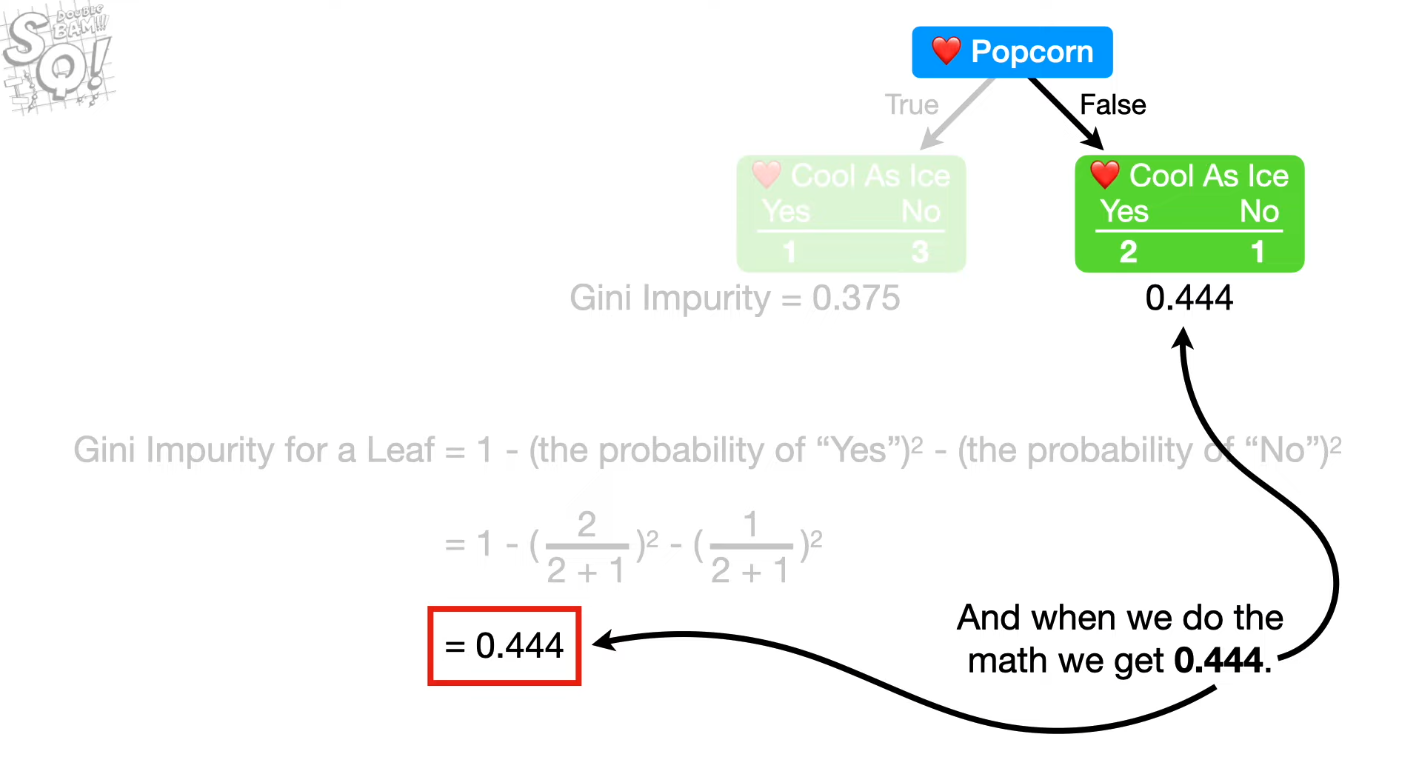
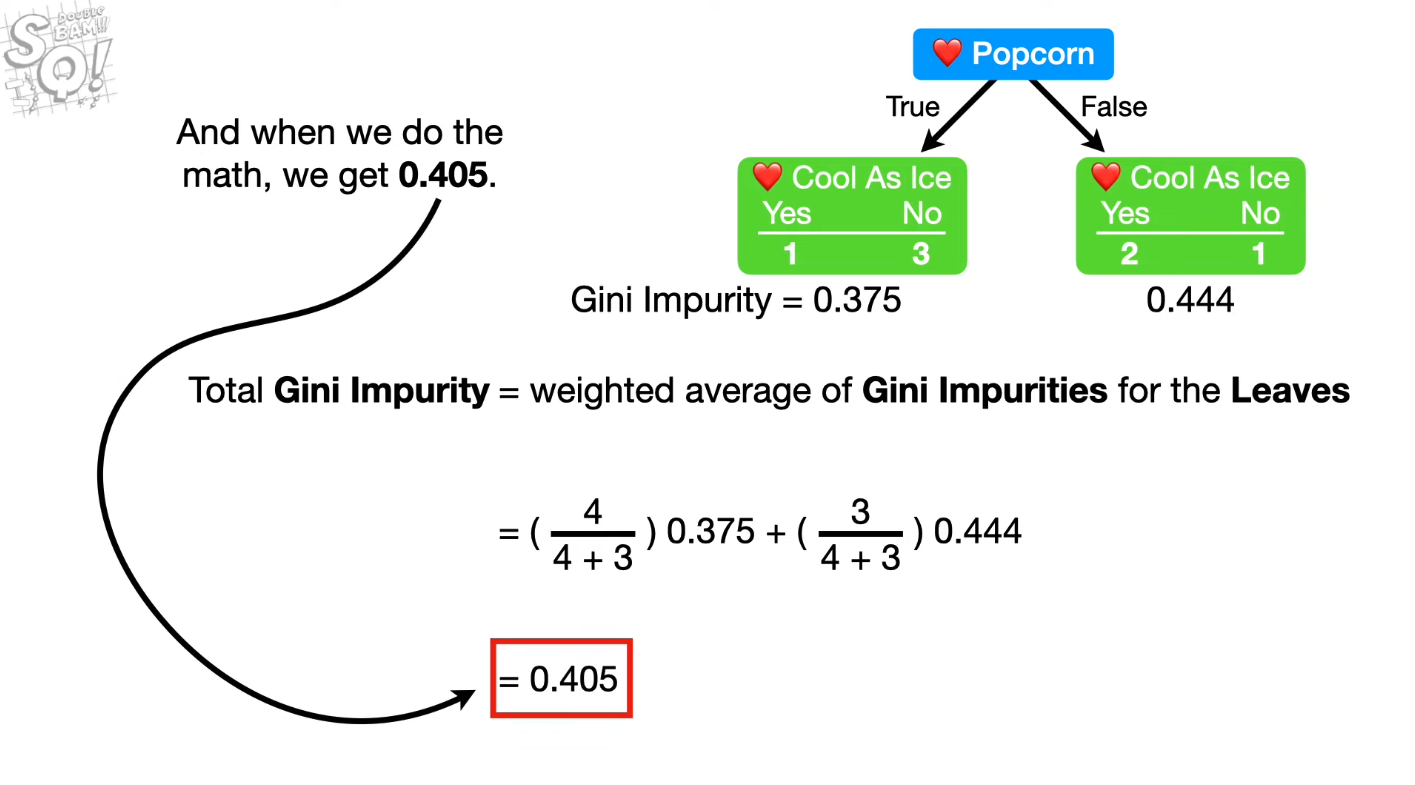
# Classification criteria: Gini Impurity

[Sklearn notes](https://scikit-learn.org/stable/modules/tree.html#mathematical-formulation).

Where *k* = target class, and is the probability of obtaining the class *k* in the leaf.

Where = each leaf.





# Regression criteria: Mean Squared Error

Where m = number of samples. = the predicted values at a specified threshold, computed using the mean values of the data after splitting by threshold. For example, a split at the left (less than the threshold) would take the mean value of the data less than the threshold.

# Steps

## Build the tree:

1. Start at root node (or root), split based on the best total Gini impurity (for classifier), or mean squared error (for regressor).
2. Search over all features and thresholds, thresholds are created using the midpoints of each adjacent sorted values in each feature.
3. Save the best split feature and threshold at each node.
4. Build the tree recursively with the same method for each node.
5. Apply some stopping criteria to stop growing, if necessary, e.g. max depth, min samples at node, or no more class distribution at node.
6. For every leaf node, save the highest voted class as the predicted class label of the node.

## Predict (traverse the built tree):

1. Traverse the tree recursively.
2. At each node, look at the split feature and the corresponding threshold and decide to go left or right. Go left if , else go right.
3. When reach the leaf node, return the highest voted class as the predicted class.