An Introduction to Statistical Learning: Chapter 8

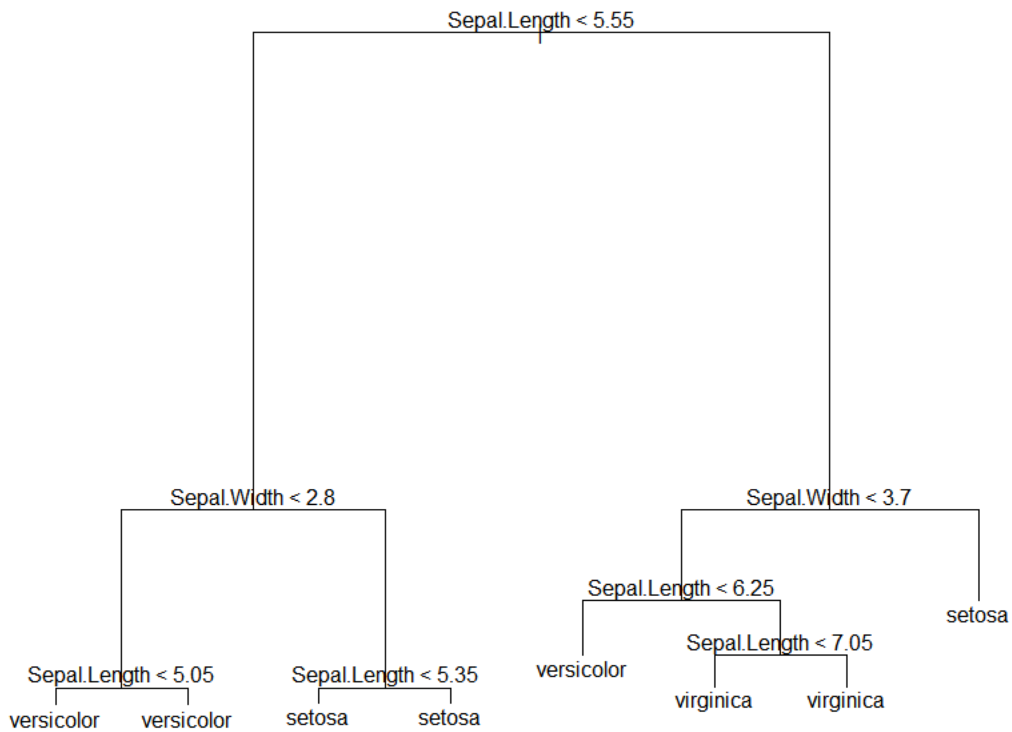
Due on January 15,2016 at 3:10pm

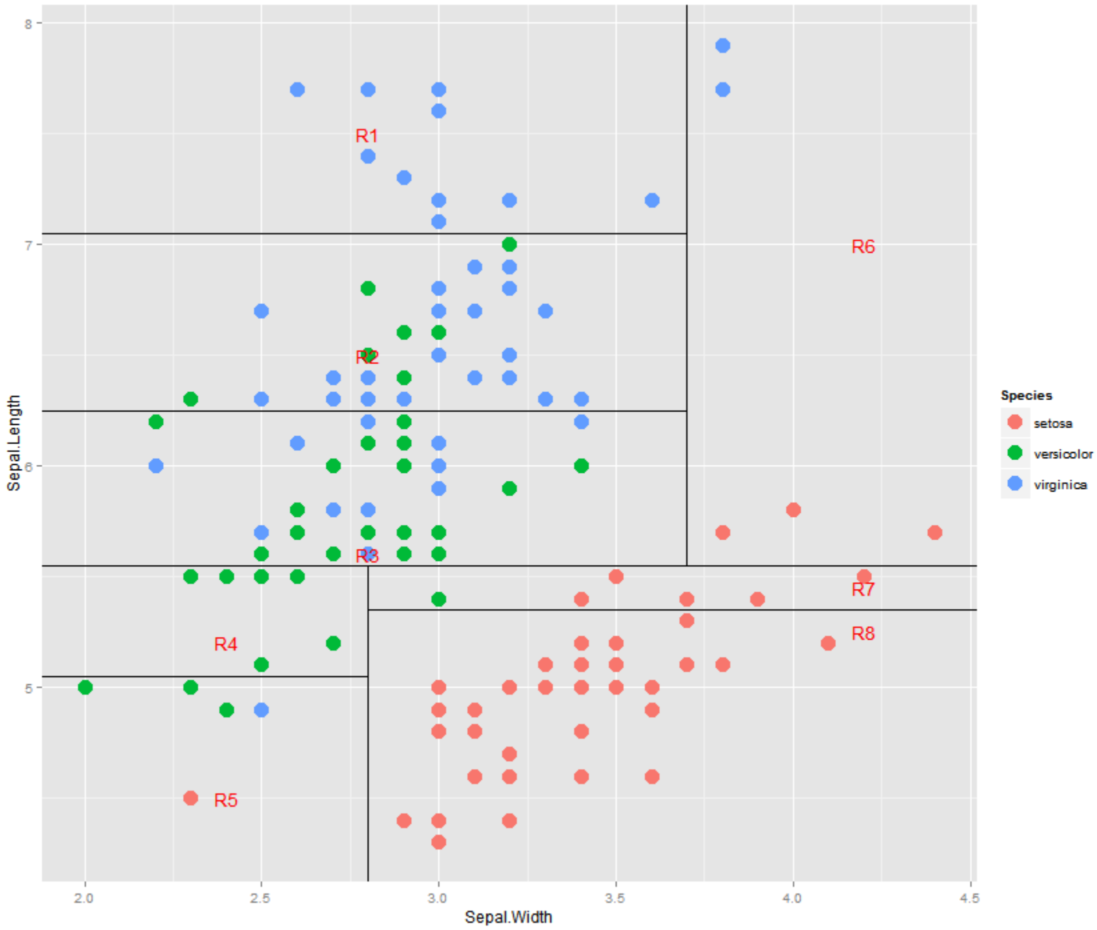
G. James et.al.

Tony Jiang

1. Draw an example (of your own invention) of a partition of two dimensional feature space that could result from recursive binary splitting. Your example should contain at least six regions. Draw a decision tree corresponding to this partition. Be sure to label all aspects of your figures, including the regions $R_1, R_2, \ldots$, the cutpoints $t_1, t_2, \ldots$, and so forth.
Hint: Your result should look something like Figures 8.1 and 8.2.

Answers:





2. It is mentioned in Section 8.2.3 that boosting using depth-one trees (or *stumps*) leads to an additive model: that is, a model of the form

$$f(X) = \sum_{j=1}^{n} f_j(X_j).$$

Explain why this is the case. You can begin with (8.12) in Algorithm 8.2.

Answers:

each split involves only one predictor at each splitx_{ip}

$$(i=1)f(x) = 0$$

$$(i=2)f(x) = \frac{\sum_{j=1}^{n} I(x_{jp})}{\sum_{j=1}^{n} I(x_{jp})} x_{jp}$$

$$(i=3)f(x) = \frac{\sum_{j=1}^{n} I(x_{jp})}{\sum_{j=1}^{n} I(x_{jp})} x_{jp} + \frac{\sum_{j=1}^{n} I(x_{jp'})}{\sum_{j=1}^{n} I(x_{jp'})} x_{jp'}$$

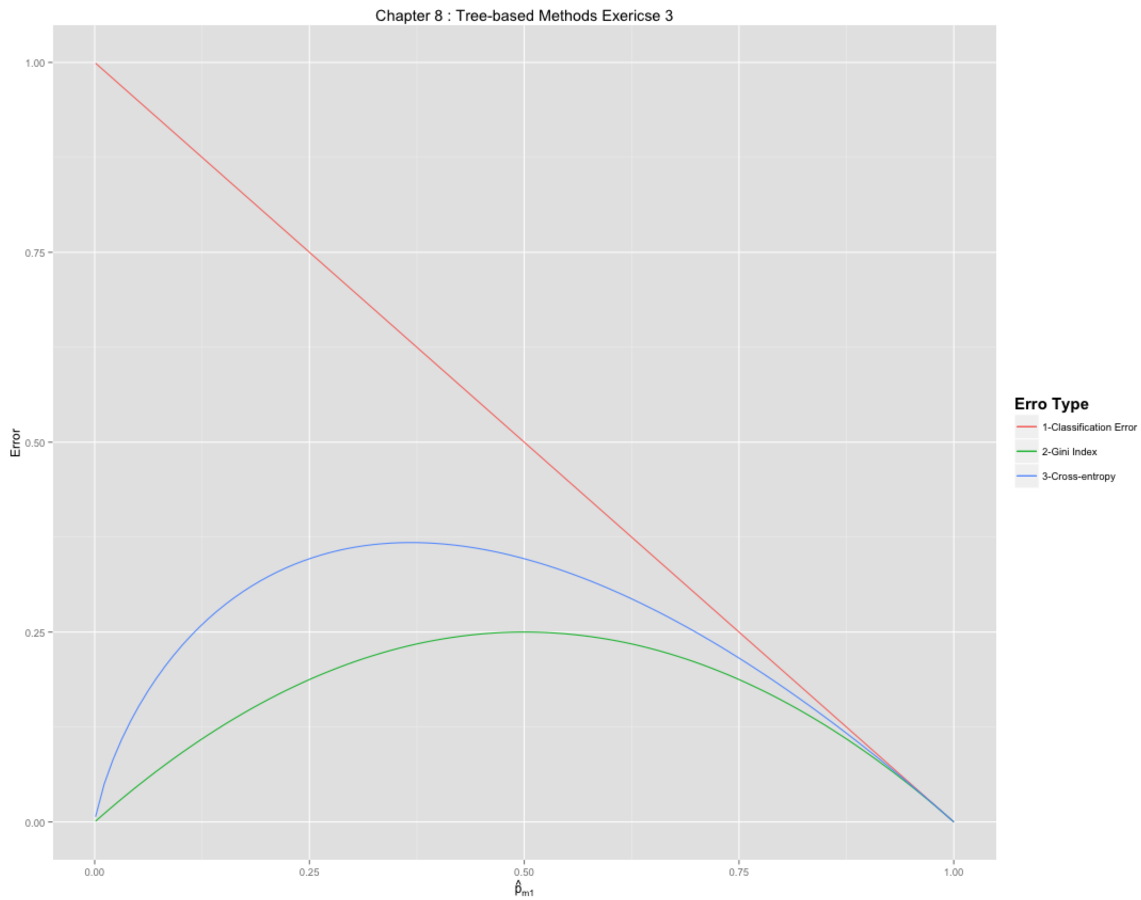
...

$$f(x) = \sum_{m=1}^{p} \frac{\sum_{j=1}^{n} I(x_{jm})}{\sum_{j=1}^{n} I(x_{jm})} x_{jm}$$

$$f(X) = \sum_{j=1}^{p} f_j(X_j)$$

3. Consider the Gini index, classification error, and cross-entropy in a simple classification setting with two classes. Create a single plot that displays each of these quantities as a function of $\hat{p}_{m1}$. The xaxis should display pm1, ranging from 0 to 1, and the y-axis should display the value of the Gini index, classification error, and entropy. Hint: In a setting with two classes, $\hat{p}_{m1} = 1 - \hat{}p_{m2}$. You could make this plot by hand, but it will be much easier to make in R.

Answers:



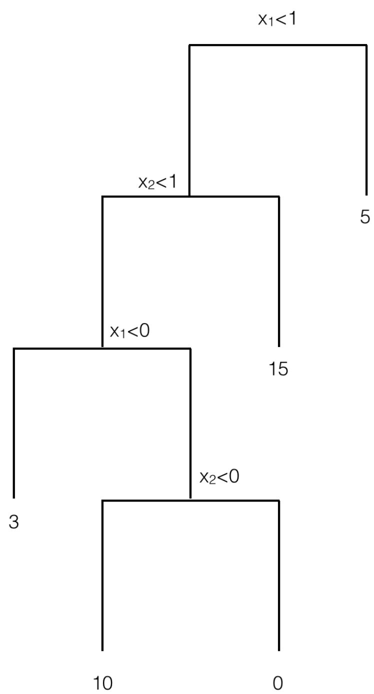
4. This question relates to the plots in Figure 8.12.

(a) Sketch the tree corresponding to the partition of the predictor space illustrated in the left-hand panel of Figure 8.12. The numbers inside the boxes indicate the mean of Y within each region.

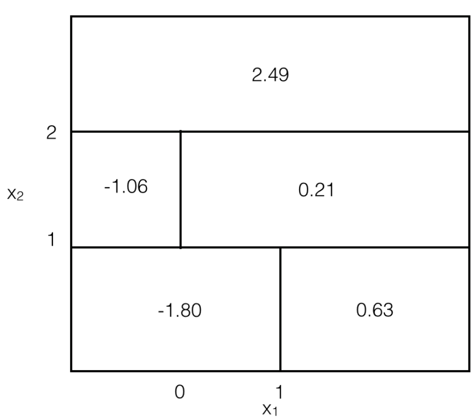
(b) Create a diagram similar to the left-hand panel of Figure 8.12, using the tree illustrated in the right-hand panel of the same figure. You should divide up the predictor space into the correct regions, and indicate the mean for each region.

Answers

(a).



(b).



5. Suppose we produce ten bootstrapped samples from a data set containing red and green classes. We then apply a classification tree to each bootstrapped sample and, for a specific value of X, produce 10 estimates of $P(\text{Class is Red}|\text{X})$:

$$0.1, 0.15, 0.2, 0.2, 0.55, 0.6, 0.6, 0.65, 0.7, \text{ and } 0.75.$$

There are two common ways to combine these results together into a single class prediction. One is the majority vote approach discussed in this chapter. The second approach is to classify based on the average probability. In this example, what is the final classification under each of these two approaches?

Answers:
majorit vote: 4 votes for Green and 6 for Red , so final classification=Red
average probability=$\frac{0.1+0.15+0.2+0.2+0.55+0.6+0.6+0.65+0.7+0.75}{10} = 0.45 > 0.5$. final classification= Green
6. Provide a detailed explanation of the algorithm that is used to fit a regression tree.

Answers:
Building a regression tree usually involves the following steps: Step1 Divide the predictor space into J distinct and non-overlapping regions. We usually divide them into boxes Step2. For every observation that falls into region R_j, we make the same prediction, which is the mean of the response value for the training observations.
A top-down, greedy approach usually starts with the top of the tree (just one region). Then for each split, we search for a split on predictor X_j that maximize :

$$\sum_{j=1}^{J} \sum_{i \in R_j} (y_i - \hat{y}_{R_j})^2,$$

We keep looking for the best predictor and best cut point to split until a stopping criterion is reached.