9.2 Bagging

Individual trees suffer from high variance.

Bootstrap aggregation (bagging) reduces variability by averaging many single tree's fits to bootstrapped replicates of the dataset.

For $b = 1, \ldots, B$

(1) Generate bootstrap dataset from $\{(\underline{x}_i, y_i)\}_{i=1}^{n}$ with replacement, to get $\{(\underline{x}_i^*, y_i^*)\}_{i=1}^{n}$.

(like paired bootstrap for regression)

(2) Fit tree to $\{(\underline{x}_i^*, y_i^*)\}$ to get $\hat{f}^b(\underline{x})$.

The bagging predictor is

$$\hat{f}(\underline{x}) = \frac{1}{B} \sum_{b=1}^{B} \hat{f}^b(\underline{x})$$

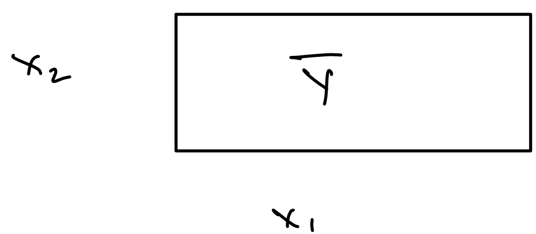
Bagged trees tend to perform better on out of sample data.

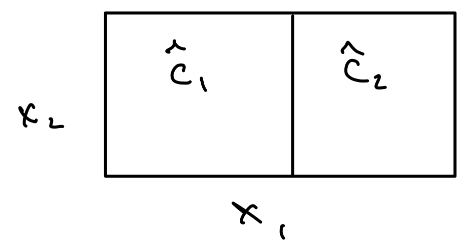
Notions of <u>variable importance</u> for bagged trees are not as straightforward.

- Track improvement in RSS for each individual feature.

For fixed feature

For each tree, descend until you hit a split over that feature, compare squared error between not splitting over that feature vs. splitting over that feature.





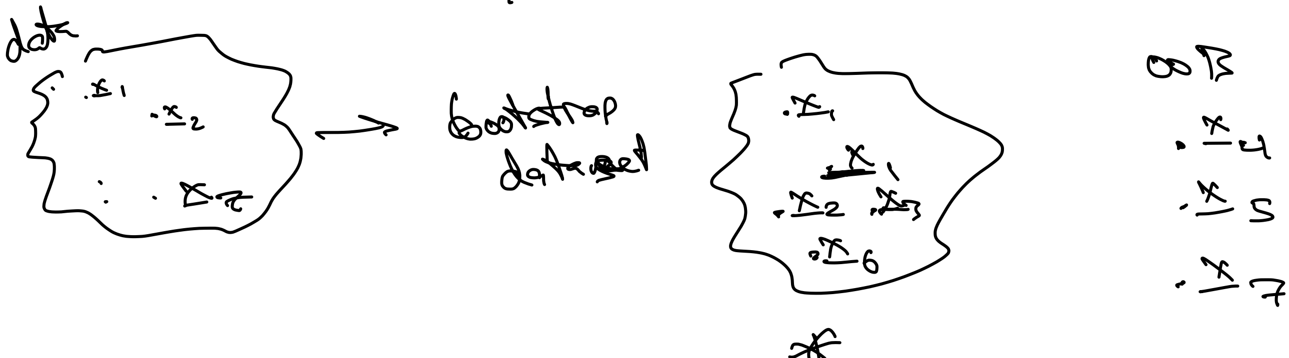
$$\sum (y_i - \bar{Y})^2 - \left(\sum_{\text{left}} (y_i - \hat{c}_1)^2 + \sum_{\text{right}} (y_i - \hat{c}_2)^2 \right)$$

Permutation-based importance

For bth bootstrapped dataset, some data points are not sampled, these are the out of bag (OOB) samples.

First, predict OOB samples based on fitted model in bag.

Then, fix feature j, permute that feature amongst all OOB samples + repredict. Do this for all $j = 1, \ldots, p$ + $b = 1, \ldots, B$, compare decrease in accuracy over the unpermuted fit.



① fit to *, predict $\hat{f}^*(\underline{x}_4), \dots, \hat{f}^*(\underline{x}_7)$

② choose a feature $\underline{\text{ex}}$ $j=2$

OOB

$$\begin{pmatrix} x_{41} \\ x_{42} \\ x_{43} \\ \vdots \\ x_{4p} \end{pmatrix} \begin{pmatrix} x_{51} \\ x_{52} \\ x_{53} \\ \vdots \\ x_{5p} \end{pmatrix} \begin{pmatrix} x_{71} \\ x_{72} \\ x_{73} \\ \vdots \\ x_{7p} \end{pmatrix} \xrightarrow[j=2]{\text{permute only}} \begin{pmatrix} x_{41} \\ x_{72} \\ x_{43} \\ \vdots \\ x_{4p} \end{pmatrix} \begin{pmatrix} x_{51} \\ x_{42} \\ x_{53} \\ \vdots \\ x_{5p} \end{pmatrix} \begin{pmatrix} x_{71} \\ x_{52} \\ x_{73} \\ \vdots \\ x_{7p} \end{pmatrix}$$

$\underline{X}_{perm,4,5,7}$

$\hat{f}^*(\underline{x}_{4perm}), \dots, \hat{f}^*(\underline{x}_{7perm})$