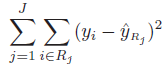
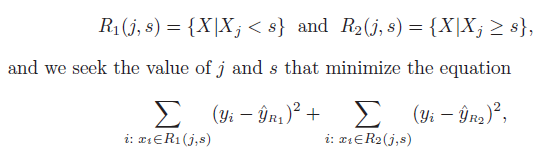
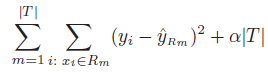
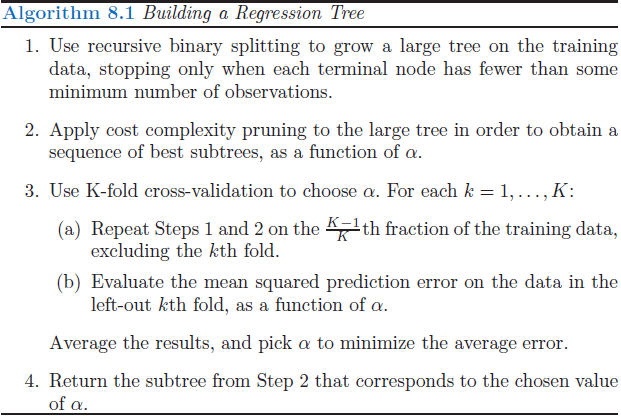
* Boostrapping – averaging a set of observations reduces variance
  + Variance =
  + Average resulting predictions, by taking repeated samples from single training set
* Regression tree 2 steps:
  + (1) Divide predictor space (X1, …, Xp) into J distinct and non-overlapping regions (R1, …, Rj)
  + (2) For every observation in region Rj 🡪 make prediction
    - Prediction = mean of response values for training observations in Rj
* Step (1):
  + Find regions to minimize RSS
    - 
  + *Greedy* approach in selecting regions –
    - Select *best* split that minimizes RSS at each step (no look-ahead)
  + 
* Prune leaves of tree:
  + 
  + α – pruning parameter, T = # of leaves
  + (+) α = less leaves
  + Similar to lasso
* 
* Regression tree vs Classification tree –
  + Predicted response:
    - Regression tree – mean response of all observations in terminal node
    - Classification tree – most commonly occurring class of all observations in terminal node
  + Criterion for making binary splits:
    - Regression tree – RSS
    - Classification tree – *classification error rate*
* **Boostrapping** – average set of observations to reduce variance
* **Bagging** – take repeated samples from single training set and average predictions
  + Step 2
* Random forests – smaller subset of predictors to choose from for each tree (as compared to **bagging**)
  + Results in uncorrelated trees
  + Step 3
* **Boosting** – alternative to **bootstrapping**
  + Bootstrapping – grows trees from same training data
  + Boosting – grow trees sequentially , using information from previously grown trees
  + Fit small trees to residuals
  + Shrinkage parameter – λ (rate at which boosting learns)
  + Number of trees – B (select w/ CV)
  + 