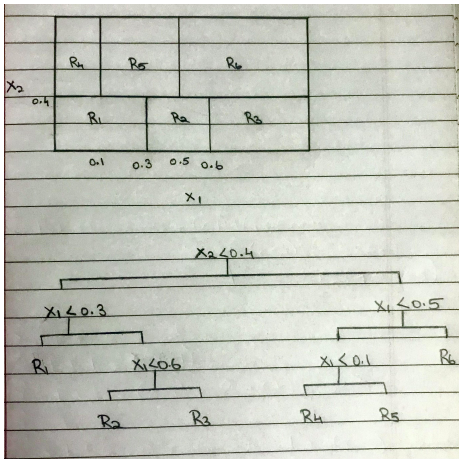
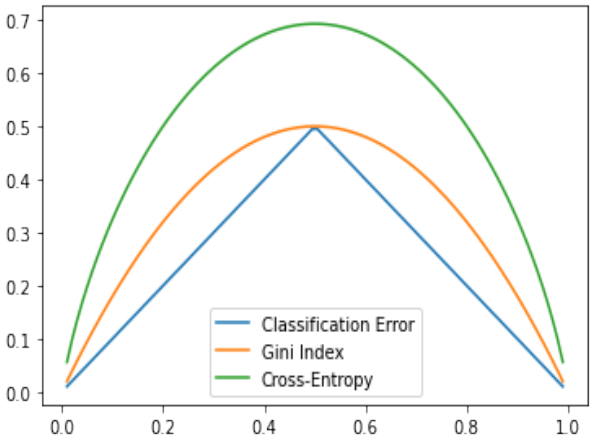
Chapter 8 - Tree-Based Models

1. The below image contains 6 regions and 6 cut-points.

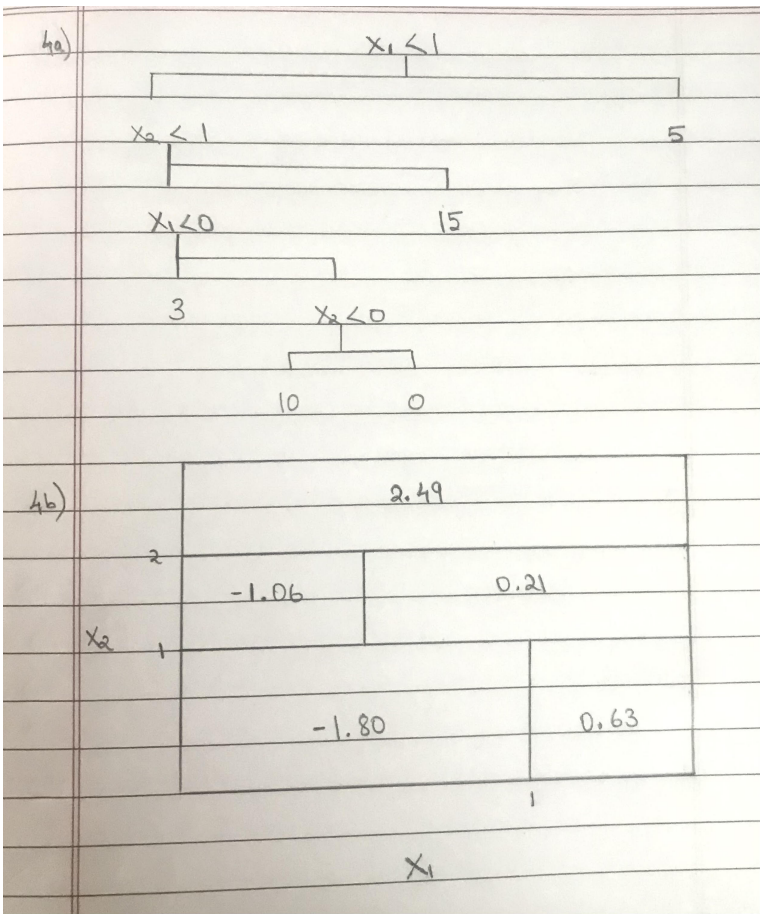


2. $f(X) = \sum_{b=1}^{B} \lambda f^b(X)$. When boosting is done using depth-one trees, then each tree only considers one variable. Thus, $f^b(X_i) = \beta_0 + \beta_1 I(X_i < S)$. Hence, $f(X) = \beta_0 + \beta_1 I(X_1 < S_1) + \beta_2 I(X_2 < S_2)$... which is additive in nature.

3. Here $\hat{p}_{m1}$ is on the x-axis, and the errors are on the y-axis.



4. The below image contains solutions for 4a and 4b.



5. $P(RED|X) = [0.1, 0.15, 0.2, 0.2, 0.55, 0.6, 0.6, 0.65, 0.7, 0.75]$.
   Average of $P(RED|X) = 0.45$, thus X does not belong to class RED.
   Max Vote is for class RED (6 vs 4), thus X belongs to class RED.

6. Firstly a large tree is created using a greedy approach called recursive binary splitting to divide the predictor space into J regions by minimizing RSS given by $\sum_{j=1}^{J} \sum_{i \in R_j} (y_i - \hat{y}_{R_j})^2$. Now prune the tree to get a set of subtrees as a function of α. Now using cross validation we can choose the best among the subtrees.