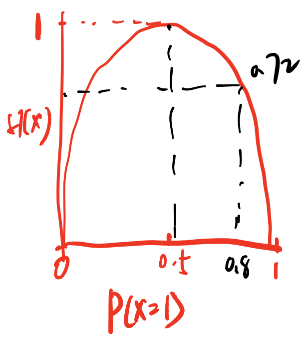
Entropy:

Entropy is one way to express the variance, or spread of a discrete random variable

$$H(X) = \sum_i P(x_i) I(x_i) = -\sum_i P(x_i) \log P(x_i)$$

ex: Coin flip. X is a RV with possible values {H, T}

Entropy of a coin: $-[P(H)\log P(H) + P(T)\log P(T)]$



Fair Coin:

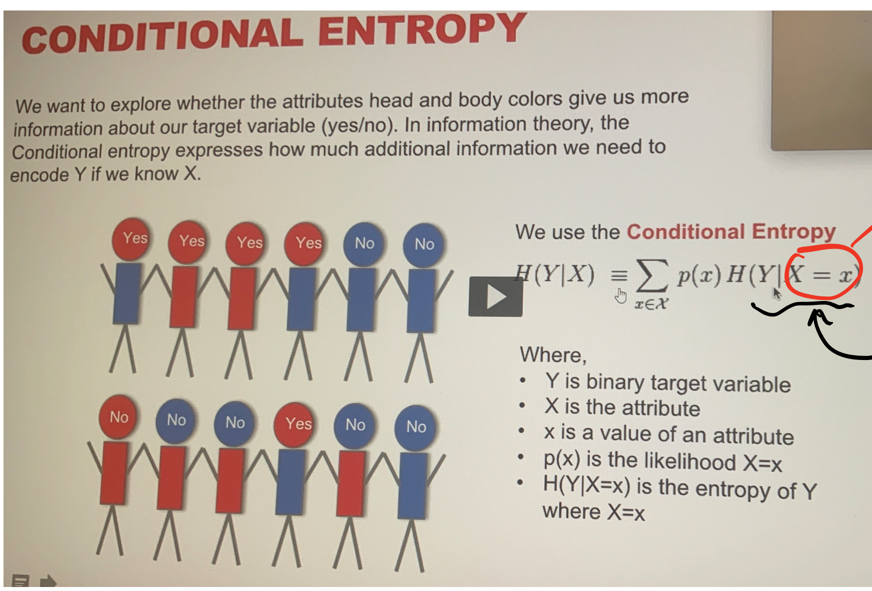
$H(X) = -[0.5\log(0.5) + 0.5\log(0.5)] = 1$ most uncertain

Single outcome coin:

$H(X) = -[1\log(1)] = 0$ most certain

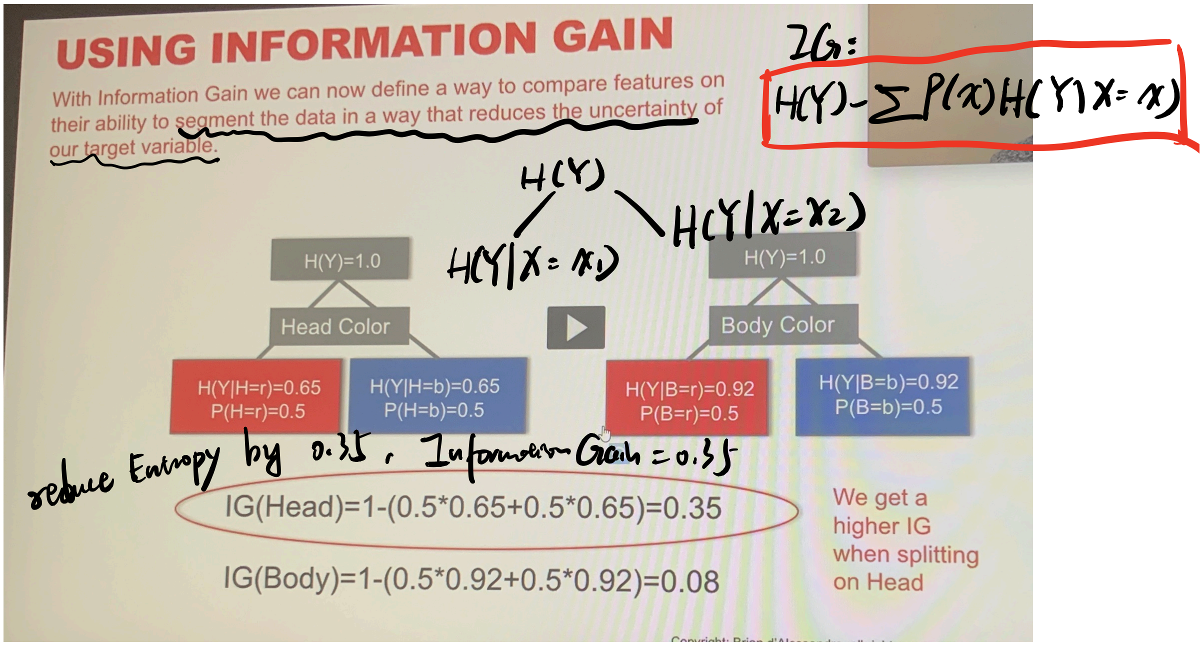
Biased coin:

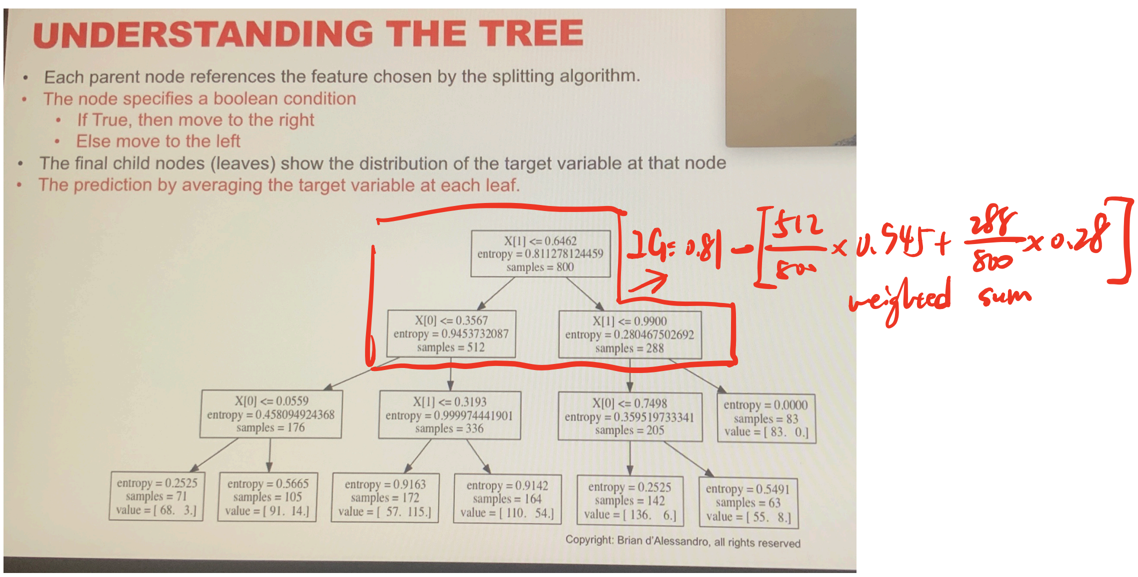
$H(X) = -[0.2\log(0.2) + 0.8\log(0.8)] = 0.72$ Less uncertain than $p=0.5$



segmenting the population by X

In certain segmentation, what is the entropy of Y



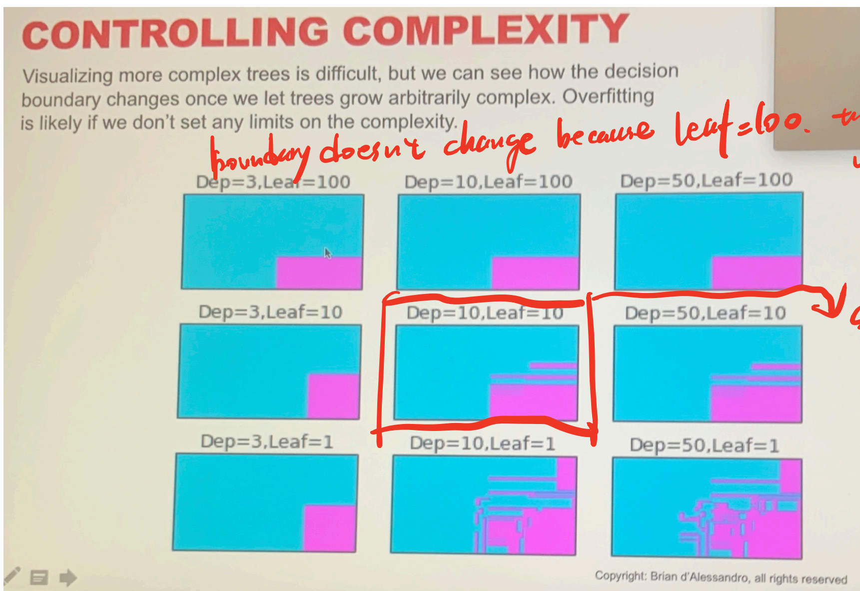


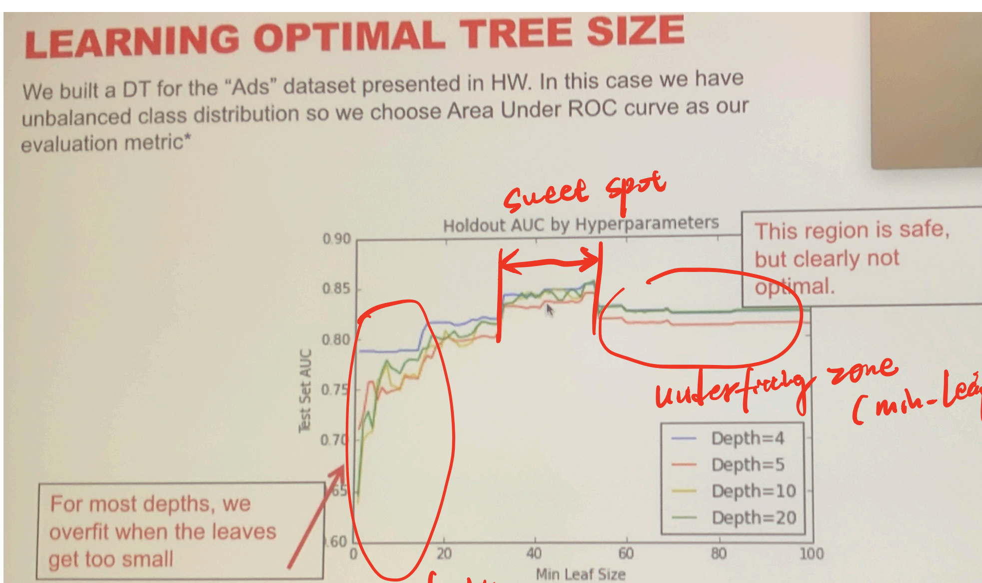
Controlling Complexity:

1. depth of trees
2. limit the size of an internal node that can be split (Min split size)
3. limit the minimum number of instances in the leaf (Min leaf size)

leaf node with very few samples → high variance
leaf node with lot of samples → high bias

- The ultimate goal of all three hyperparameters is to control the sample size in the leaf node





(min leaf ~~size~~ zone too small)

Feature Importance :

Returns the normalized information gain for each feature, can deal with collinearity of variables as well as the relationship to the target variable.