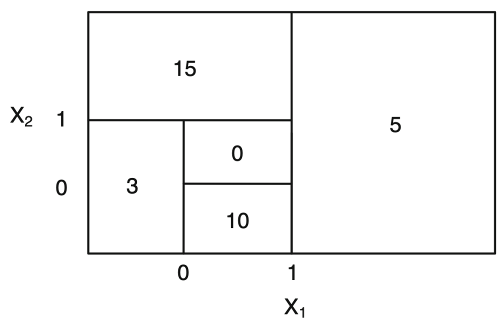
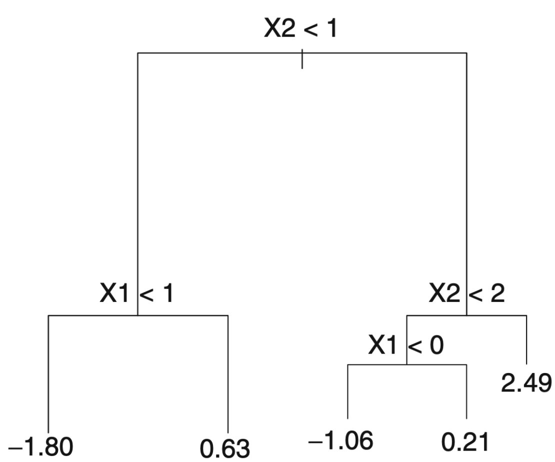
Biostat 626 Assignment 3

Due: 11:59 PM, Friday, April 21

1. Find the decision tree corresponding to the partition of the feature space illustrated in the following figure. The numbers inside the boxes indicate the mean of outcome, Y, within each region.



2. Find the partition of the feature space corresponding to the decision tree illustrated in the following figure. The leafs of the tree indicate the mean of outcome, Y, within each branch.



3. Consider a committee in which we allow unequal weighting of the constituent models, so that

$$y_{COM}(\boldsymbol{x}) = \sum_{m=1}^{M} \alpha_m y_m(\boldsymbol{x}).$$

In order to ensure that the predictions $y_{COM}(\boldsymbol{x})$ remain within sensible limits, suppose that we require that they be bounded at each value of $\boldsymbol{x}$ by the minimum and maximum values given by any of the members of the committee, so that

$$y_{min}(\boldsymbol{x}) \leq y_{COM}(\boldsymbol{x}) \leq y_{max}(\boldsymbol{x}).$$

Show that a necessary and sufficient condition for this constraint is that the coefficients satisfy

$$\alpha_m \geq 0, \qquad \sum_{m=1}^{M} \alpha_m = 1.$$

4. For the maximum variance formulation of principal component analysis, show that the linear projection on a q-dimensional subspace that maximizes the variance of the projected data is defined by the q eigenvectors of the data covariance matrix $\boldsymbol{S}$, corresponding to the q largest eigenvalues. In class, this result is proven for $q = 1$. Now generalize this result for general M by induction.