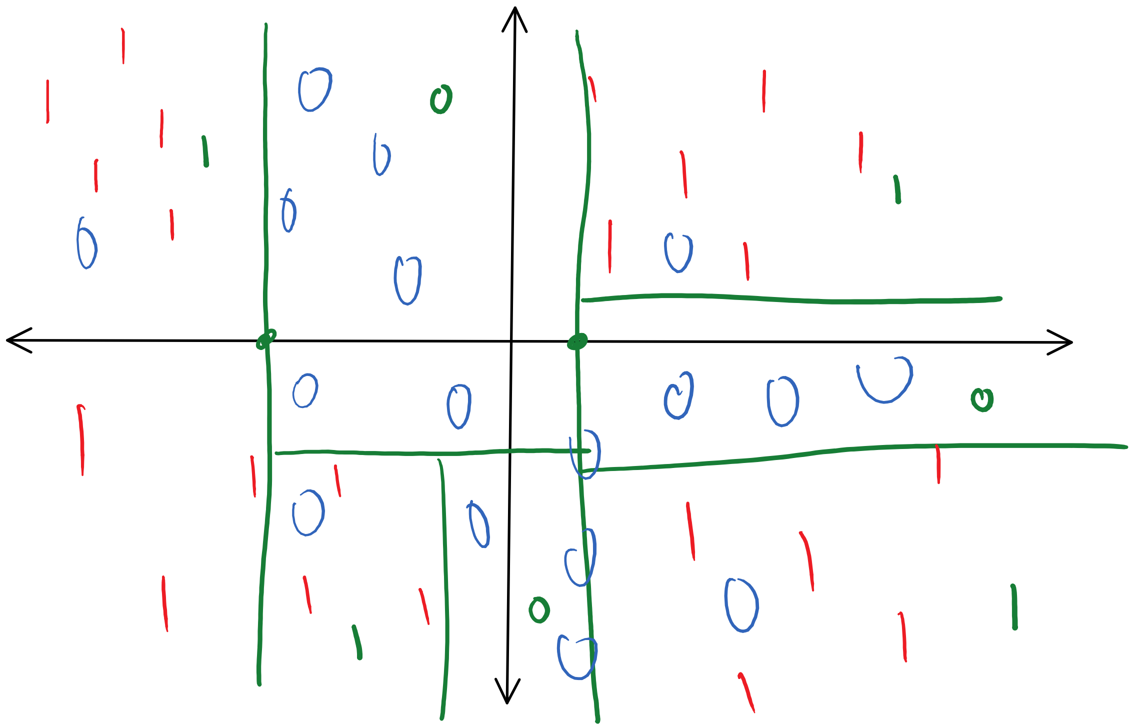
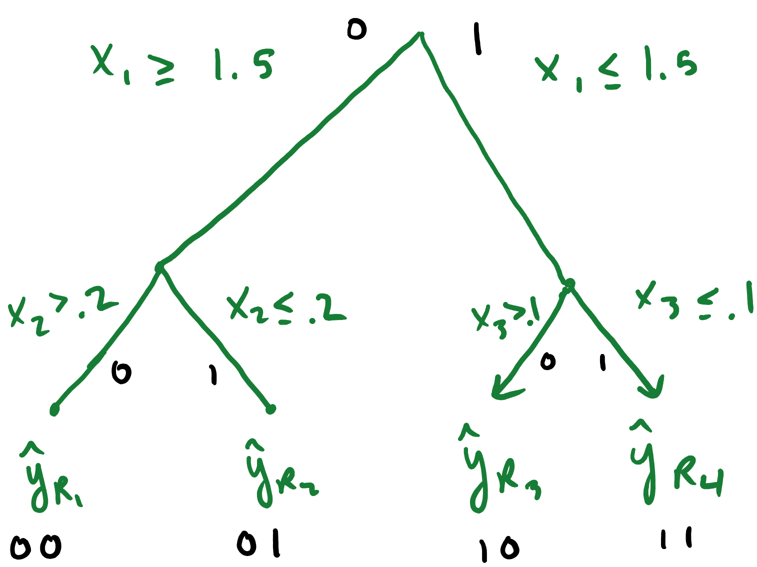
Decision Trees

Monday, May 13, 2019 1:27 PM



ex covariates x_1, x_2, x_3



predict

$$\hat{f}(x) = \sum_{m=1}^{r} \hat{y}_{R_m} \mathbb{1}\{x \in R_m\}$$

r : # of region ≡ leaf nodes

fit (regression)

(1) Try split into halfplanes

variable ↓ ↙ split

$$R_1(j, s) = \{x : x_j \leq s\}$$

$$R_2(j, s) = \{x : x_j > s\}$$

(2) $\min_{j,s}\left[\min_{\hat{y}_1}\sum_{x_i\in R_1(j,s)} \ell(y_i,\hat{y}_1) + \min_{\hat{y}_2}\sum_{x_i\in R_2(j,s)} \ell(y_i,\hat{y}_2)\right]$

(3) Recurse (1-2) w/in R_1, R_2

<u>Classification</u> multiclass $k \in \{1,\dots,K\}$

- use $\ell(y_i,\hat{y}) = \mathbb{1}\{y_i \neq \hat{y}\}$

  calculate emp. prob: $\hat{p}_{mk} = \frac{1}{|\{x_i : x_i \in R_m\}|}\sum_{x_i\in R_m} \mathbb{1}\{y_i = k\}$

  $\hat{y}_i = \arg\max_k \hat{p}_{mk}$ for $x_i \in R_m$.

- Gini purity: $\sum_{k\neq k'} \hat{p}_{mk}\hat{p}_{mk'} = \sum_k \hat{p}_{mk}(1-\hat{p}_{mk})$

- Cross-entropy: $-\sum_k \hat{p}_{mk}\log\hat{p}_{mk}$

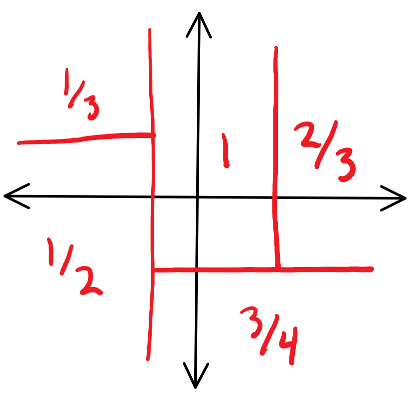
Simplest stopping criteria - fix a depth of tree

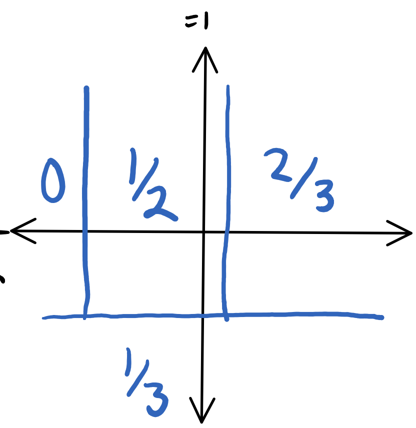
Random Forests

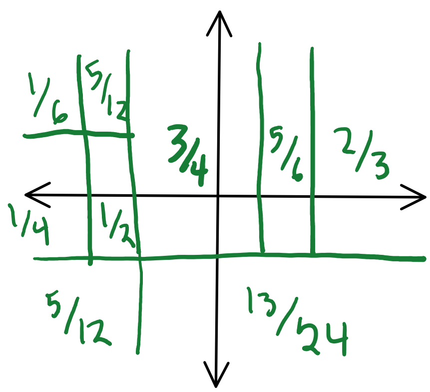
Monday, May 13, 2019 1:51 PM

Aggregation: $\hat{f}(x) = \sum_{b=1}^{B} \overset{\text{weight}}{\alpha_b} \underset{\text{base classifier}}{\hat{f}_b(x)}$

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

$\frac{1}{2}$ 

$+ \frac{1}{2}$ 

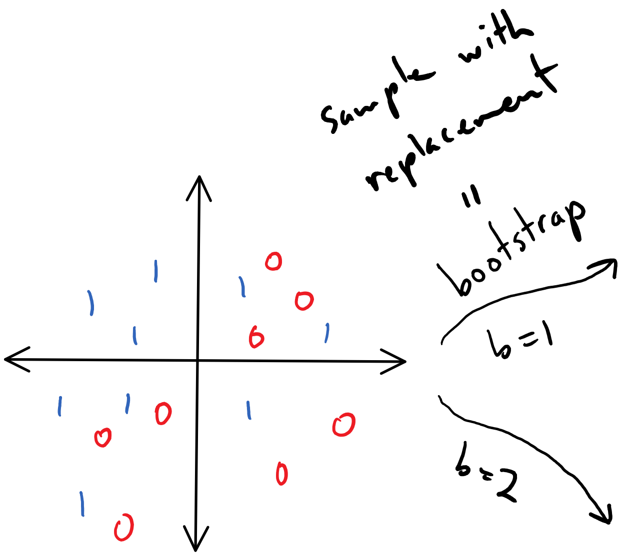
$=$ 

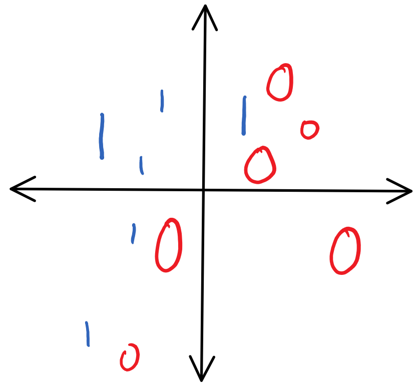
Bias: $\mathbb{E}[\hat{f}_b(x) - Y]$

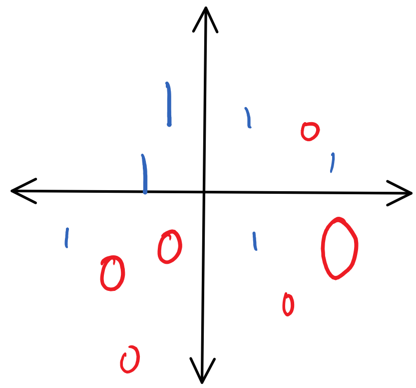
$$\mathbb{E}[\hat{f}(x) - Y] = \frac{1}{2}\Big[\mathbb{E}[\hat{f}_1(x) - Y] + \mathbb{E}[\hat{f}_2(x) - Y]\Big]$$

Var: suppose $\hat{f}_1 \perp\!\!\!\perp \hat{f}_2$ $\quad \mathbb{V}(\hat{f}(x)) = \mathbb{V}\left[\frac{1}{2}\hat{f}_1(x) + \frac{1}{2}\hat{f}_2(x)\right]$

$$= \frac{1}{4} \sum_b \mathbb{V} \hat{f}_b(x)$$







Bagging

- ▷ bootstrap $D_1, \dots, D_B$
- ▷ train $\hat{f}_b$ of D_b $\forall b$
- ▷ average $\hat{f}_b$

How can we reduce bias?

- ▷ Choose classifier w/ low bias (large depth dec. tree)

How can we reduce V?

- ▷ reduce subsampling rate / increase B

Random Forests

For $b = 1, \dots, B$

Bootstrap m rows w/ repl.

Sample d predictors and grow (nearly) max-depth tree on $m \times d$ dataset

Save trees and sample frequencies

Predict w/ $\frac{1}{B} \sum_{b=1}^{B} \hat{f}_b(x)$

Out-of-bag error : for each x_i, y_i

Let B_i be the bootstrap samples w/ out i

Predict w/ $\hat{y}_i^{o} = \frac{1}{|B_i|} \sum_{b \in B_i} \hat{f}_b(x_i)$

Compute OOB error : $\frac{1}{n} \sum_{i=1}^{n} \ell(y_i, \hat{y}_i^{o})$