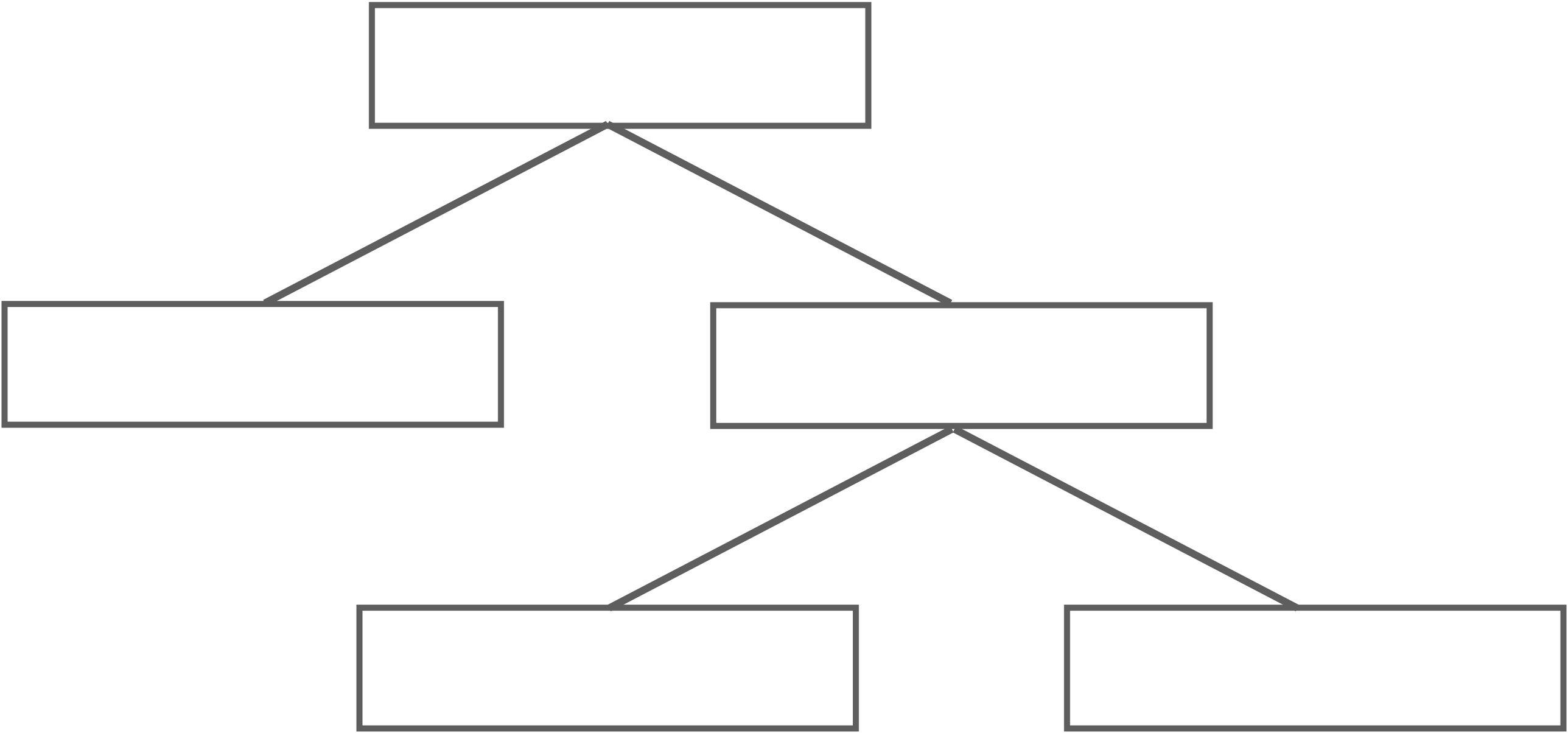
Review: Basic Steps

1. Compute Gini index or Entropy as measure of impurity for each node

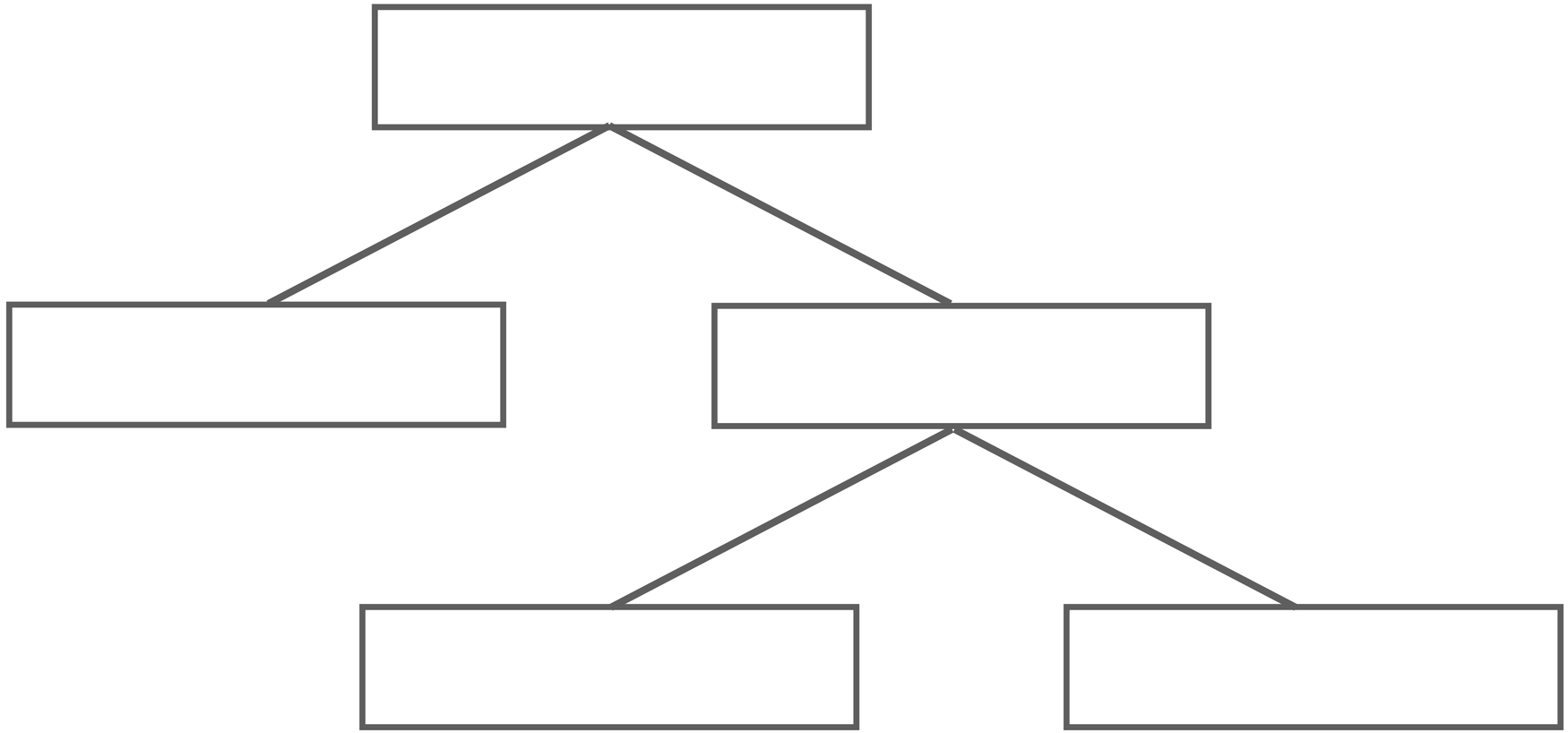
2. Choose node with lowest score

3. If the parent node has the lowest score, it is a leaf



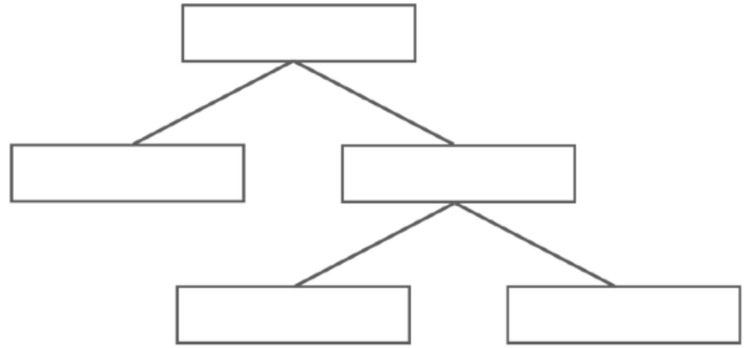
Review: Basic Steps

1. Compute Gini index or Entropy as measure of impurity for each node
2. Choose node with lowest score
3. If the parent node has the lowest score, it is a leaf



Variable Importance Measure: Gini Importance

1. How much does this feature reduce node impurity?



weighted parent node impurity

$$\underbrace{\text{node-impo}_j}_{\text{importance of node j}} = \overbrace{w_j C_j}^{} - \underbrace{\left(w_{\text{left}_j} C_{\text{left}_j} + w_{\text{right}_j} C_{\text{right}_j} \right)}_{\text{weighted child node impurity}}$$

feature importance (fi):

$$\text{fi}_j = \frac{\sum_{j \in S_i} \text{node-impo}_j}{\sum_{k \in S_{all}} \text{node-impo}_k}$$ where S_i is set of all nodes that split on feature$_i$