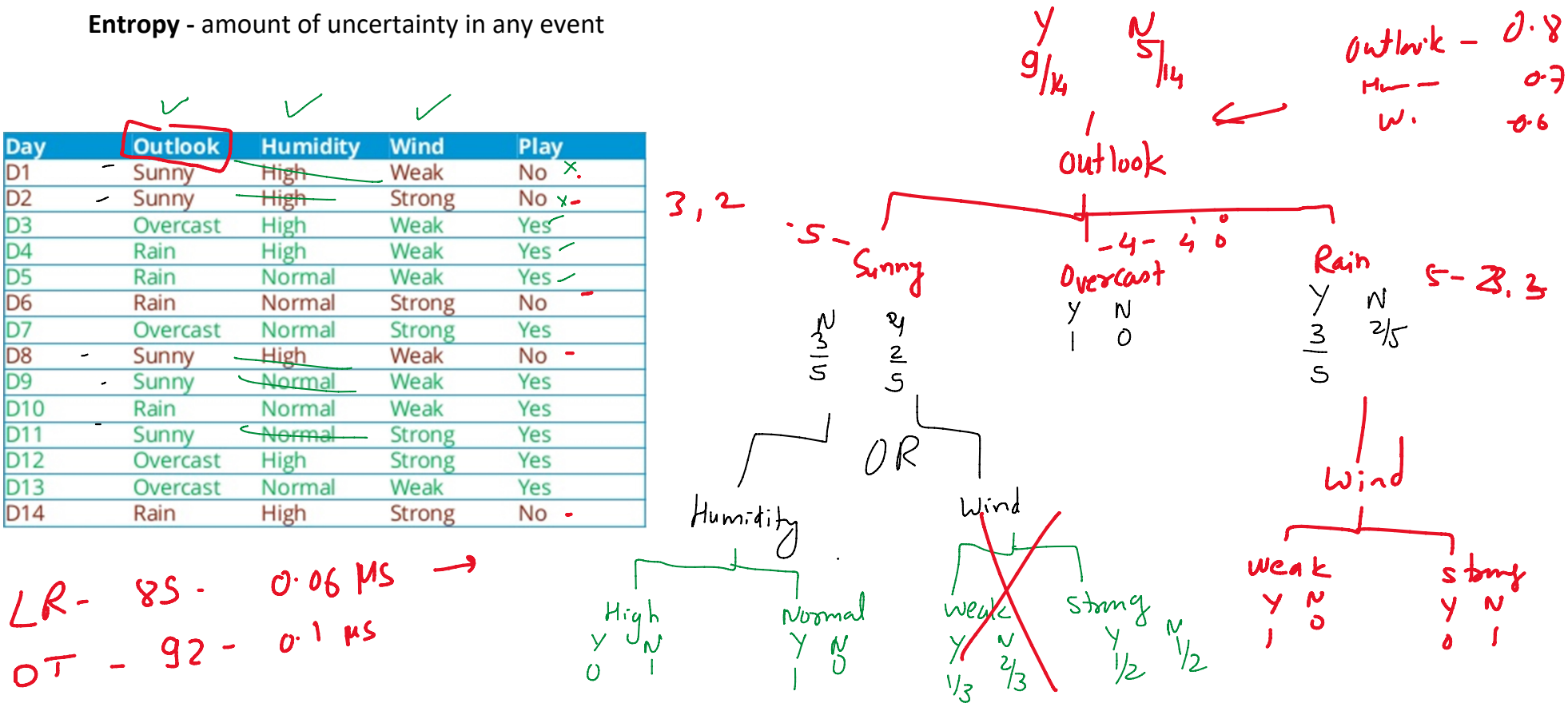
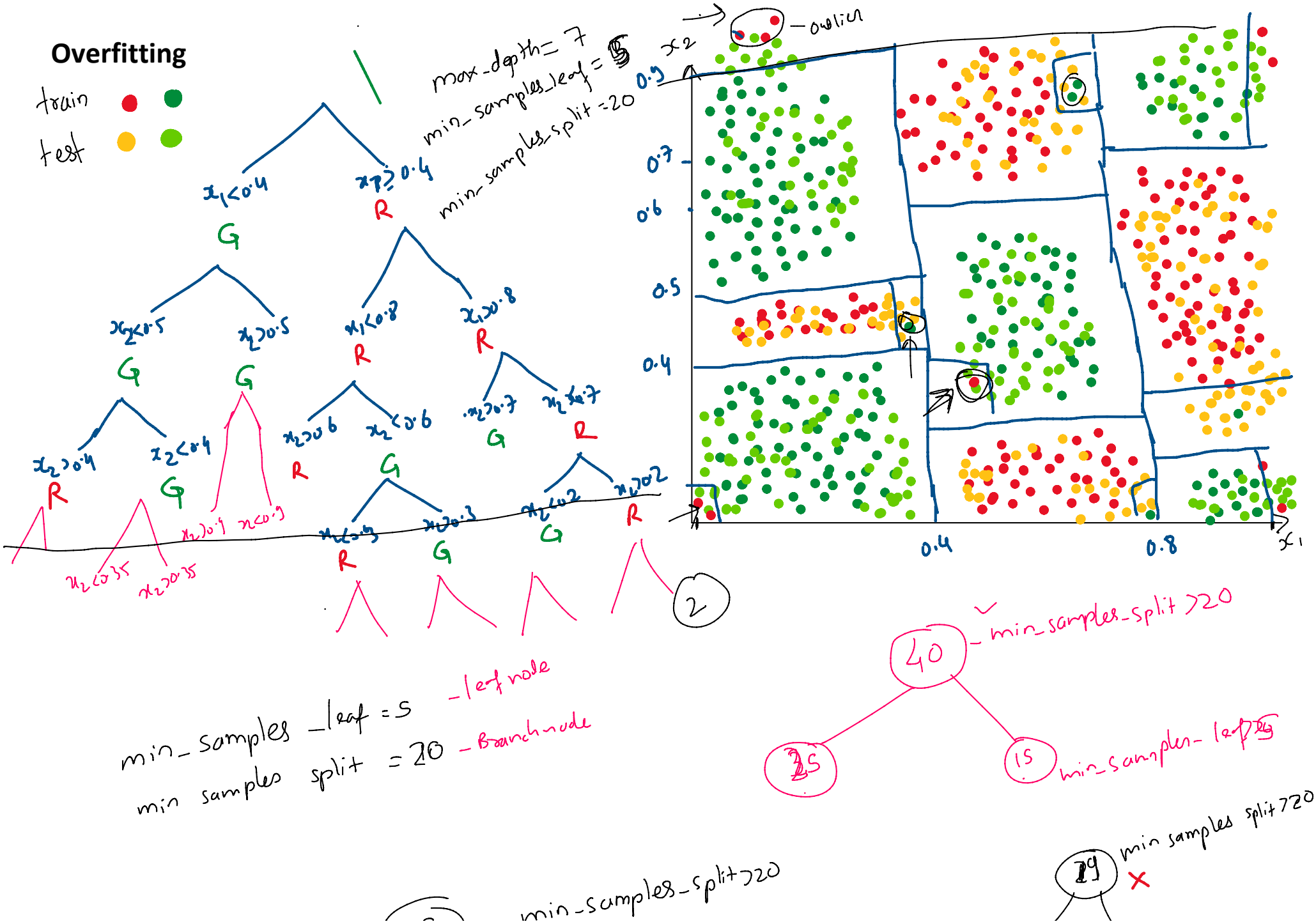
Decision Tree

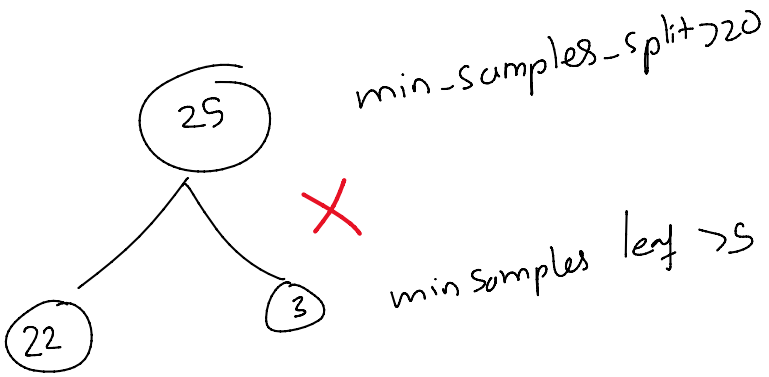
- It is used for both classification and regression problems
- In case of classification problems, decision tree can also be used for multiclass classification
- Decision Tree algorithm is a non linear algorithm, it can be used and provide good results even in case of features not having good linear relationship with the target attribute.
- Decision Tree algorithms are explainable, interpretable models - white box models

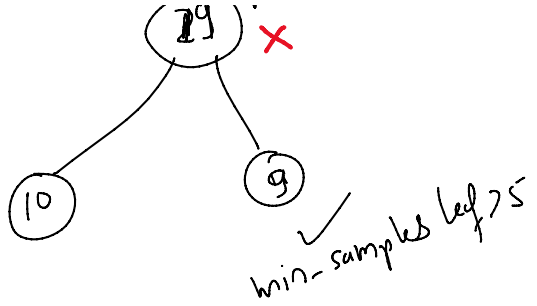
Entropy - amount of uncertainty in any event

Day	Outlook	Humidity	Wind	Play
D1	Sunny	High	Weak	No
D2	Sunny	High	Strong	No
D3	Overcast	High	Weak	Yes
D4	Rain	High	Weak	Yes
D5	Rain	Normal	Weak	Yes
D6	Rain	Normal	Strong	No
D7	Overcast	Normal	Strong	Yes
D8	Sunny	High	Weak	No
D9	Sunny	Normal	Weak	Yes
D10	Rain	Normal	Weak	Yes
D11	Sunny	Normal	Strong	Yes
D12	Overcast	High	Strong	Yes
D13	Overcast	Normal	Weak	Yes
D14	Rain	High	Strong	No




Overfitting
train
test
max_depth= 7
min_samples_leaf = 5
min_samples_split =20
x1<0.4
x1≥0.4
G
R
x2<0.5
x2≥0.5
x1<0.8
x1≥0.8
x2≥0.4
x2<0.4
x2≥0.6
x2<0.6
x2≥0.7
x2<0.7
x2<0.35
x2≥0.35
x2≥0.9
x<0.9
x2<0.3
x2≥0.3
x2<0.2
x2≥0.2
2
outlier
x2
x1
0.9
0.7
0.6
0.5
0.4
0.4
0.8
min_samples_leaf =5
min samples split =20
- leaf node
- Branch node
40
25
15
min_samples_split>20
min_samples_leaf
min_samples_split>20
19
min samples split>20
×


25
22
3
min_samples_split>20
min samples leaf >5


19
10
9
min_samples leaf>5