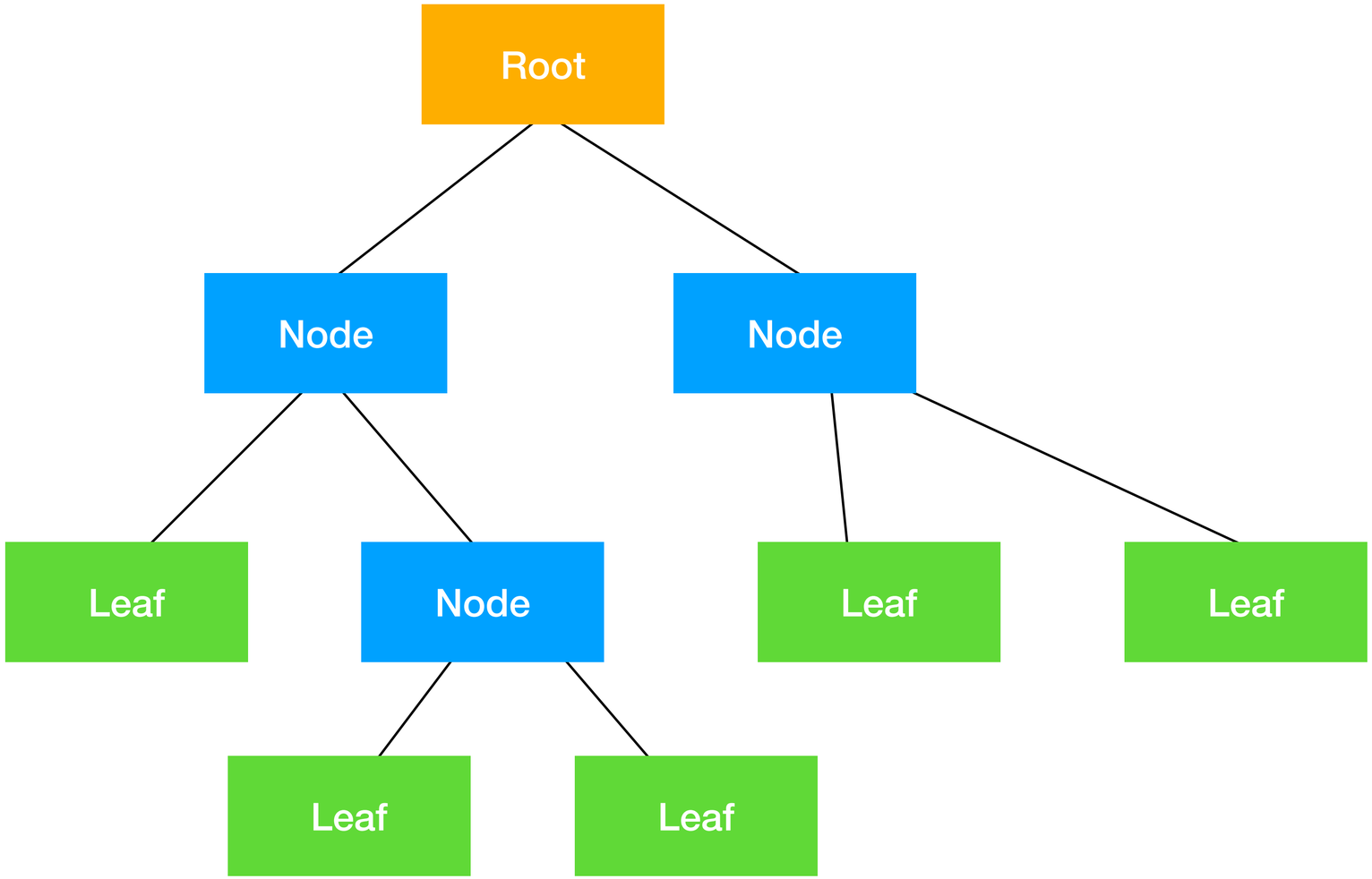
Decision Trees

Daniel Capurro, MD, PhD

Decision Trees

- A type of supervised Machine Learning (ML)
- One of the most frequently used ML algorithms
- They are both powerful and interpretable
    - When it classifies, we can see which attributes are driving the classifications

Decision Trees are upside down



How do they work?

	Age	Diabetes	Chest Pain	Fever	HOSP
Patient 1	68	Yes	Yes	No	Yes
Patient 2	73	No	Yes	Yes	Yes
Patient 3	81	No	No	Yes	No
Patient 4	78	Yes	No	No	No
Patient 5	47	No	No	No	No
Patient 6	64	Yes	Yes	No	Yes

Step 1: everyone is in the root

Root

	Age	Diabetes	Chest Pain	Fever	HOSP
Patient 1	68	Yes	Yes	No	Yes
Patient 2	73	No	Yes	Yes	Yes
Patient 3	81	No	No	Yes	No
Patient 4	78	Yes	No	No	No
Patient 5	47	Yes	No	No	No
Patient 6	64	Yes	Yes	No	Yes

Root node is ‘impure’, we can measure impurity using different metrics. The algorithm decides which attribute is best to reduce average impurity of the resulting nodes.

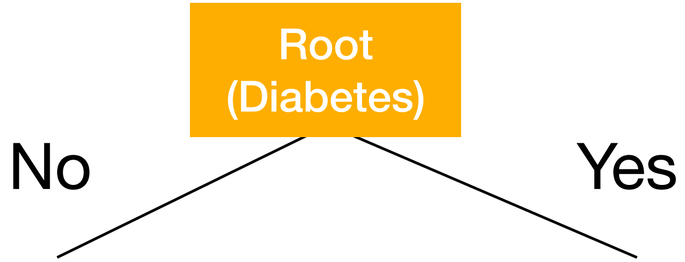
Entropy $H(X) = -\sum_{i=1}^{k} P(x)\log_2 P(x)$

- (P(yes) $\log_2$P(yes) + P(no)$\log_2$P(no))

=1

Step 2: split the root

Using the attribute that reduces impurity the most



	Age	Diab	Chest Pain	Fever	HOSP
P2	73	No	Yes	Yes	Yes
P3	81	No	No	Yes	No
P5	47	No	No	No	No
P4	78	No	No	No	No

Entropy = 0.81

	Age	Diab	Chest Pain	Fever	HOSP
P1	68	Yes	Yes	No	Yes
P6	64	Yes	Yes	No	Yes

Entropy = 0.81

The weighted average of entropy is the entropy for the tree as it is right now

$$4/6 \times 0.81 + 2/6 \times 0 = 0.54$$

We moved from entropy 1 to entropy 0.54, so information gain is 0.46

Step 3: keep splitting

- The algorithm splits using the attribute that maximizes information gain
- Continues to do so until it cannot split anymore or until it meets conditions that we have pre defined (depth of the tree, number of elements in the leaf)

- Now let’s work an example