Play Tennis HW: Solution

Rao Vmuri

Training Examples

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	Weak	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Strong	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

Step1: Entropy of Set S

- 14 examples
- 9 + examples
- 5 – examples
- Entropy [9+, 5 -] = $- (9/14)\log_2 (9/14)$

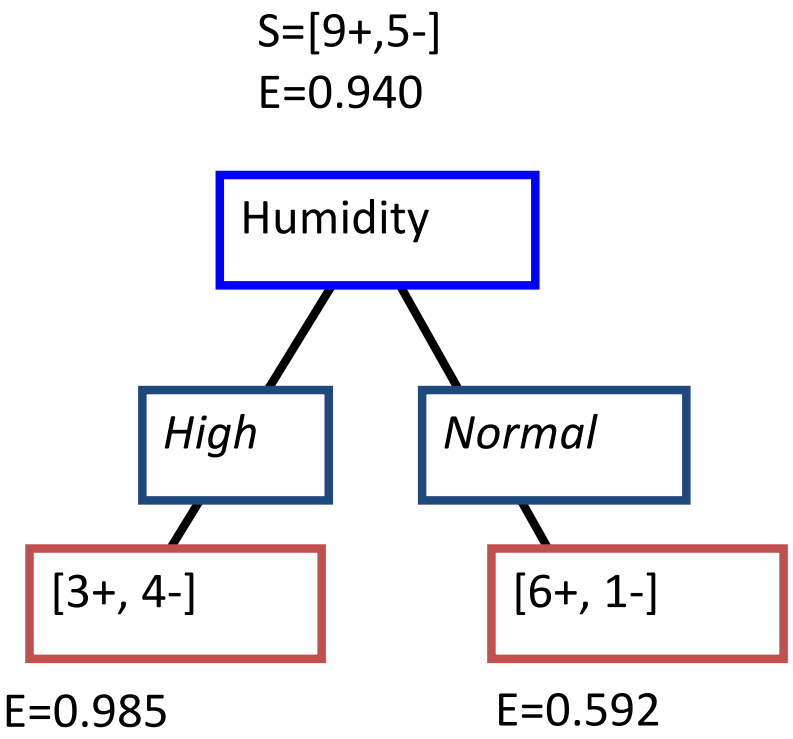
$$– (5/14) \log_2 (5/14)$$

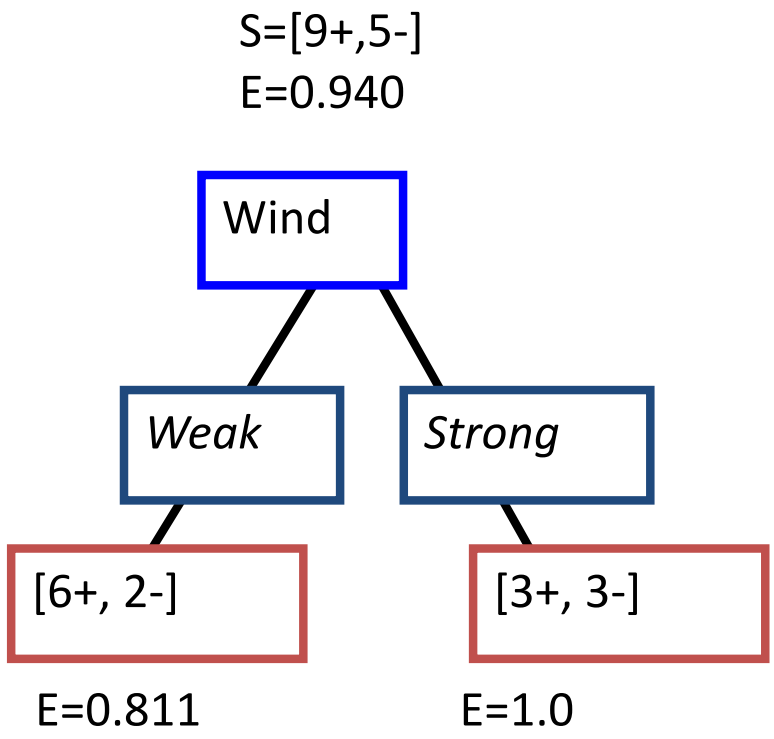
$$= 0.940$$

STEP 2: Next Attribute?

- We calculate the Gain in entropy for each of the 4 attributes: Outlook, Temperature, Humidity, Wind

Gain for Humidity & Wind





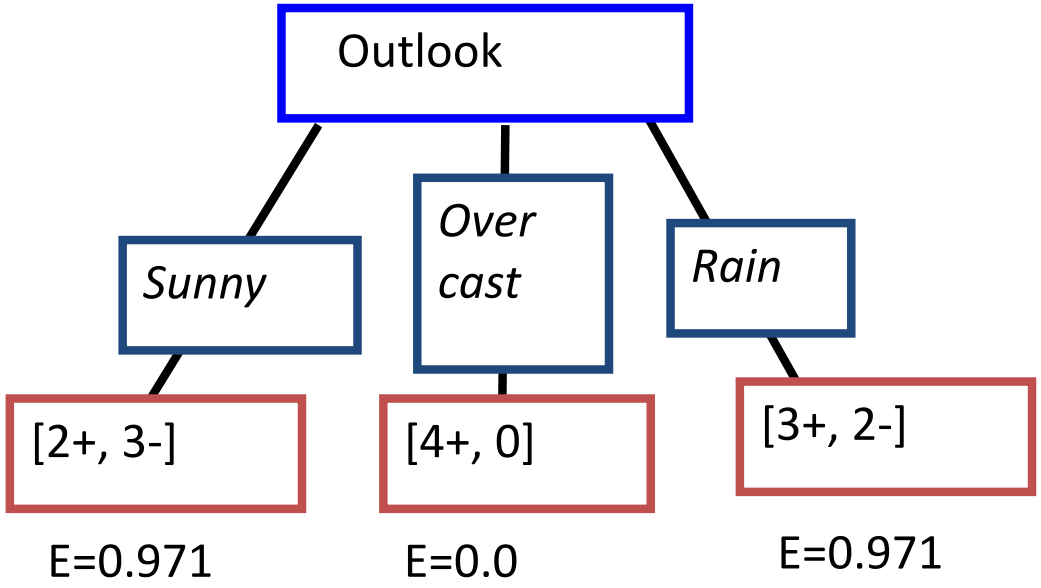
Gain(S,Humidity)
=0.940-(7/14)*0.985
– (7/14)*0.592
=0.151

Gain(S,Wind)
=0.940-(8/14)*0.811
– (6/14)*1.0
=0.048

Humidity provides greater info. gain than Wind, w.r.t target classification.

Gain for Outlook

S=[9+,5-]
E=0.940



Gain(S,Outlook)
=0.940-(5/14)*0.971
-(4/14)*0.0 – (5/14)*0.0971
=0.247

Outlook provides greater info. gain than humidity & Temp.

Info Gain for all 4 Attributes

The information gain values for the 4 attributes are:

- Gain(S,Outlook) =0.247
- Gain(S,Humidity) =0.151
- Gain(S,Wind) =0.048
- Gain(S,Temperature) =0.029

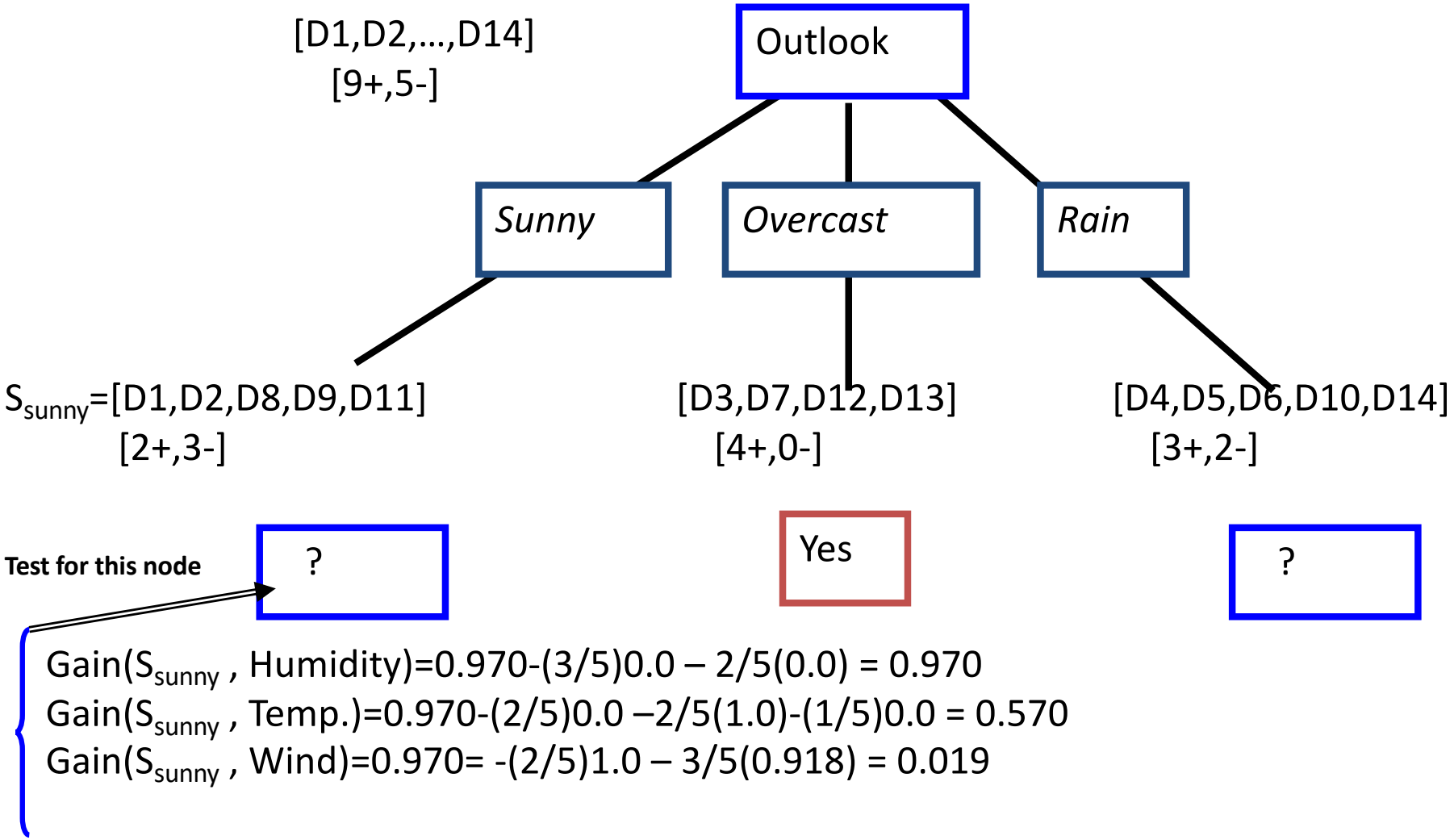
where S denotes the collection of training examples

So Choose Outlook for the first split

Note: $0\text{Log}_2 0 = 0$

ID3 Algorithm

Note: $0\text{Log}_2 0 = 0$



Entropy for S_{sunny}

- S_{sunny} = D1, D2, D8, D9, D11
- Entropy = $-(2/5)\log_2(2/5) - (3/5)\log_2(3/5)$

$$= 0.528 + 0.4422 = 0.970$$

- Home work:
  - Similarly, do the calculations for other sets and verify the solution shown on the next slide

ID3 Algorithm