

Regression Tree

Y numeric, want to predict Y using X

↳ good MSPE

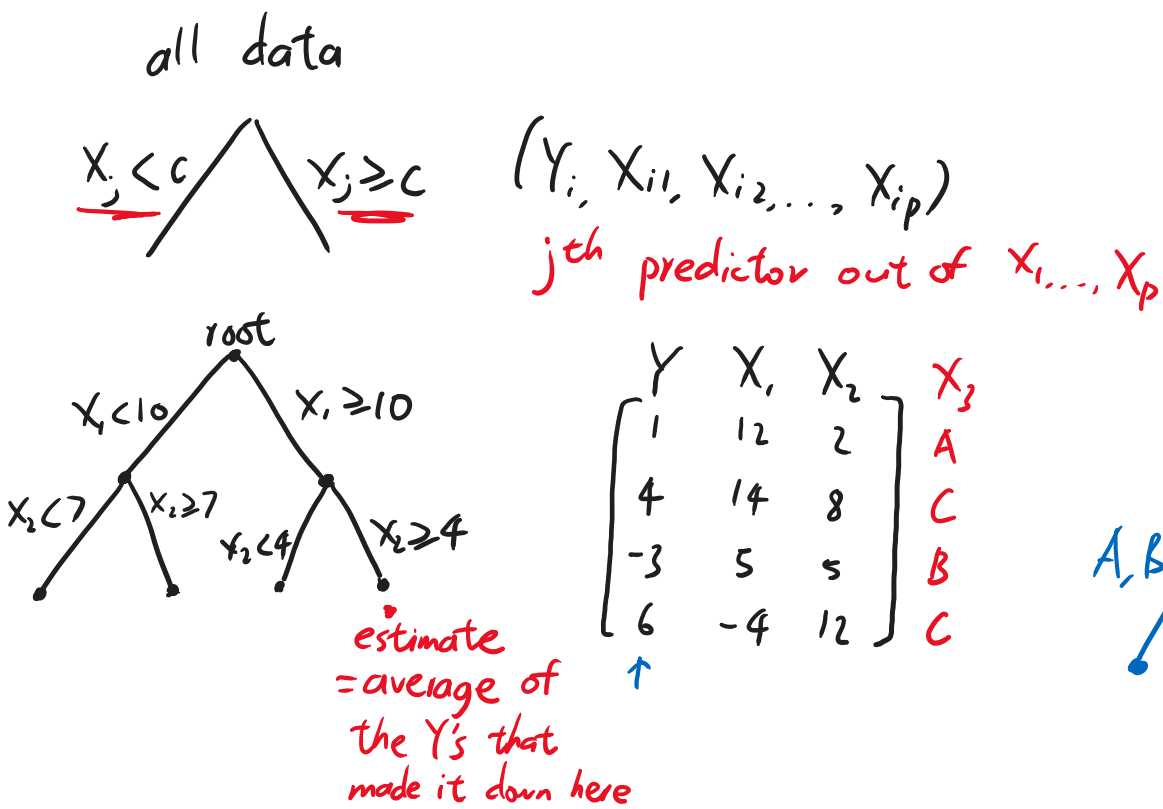
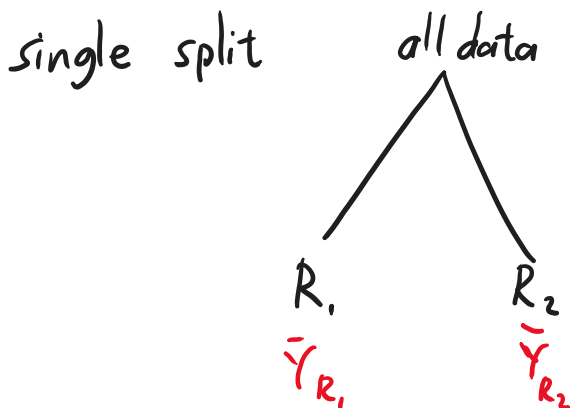
want \hat{Y} to be close to Y

Y, X_1, \dots, X_p predictors numeric or categorical

regression \rightarrow set of rules \rightarrow outcome determines \hat{Y}

simplest: no tree $\hat{Y} = \bar{Y}$, $RSS = \sum (Y_i - \bar{Y})^2$

want to split on X_1, \dots, X_p to give better predictions
split data to minimize RSS



can keep splitting data
- can get single observation in each node

↳ $\bar{Y} = Y \Rightarrow RSS = 0$

overfitting

stop growing tree \rightarrow guidelines

prune

