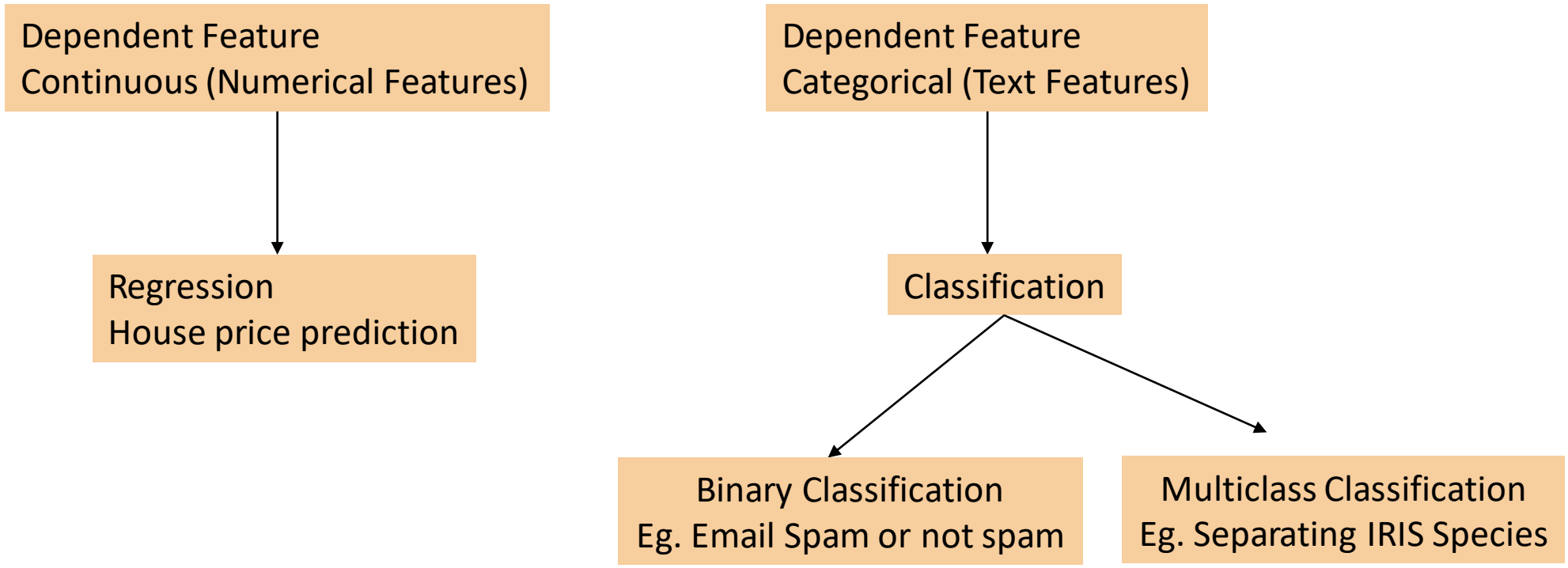
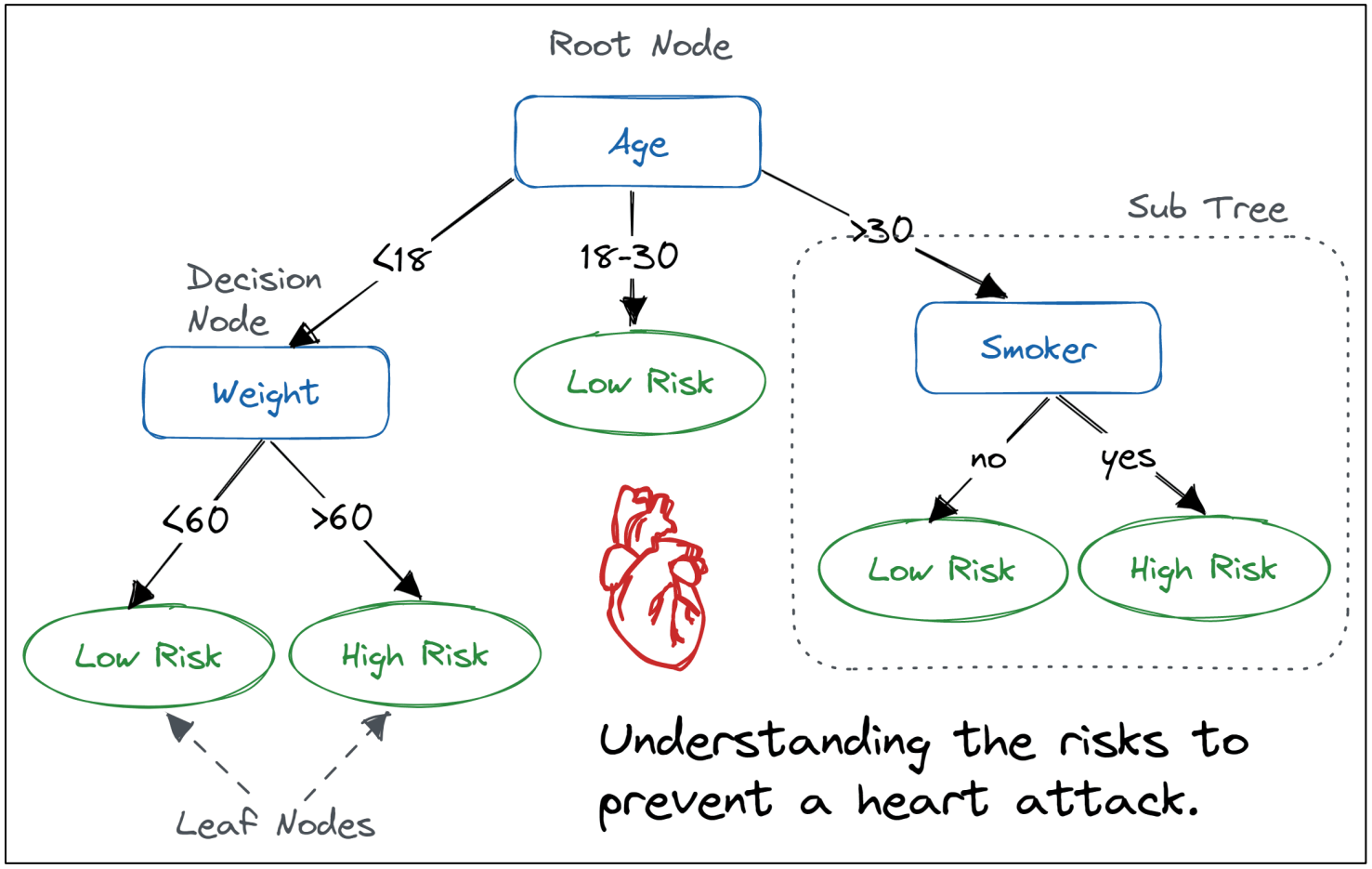
Decision Trees Classification

UTKARSH GAIKWAD

Classification vs Regression



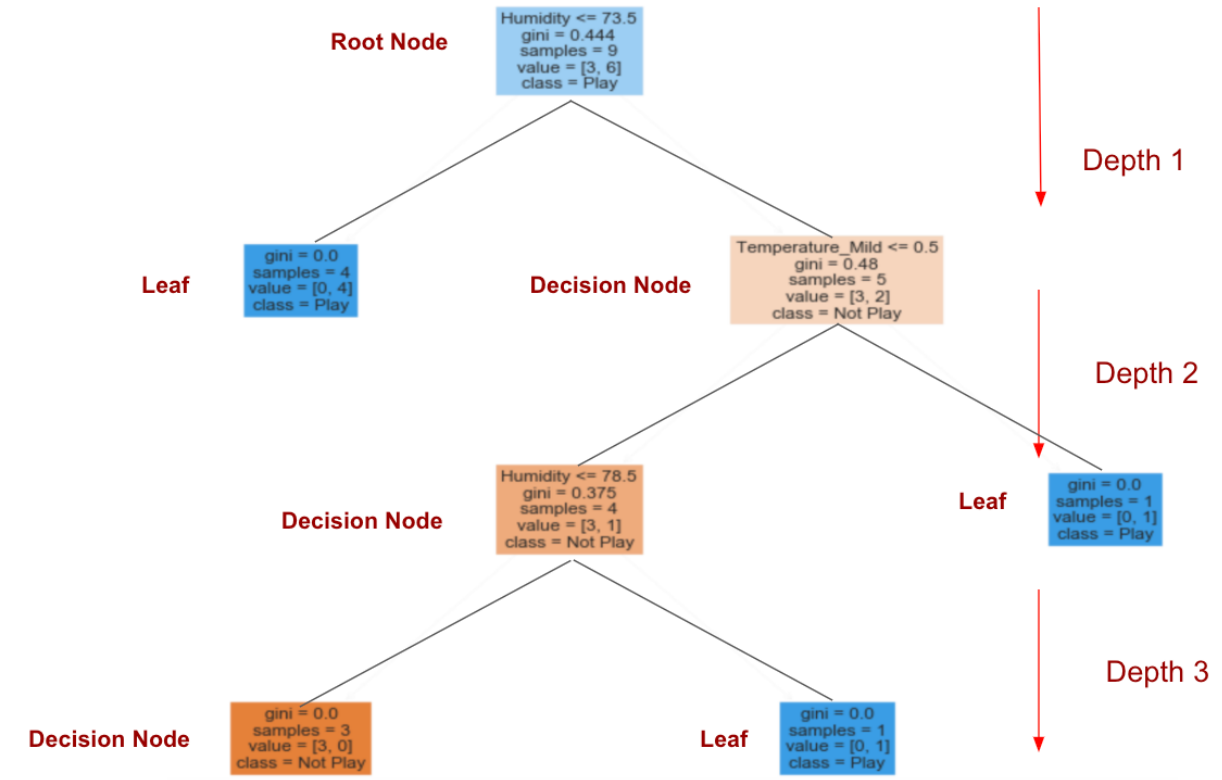
Example of a Classification Tree



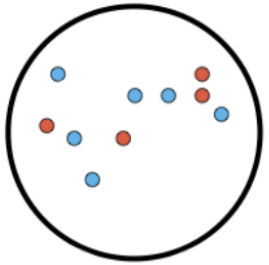
Risk for Cardiac Arrest

Age = 15
Weight = 65
High Risk

Decision Tree Max Depth



Decision Tree Min Samples Split

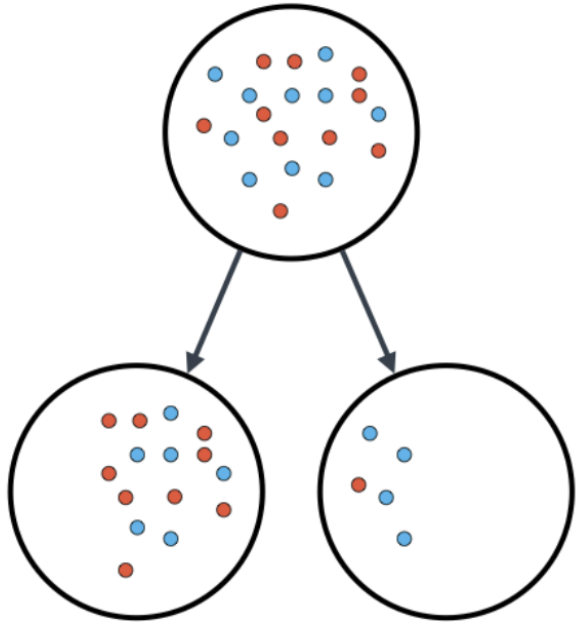


No split!

P(Low Risk) = 6/10 = 0.6

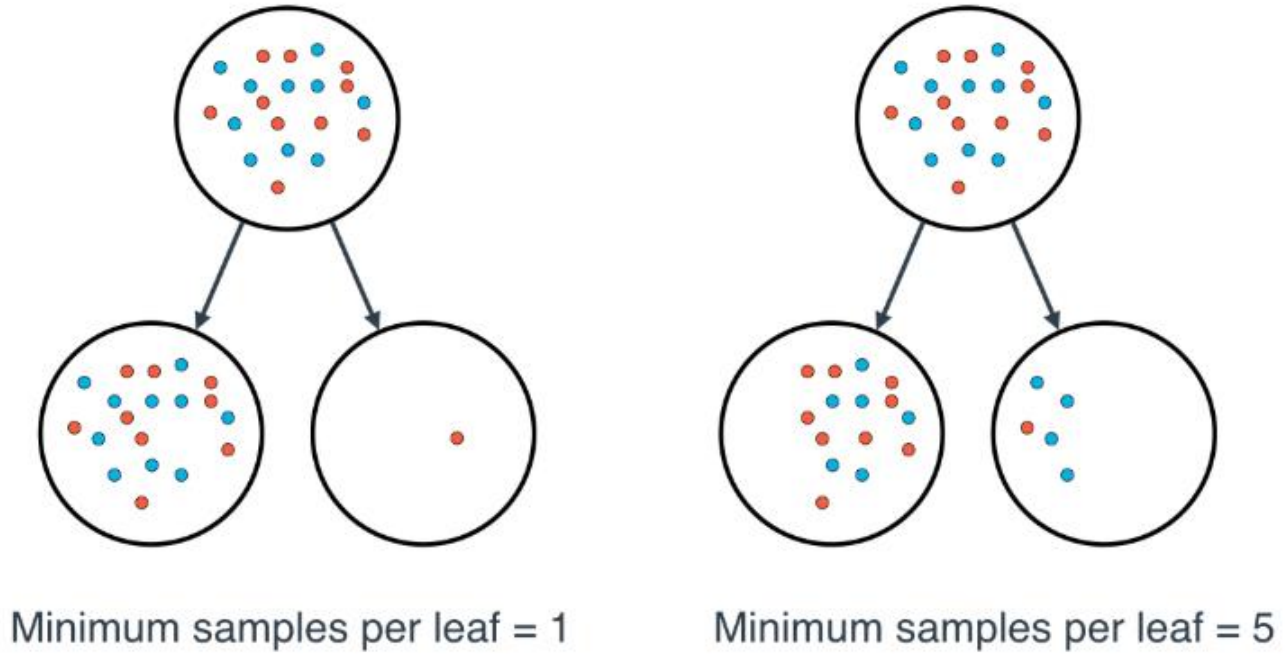
P(High Risk) = 4/10 = 0.4

Minimum number of samples to split = 11



Minimum number of samples to split = 11

Min Samples Leaf



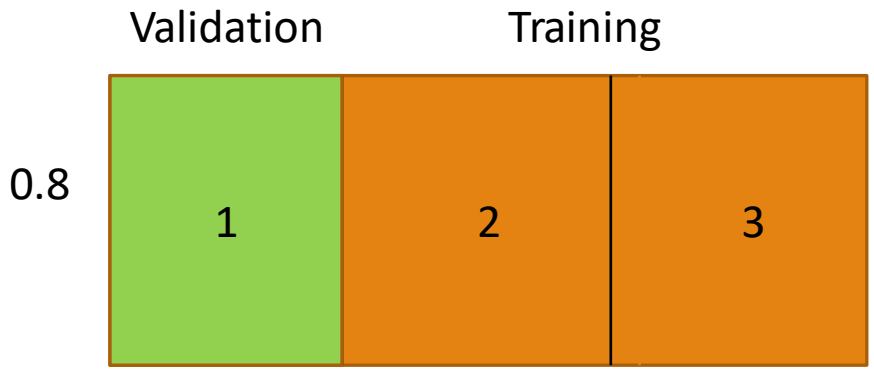
Minimum number of samples per leaf

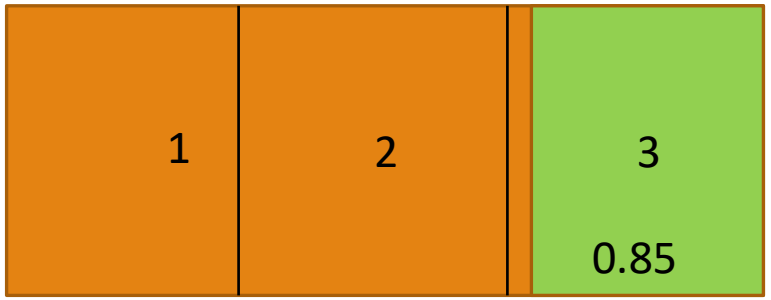
Gini vs Entropy

$$Gini: Gini(E) = 1 - \sum_{j=1}^{c} p_j^2$$

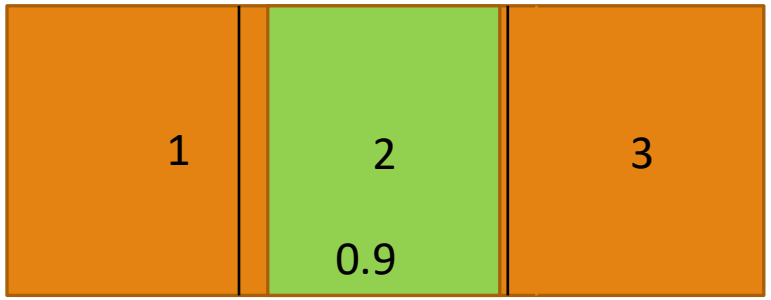
$$Entropy: H(E) = -\sum_{j=1}^{c} p_j \log p_j$$

K Fold Cross validation





3 fold validation



(0.8+0.9+0.85)/3

Thank you

PING ME ON SKYPE FOR ANY QUERIES