Post Midsem L6

▷ The algorithm we use depends upon the data?

→ Easy to explain
→ Easily handle qualitative predictors
→ Predictive accuracy of tree is low

→ ▪ Bagging | Boosting | Random forest | BART

BART → Bayesian Additive Regression Tree

Bagging → Multi-training set → multidecision tree (optimize each)

→ Aggregate or Average them

→ Averaging the variance from σ^2 to σ^2/n

B different bootstrapped data sets

$\hat{f}^{*b}(x) \longrightarrow$ prediction of b^{th} tree at point x

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

In ~~logistics~~ classification we take majority vote from all B trees.

Random forest

Total = 'p' predictors

→ We use some attributes $m = \sqrt{P}$

random subset of predictors

□ Can we use the result from 2 egg CS Problem for taking 'm' value instead of $\sqrt{P}$?

→ to optimal

Boosting

→ Similar way except that trees are grown sequentially, each tree using information from previously grown trees.

residual

1. $\hat{f}(x) = 0$, $r_i = y_i \ \forall i$

2. for $b = 1, 2, \dots B$ repeat:

2.1 → tree $\hat{f}^b$ with d splits ($d+1$ terminal nodes) to training data (X, r)

2.2 update $\hat{f}$ by shrunken version of the new tree

$$\hat{f}(x) \leftarrow \hat{f}(x) + \lambda \hat{f}^b(x)$$

2.3 residual updation

$$r_i \leftarrow r_i - \lambda \hat{f}^b(x_i)$$

3. output the boosted mode

$$\boxed{\hat{f}(x) = \sum_{b=1}^{B} \lambda \hat{f}^b(x)}$$

typical values 0.01 to 0.001