$$g(x_1, x_2) = 10.5\,\mathbb{1}_{x_1 > 5} + 55\,\mathbb{1}_{x_1 \le 5}\mathbb{1}_{x_2 \le 5} + 10.5\,\mathbb{1}_{x_1 \le 5}\mathbb{1}_{x_2 \ge 5}$$

Binary Tree Models are linear models where the variables are carved into partial spaces like categorical variables.

where all categorical variables are binarized already

One regression tree algorithm (there are many)

1. Begin with $\mathbb{D} = \langle X, \vec{y} \rangle$, $X = \begin{bmatrix} \vec{x}_{\cdot 1} & \vec{x}_{\cdot 2} & \cdots & \vec{x}_{\cdot p} \end{bmatrix} = \begin{bmatrix} \vec{x}_1 \\ \vec{x}_2 \\ \vdots \\ \vec{x}_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & & & \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$

2. Consider all possible orthogonal-to-axis splits i.e.

$$X_1 \le x_{(1)1},\ X_1 \le x_{(2)1}, \ldots, X_1 \le x_{(n-1),1},$$

Sorted $\vec{x}_{\cdot 1}$ values from min - max.

$$X_2 \le x_{(1)2},\ X_2 \le x_{(2)2}, \ldots, X_2 \le x_{(n-1),2},$$

$$\vdots$$

$$X_p \le x_{(1)p},\ X_p \le x_{(2)p}, \ldots, X_p \le x_{(n-1),p} \implies (n-1)p \text{ possible splits}$$

For each split there are two putative daughter nodes.
For each split we assign yhat = ybar of the observations that landed in the node. Calculuate SSE in both daughters.

3. Locate the split rule with smallest #-obs-weighted average SSE:

$$SSE_w := \frac{n_L\, SSE_L + n_R\, SSE_R}{n_L + n_R}$$

$n_L :=$ # of obsevations in left daughter node
$n_R :=$ # of obsevations in right daughter node

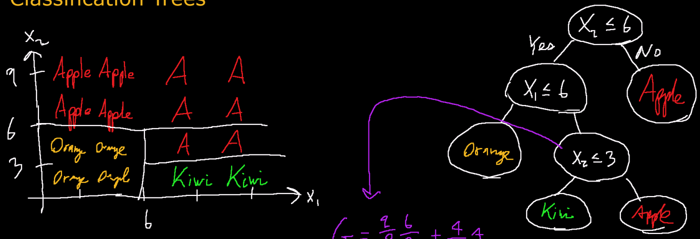
This algorithm is "greedy" because at each juncture it takes the locally optimal solution and these algorithms never find global optimums.

4. Create the optimal split and the two daughter nodes.

5. Repeat steps 2-4 on the daughter nodes unless daughter node has N_0 observations (e.g. N_0 = 5) and unless there's only one unique y.

Let $y = \{C_1, C_2, \ldots, C_K\}$ classification problem

Classification Trees



$$G = \frac{2}{8}\frac{6}{8} + \frac{4}{8}\frac{4}{8} + \frac{2}{8}\frac{6}{8} = \frac{40}{64}$$

One classification tree algorithm is as follows:

1. Same as Regression Tree Algorithm above

2. Consider all splits as in the Regression Tree Algorithm.
For all daughter nodes set
yhat = Mode[y's in the daughter node]
and calculate a heterogeneity metric called the "Gini"

$$G := \sum_{l=1}^{K} \hat{p}_l (1 - \hat{p}_l) \quad \text{where} \quad \hat{p}_l := \frac{\#\{y_i \text{ in category } l\}}{\#\text{ obs. in node}}$$

3. Same as Regression Tree Algorithm above. Take the #-obs-weighted Gini average:

$$G_w := \frac{n_L G_L + n_R G_R}{n_L + n_R}$$

4. Same as Regression Tree Algorithm above.

5. Same as Regression Tree Algorithm above. Stop at default N_0 = 1 or if node is completely homogeneous.