

# Decision Tree Exercises

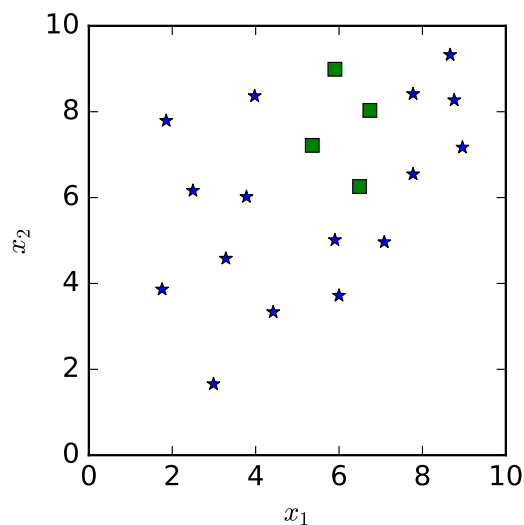
## 1. Gini Impurity

The goal in building a decision tree is to create the smallest possible tree in which each leaf node contains training data from only one class. In evaluating possible splits, it is useful to have a way of measuring the *purity* of a node. The purity describes how close the node is to containing data from only one class. Gini purity is defined as follows<sup>1</sup>:

$$\phi(\mathbf{p}) = \sum_j p_j(1 - p_j)$$

Where  $\mathbf{p} = (p_1, \dots, p_n)$  and each  $p_i$  is the fraction of elements from class  $i$ . This expresses the fractions of incorrect predictions in the node if the class of each element was predicted by randomly selecting a label according to the distribution of classes in the node. This value will be 0 if all elements are from the same class, and it increases as the mix becomes more uniform.

Calculate the Gini impurity of the following data set:



---

<sup>1</sup>Notation follows: Breiman, Leo. "Technical note: Some properties of splitting criteria." Machine Learning 24.1 (1996): 41-47.

## 2. Tree Construction

The decision tree construction algorithm proceeds by recursively splitting the training data into increasingly smaller subsets. When splitting a node in the tree we search across all dimensions and all split points to select the split that results in the greatest decrease in impurity. This goodness-of-split value can be expressed as:

$$\Theta(s, t) = \phi(\mathbf{p}) - P_L\phi(\mathbf{p}_L) - P_R\phi(\mathbf{p}_R)$$

Where  $s$  is a possible split,  $t$  is the node and  $P_L$  and  $P_R$  represent the fraction of elements of nodes that ended up in the left child and right child respectively. Higher values represent better splits.

Execute the recursive tree-construction algorithm on the data above and draw the resulting tree. Calculate the impurity of each node and the goodness-of-split for each split.

## 3. Classification

Classify the following three points using your decision tree.

(4, 10)  
(6, 10)  
(6, 0)