
- Then, in hamming distance, if the categorical values are same or matching that is x1 is same as x2 then distance is 0, otherwise 1.
- For example,
- If value of **x1 is blue** and **x2 is also blue** then the distance between x1 and x2 is **0**.
- If value of **x1 is blue** and **x2 is red** then the distance between x1 and x2 is **1**.

K-Nearest Neighbors Algorithm Solved Example - 3

	Pepper	Ginger	Chilly	Liked	Distance
A	True	True	True	False	1 + 0 + 0 = 1
B	True	False	False	True	1 + 1 + 1 = 3
C	False	True	True	False	0 + 0 + 0 = 0
D	False	True	False	True	0 + 0 + 1 = 1
E	True	False	False	True	1 + 1 + 1 = 3

New Example - Q: pepper: false, ginger: true, chilly : true

Use Hamming Distance and

find the distance from Query Example (Q) to training examples (A-E)

K-Nearest Neighbors Algorithm Solved Example - 3

	Pepper	Ginger	Chilly	Liked	Distance	3NN
A	True	True	True	False	1 + 0 + 0 = 1	2
B	True	False	False	True	1 + 1 + 1 = 3	
C	False	True	True	False	0 + 0 + 0 = 0	1
D	False	True	False	True	0 + 0 + 1 = 1	2
E	True	False	False	True	1 + 1 + 1 = 3	

New Example - Q: pepper: false, ginger: true, chilly : true