

Question 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 (R1, R2...) and the cut points (t1, t2...).

Question 2

In this exercise we will use regression trees and related approaches to predict the quantitative variable `sales` from the `carseat` dataset.

- a. Split the data set into a training set and a test set.
- b. Fit a regression tree to the training set. Plot the tree and interpret the results. What test MSE do you obtain?
- c. Use cross-validation in order to determine the optimal level of tree complexity. Does pruning the tree improve the test MSE?
- d. Use the bagging approach in order to analyze this data. What test MSE do you obtain?
- e. Use random forests to analyze this data. What test MSE do you obtain?
- f. Report the important features in your random forest. One can do this by using `importance()` function in R or `feature_importance_` on a fitted model in sklearn in python.
- g. Describe the effect of m (the number of variables considered at each split) on the error rate.