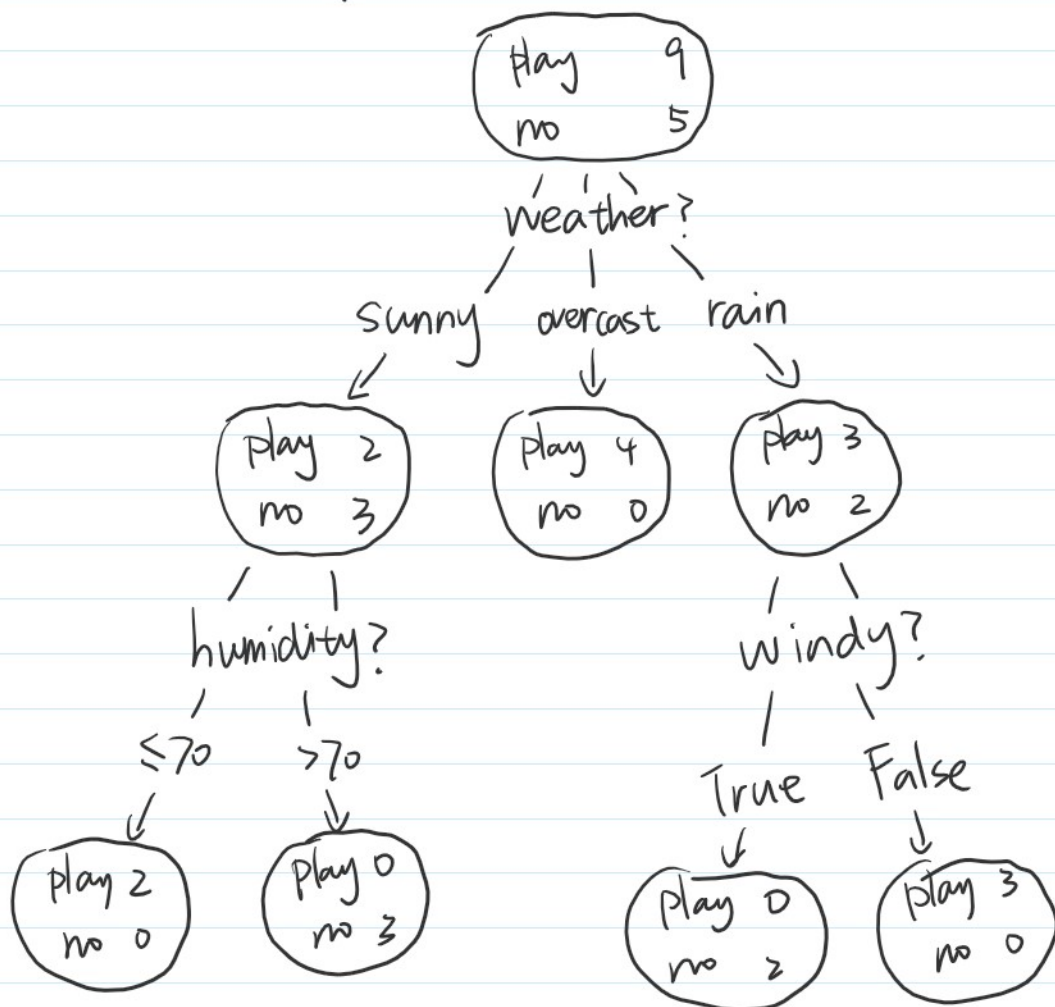


决策树算法.

类似于流程图的树型结构, 每一个内部结点表示在一个属性上的测试, 每一个分支代表一个属性输出.

Dependent Variable: PLAY.



信息熵 (entropy)

一条信息的信息量大小和它的不确定性有直接关系
变量的不确定性少, 信息熵少

变量的不纯度，信息熵

$$H(x) = - \sum_{i=1}^n P(x_i) \cdot \log_2 [P(x_i)].$$

ID3 算法.

依信息熵来选择属性节点(树节点)

信息熵

$$\text{Gain}(x) = H(D) - H(D-x)$$

其中 D 为所用数据集， x 为其中的一个属性。

在一个节点处，将所有属性的 Gain 值全部计算一次，
哪一个属性的 Gain 高，则作为节点。