

# 1 Decision Tree Classification

Choose a variable at each step that best split the sets of items. To build the tree, calculate Information Gain of each step. The best split is one that provides the **MOST** Information Gain

1. **Entropy:**

$$H(T) = I_E(p_1, p_2, \dots, p_J) = - \sum_{i=1}^J p_i \log_2 p_i$$

2. **Information Gain:** used to decide which feature to split on at each step in building the tree

## 1.1 Regression Tree

If the target is continuous, use Regression Decision Tree To build the tree, choose a split, then calculate weighted variance at each step, then calculate the variance reduction. The root split will be the node with the **MOST** Variance Reduction

1. **Variance:**

$$\text{Var}(X) = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

2. **Variance reduction:** Used as a criterion to determine the quality of a split