PS5841

Data Science in Finance & Insurance

Ensemble Learning

Yubo Wang

Autumn 2022

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Many Trees: Bagging

- Bagging: Bootstrap Aggregation
  - Generate N different bootstrapped (sample with replacement) training sets from a single data set
  - Fit a tree to each training set
  - Combine the trees to form a single predictive model

$$\hat{f}_{bag}(\boldsymbol{x}) = \frac{1}{N}\sum_{n=1}^{N}\hat{f}^{n}(\boldsymbol{x})$$

Many Trees: Random Forest

- Random Forest: Bagging with a twist that decorrelates trees
  - Each split use one optimal predictor out of a random subset (of size m) of the full set of predictors (of size p)
  - Typically, $m \approx \sqrt{p}$

Many Trees: Boosting (1)

- Boosting: Sequentially grown trees
  - Each tree is grown using information from previously grown trees
  - Each tree is fit on a modified version of the original data set
  - Learns slowly: slowly improves prediction in areas where it does not perform well (residuals).

Many Trees: Boosting (2)

$$\hat{f}(x)^{(0)} = 0;\ \ r_i^{(0)} = y_i \ \ \forall i$$

$\hat{f}^{(b)}$ is fit to the training data $(\mathbf{X}, \boldsymbol{r}^{(b-1)})$

$$\hat{f}(x)^{(1)} = \hat{f}(x)^{(0)} + \lambda \hat{f}^{(1)}(x) = \lambda \hat{f}^{(1)}(x)$$

$$r_i^{(1)} = y_i - \hat{f}(x_i)^{(1)} = r_i^{(0)} - \lambda \hat{f}^{(1)}(x_i)$$

$$\hat{f}(x)^{(2)} = \hat{f}(x)^{(1)} + \lambda \hat{f}^{(2)}(x)$$

$$= \lambda \hat{f}^{(1)}(x) + \lambda \hat{f}^{(2)}(x) = \sum_{b=1}^{2} \lambda \hat{f}^{(b)}(x)$$

$$r_i^{(2)} = y_i - \hat{f}(x_i)^{(2)} = y_i - \lambda \hat{f}^{(1)}(x_i) - \lambda \hat{f}^{(2)}(x_i)$$

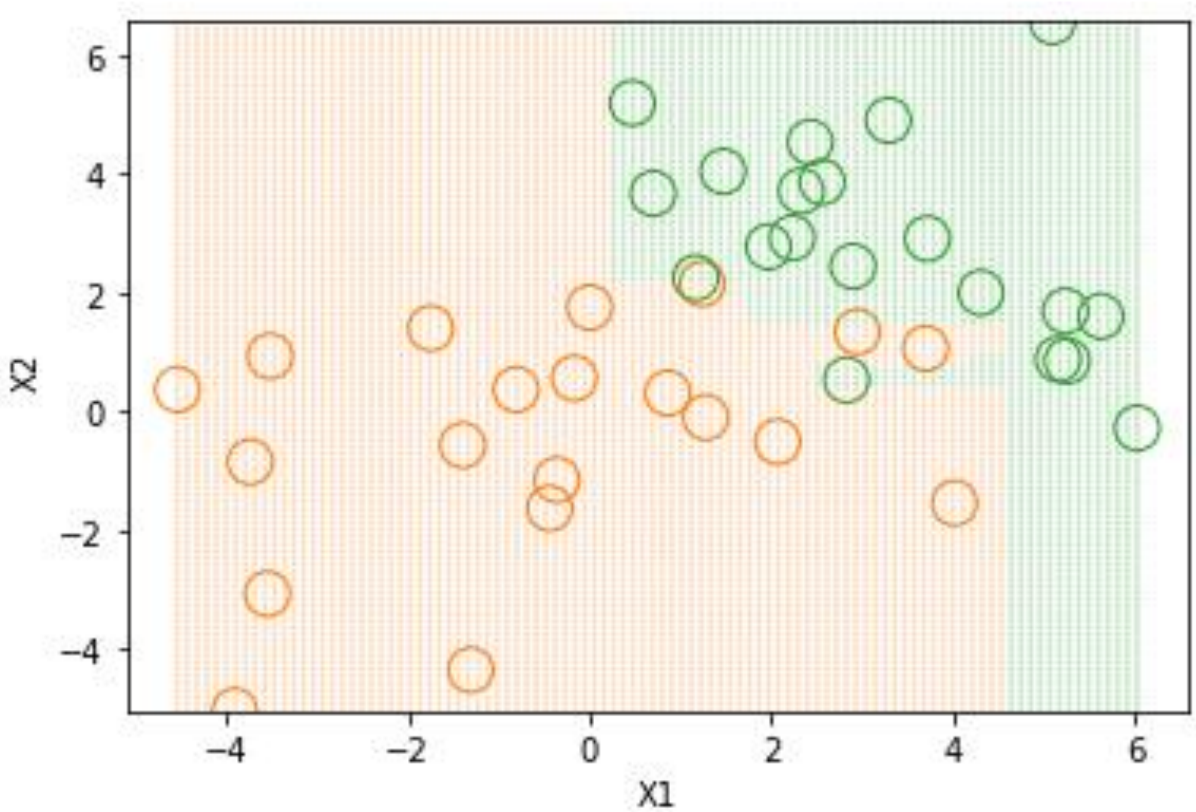
$$= r_i^{(0)} - \lambda \hat{f}^{(1)}(x_i) - \lambda \hat{f}^{(2)}(x_i) = r_i^{(1)} - \lambda \hat{f}^{(2)}(x_i)$$

Many Trees: Boosting (3)

$$\vdots$$

$$\hat{f}(x) = \hat{f}(x)^{(N)} = \sum_{n=1}^{N} \lambda \hat{f}^{(N)}(x)$$

Decision Boundary
random forest



That was