의사결정나무ID3

Decision tree ID3

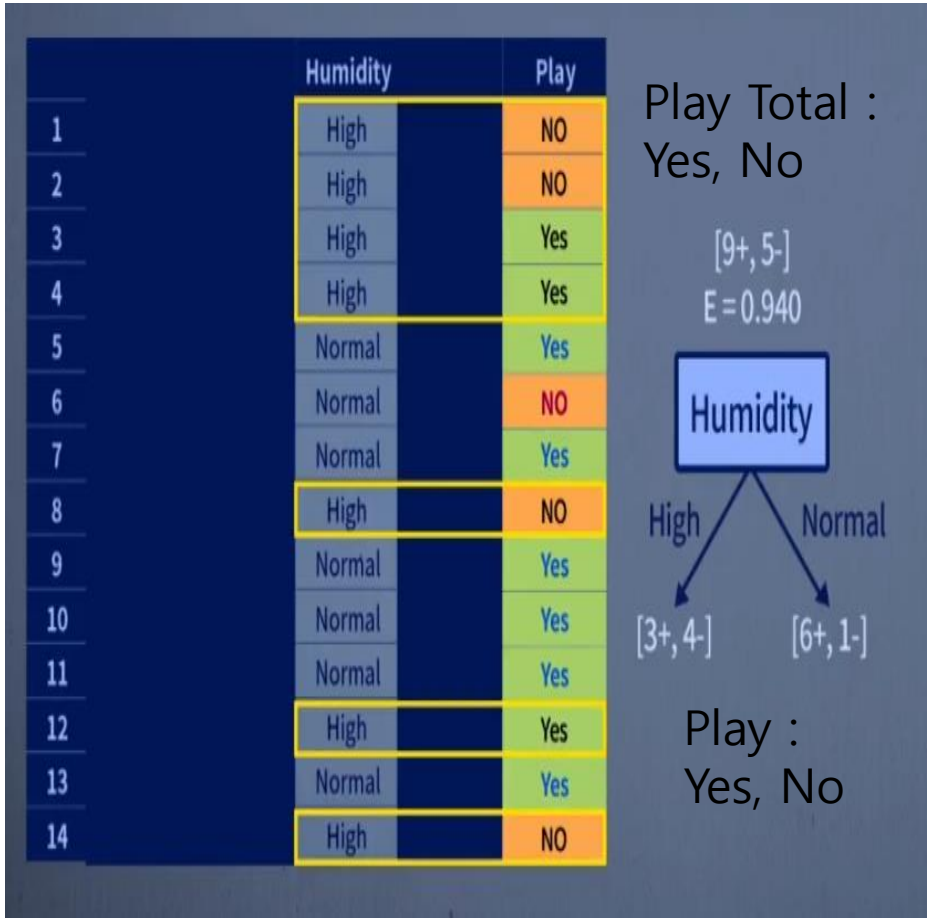
Which is the Most Predictive Input?

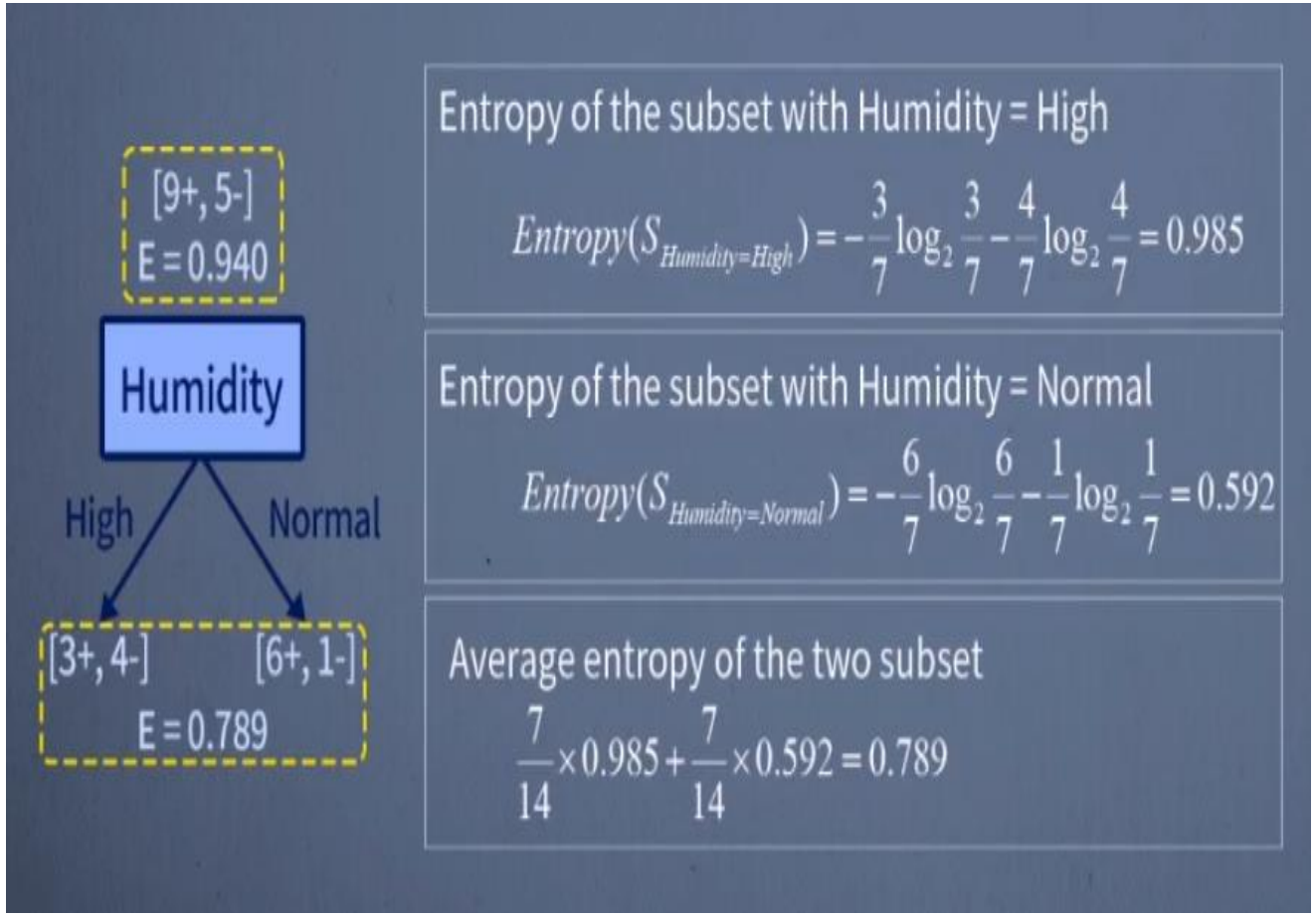
- Which input split Output into the most homogeneous subset?

	Outlook	Temp	Humidity	Wind	Play
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

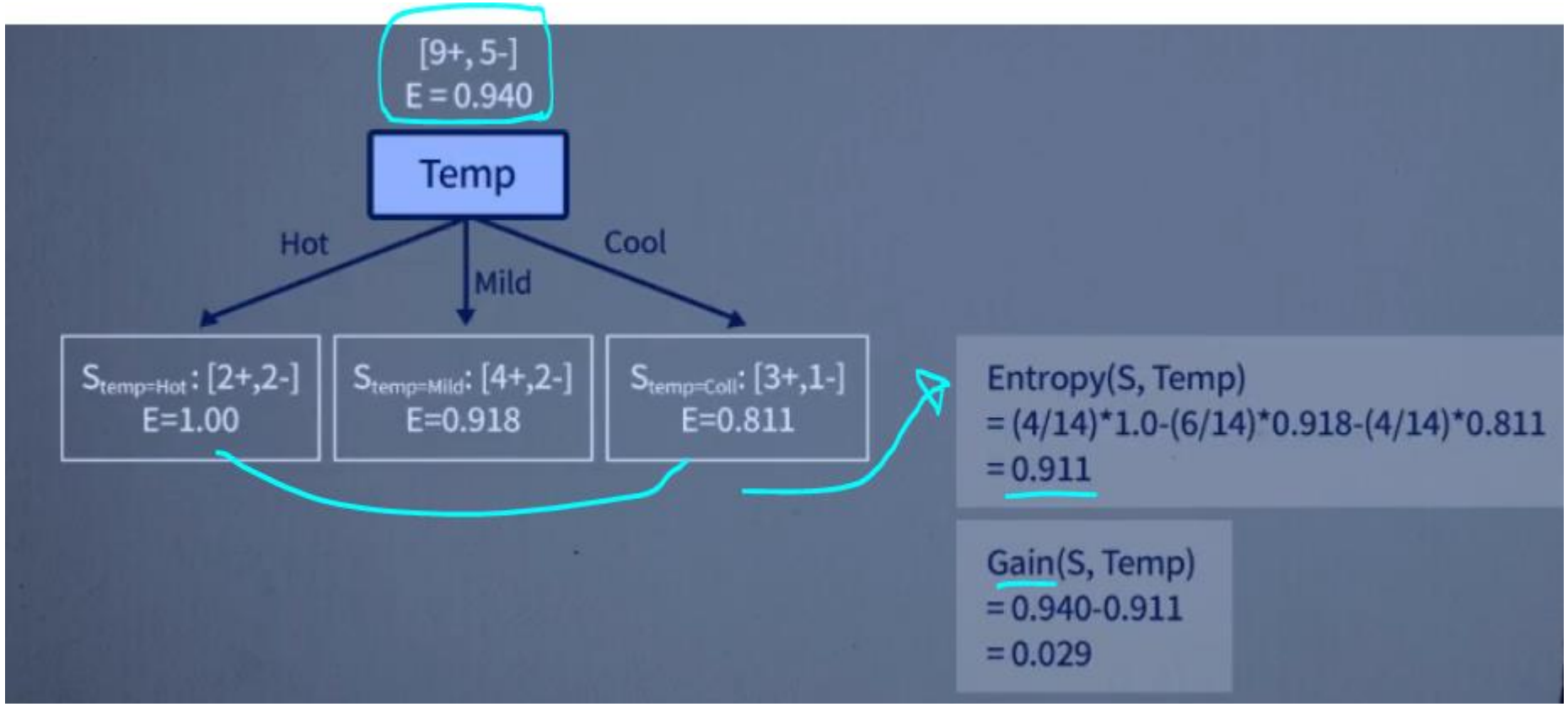
어떤 입력이 Play를 가장 homogeneous 하게 split하는가?

Decision tree ID3





Decision tree ID3

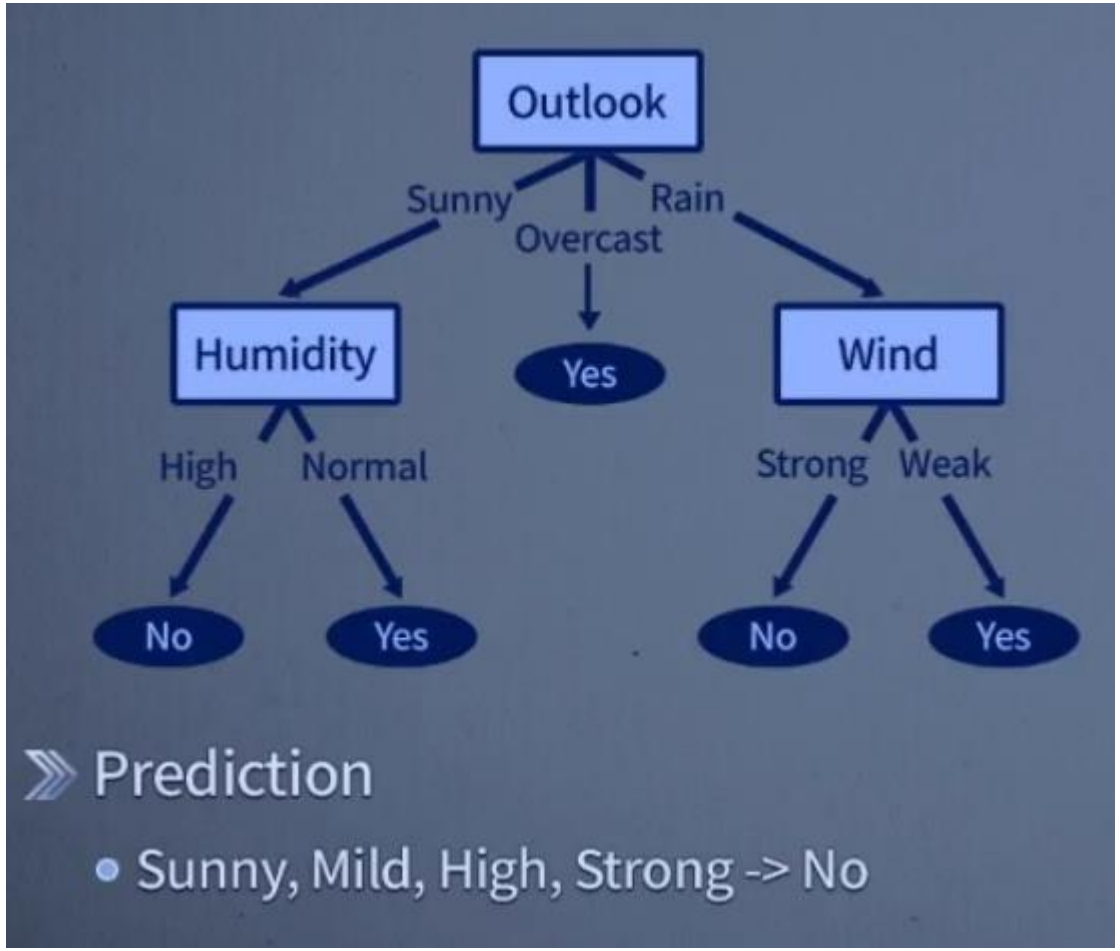


Decision tree ID3

Gain(S, Outlook) = 0.246
Gain(S, Humidity) = 0.151
Gain(S, Wind) = 0.048
Gain(S, Temp) = 0.029

	Outlook	Temp	Humidity	Wind	Play
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

Decision tree ID3



If Outlook = Sunny & Humidity = High
then Tennis = No

If Outlook = Sunny & Humidity = Normal
then Tennis = Yes

If Outlook = Overcast
then Tennis = Yes

If Outlook = Rain & Wind = Strong
then Tennis = No

If Outlook = Sunny & Wind = Weak
then Tennis = Yes

Decision tree ID3

- 어떤 것을 기준으로 나누느냐에 따라 Decision tree가 여러 개 나올 수 있다.
- 예측 정확도는 모두 100%이지만, Node수가 가장 적을수록 더 좋은 Decision tree이다.(모델의 복잡도가 낮다.)
- ID3는 가장 좋은 Decision tree는 고르지 못함(Greedy Algorithm을 사용하기 때문)