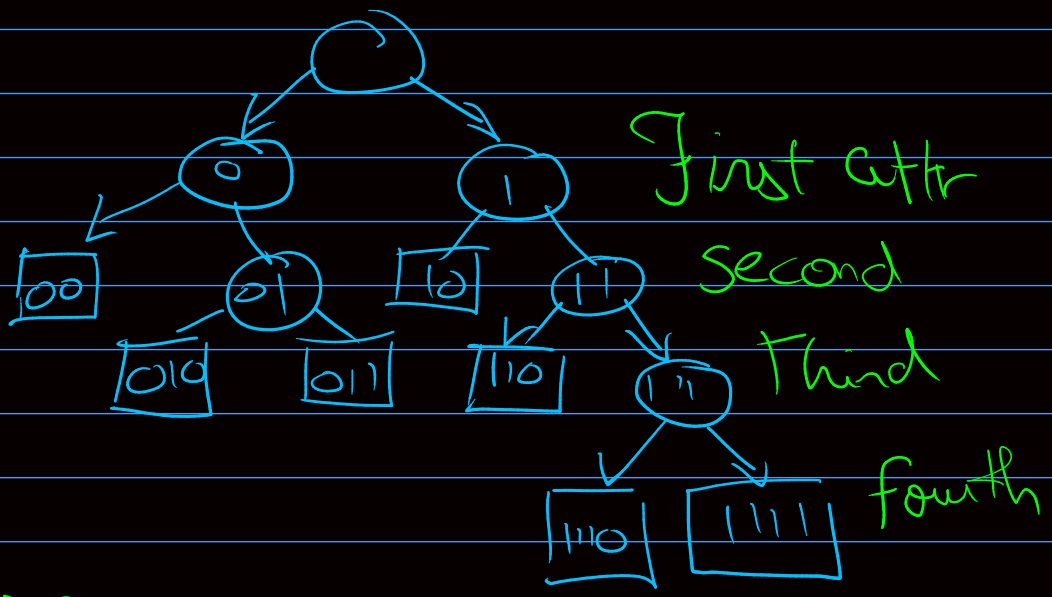
Decision Tree

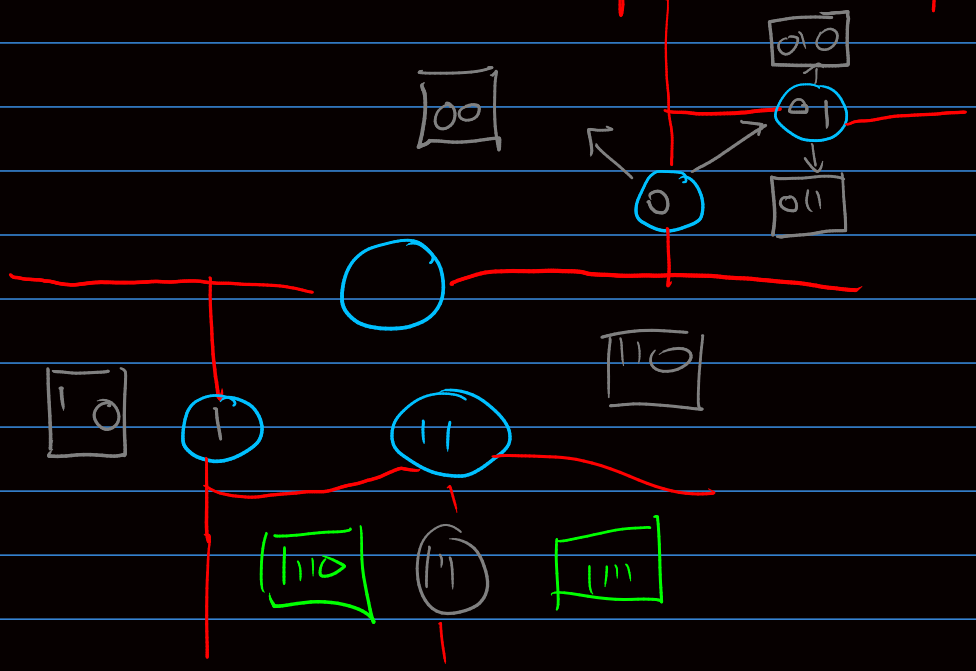


→ Considering a full blown tree, no overfitting

2000 training examples ⇒ 2000 leaves

to look at the decision boundaries

~~Or~~ divide the 2-D space into 2 parts



→ In this case, the max features a decision tree can take is ~~the~~ based on the number of training examples.

→ In reality there will be exp. no of leaves

→ Decision tree doesn't work

→ you lose features

Curse of Dimensionality:

Imagine 1 attr taking 10 values

You see the samples for each 10

So now You need atleast 10 samples

if there are 2 attr. ⇒ 10x10 = 100 samples

3 attr = 1000 samples

30 attr ⇒ 10^{30} samples

⇒ Thats why local constancy assumption for a good number of classical ML the curse of dimensionality holds