



Best Splits

- Pick best attributes and conditions on which to partition
- The purity of a set S of training instances can be measured quantitatively in several ways.
 - Notation: number of classes = k , number of instances = $|S|$, fraction of instances in class $i = p_i$.
- The **Gini** measure of purity is defined as

[

$$\text{Gini}(S) = 1 - \sum_{i=1}^k p_i^2$$

- When all instances are in a single class, the Gini value is 0
- It reaches its maximum (of $1 - 1/k$) if each class the same number of instances.