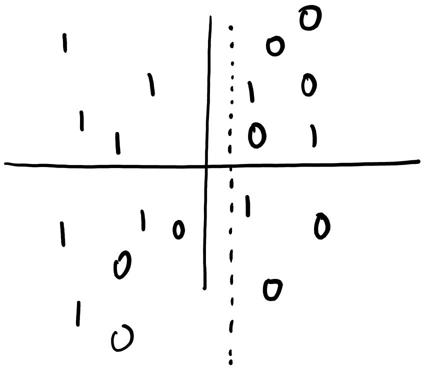
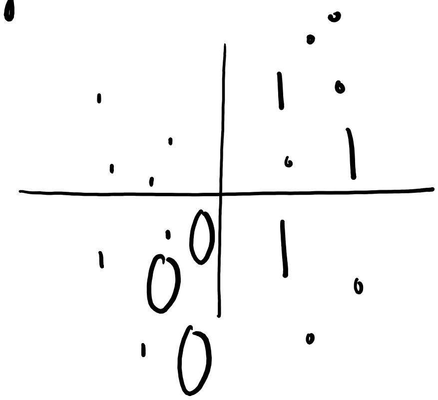
Boosting

Wednesday, May 15, 2019 2:06 PM

Bagging: Random data augmentation

Boosting: Adaptive data augmentation





Adaboost ($y_i \in \{-1, 1\}$)

Idea: increase weights on misclassified points

Init $w_i = \frac{1}{n}$ $i = 1, \dots, n$

For b in $1, \dots, B$:

Fit classifier to training data ($\hat{y}_b$) using weights w_i

Compute weighted misclassification error ↑ produces f_b

$$\varepsilon_b = \frac{\sum_i w_i \mathbb{1}\{y_i \neq \hat{y}_{bi}\}}{\sum_i w_i}$$

Set $\alpha_b = \log \frac{1 - \varepsilon_b}{\varepsilon_b}$

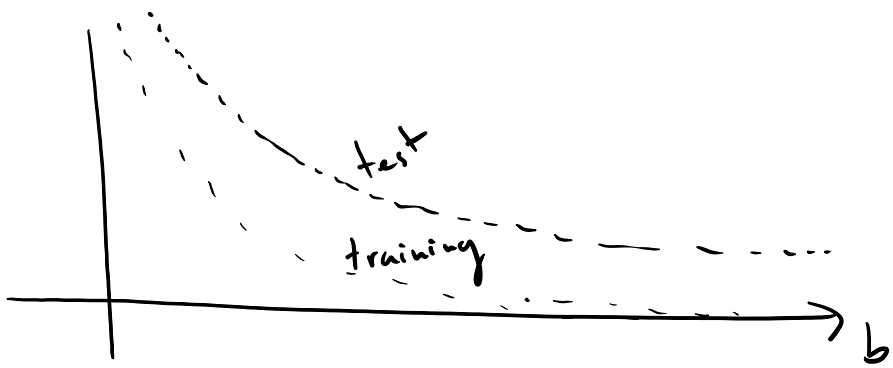
Update weights $w_i \leftarrow w_i \cdot \exp\left(\alpha_b \mathbb{1}\{y_i \neq \hat{y}_{bi}\}\right)$

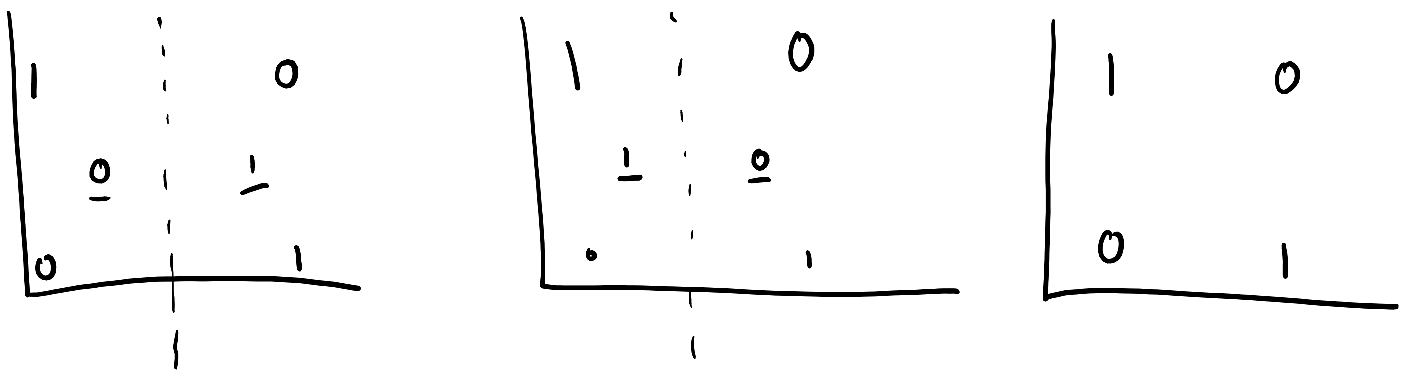
$f(x) = \sum_{b=1}^{B} \alpha_b f_b(x)$ final classifier: $\hat{y}(x) = \text{sign}(f(x))$

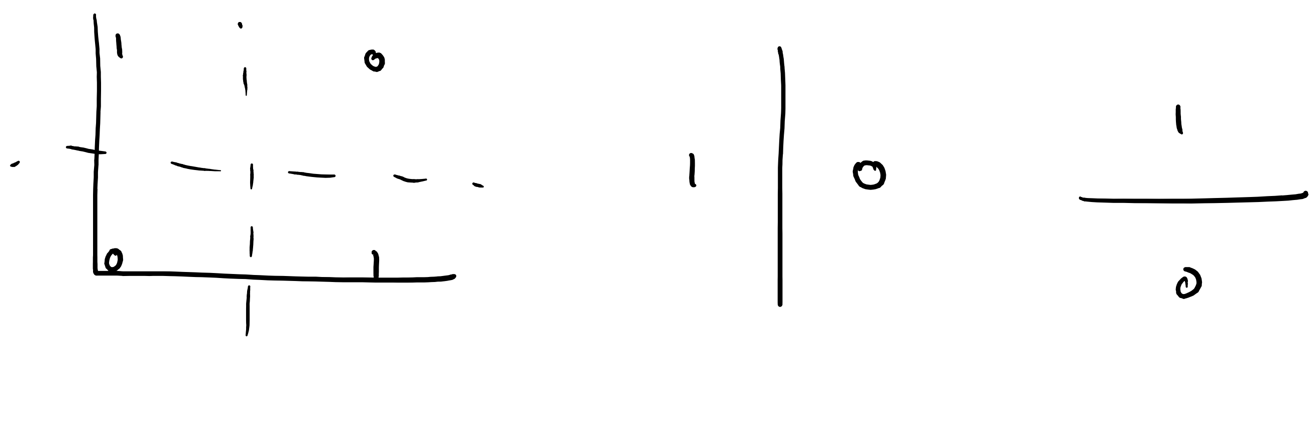
<u>def</u> weak learner: base classifier used in boosting

$\frac{1}{n}\sum_i \mathbb{1}\{\hat{y}_i \neq y_i\}$

prop The training error of adaboost is bounded by $e^{-2\sum_{b=1}^{B} \gamma_b^2}$ w/ $\gamma_b = \frac{1}{2} - \varepsilon_b$







Gradient Boosting

Wednesday, May 15, 2019 2:07 PM

Want to fit a given loss, ℓ, but we want to use regression trees

<u>Alg</u>

$$F_0(x) = \underset{\gamma}{\arg\min} \sum_{i=1}^{n} \ell(y_i, \gamma)$$

For b in $1, \dots, B$:

$$\tilde{y}_{bi} = -\left.\frac{\partial \ell(y_i, \gamma)}{\partial \gamma}\right|_{\gamma = F_b(x_i)} \quad \forall i = 1, \dots, n$$

↳ pseudo residuals (may be subgrad)

$$\{R_{b\ell}\}_{\ell=1}^{L} = \text{Decision Tree}\left(\{\tilde{y}_{bi}, x_i\}_{i=1}^{n}, \ \# \text{ of leaves} = L\right)$$

(regions)

$$\gamma_{b\ell} = \underset{\gamma}{\arg\min} \sum_{i : x_i \in R_{b\ell}} \ell(y_i, F_{b-1}(x_i) + \gamma)$$

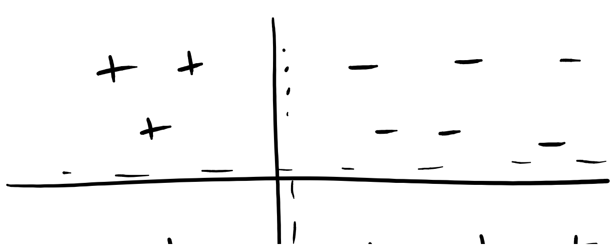
$$F_b(x) = F_{b-1}(x) + \nu \sum_{\ell=1}^{L} \gamma_{b\ell} \mathbb{1}\{x \in R_{b\ell}\} \quad 0 \le \nu \le 1$$

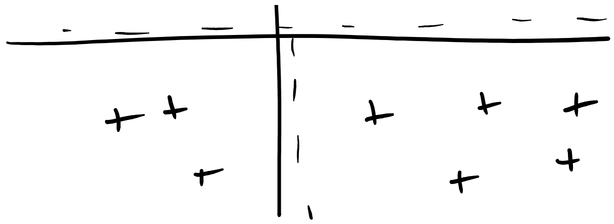
<u>eg</u> $\ell(y_i, \gamma) = (1 - y_i\gamma)_+ \quad y_i \in \{-1, 1\}$

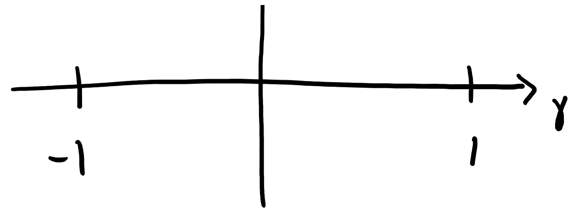
$$\partial \ell(y_i, \gamma) \ni -y_i \mathbb{1}\{1 > y_i\gamma\} = -\tilde{y}_{bi}$$

$$\tilde{y}_{bi} = \begin{cases} y_i, & y_i \cdot \gamma < 1 \\ 0, & \text{otherwise} \end{cases}$$

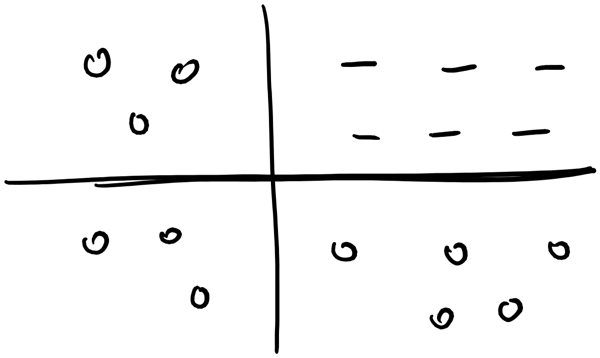
$$\sum_i (1 - y_i\gamma)_+ = \underbrace{N_+}_{\#\{i : y_i = 1\}} \cdot (1 - \gamma)_+ + N_-(1 + \gamma)_+$$



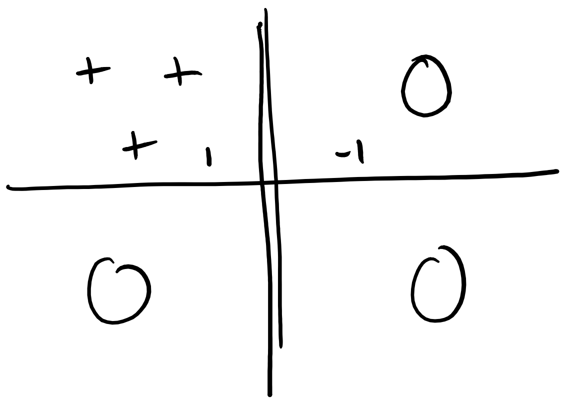




$\Rightarrow$ minimize hinge $\equiv$ majority vote



$\tilde{y}_1$



$\tilde{y}_2$

Stochastic Gradient Tree Boosting

For b in $1, \ldots, B$:

S is a random sample of $1, \ldots, n$ of size m w/out replacement

$$\tilde{y}_{bi} = -\frac{\partial l(y_i, \gamma)}{\partial \gamma}\Bigg|_{\gamma = F_{b-1}(x)} \quad \forall i \in S$$

▷ eval Dec. tree over $\{x_i, \tilde{y}_{bi}\}_{i \in S}$

↳ $\{R_{bl}\}$

$$\gamma_{bl} = \arg\min_{\gamma} \sum_{i \in S:\, x_i \in R_{bl}} \ell(y_i, F_{b-1}(x_i) + \gamma)$$

$$F_b(x) = F_{b-1}(x) + \nu \sum_{l=1}^{L} \gamma_{bl} \mathbb{1}\{x \in R_{bl}\}$$