AdaBoost

Wednesday, 12 April 2023   3:54 PM

Boosting:- Combining 'base' classifiers to form a "committee" whose performance is better than any of base classifiers.

It works even for "Weak Learners" i.e. learners whose error is slightly better than $1/2$

Idea:- Train classifiers in sequence by weighted data sets with weights updated in every iteration.

Setting: Data : $\{(x_1, y_1), (x_2, y_2) \ldots (x_n, y_n)\}$

$$y_i \in \{-1, +1\}$$

$$w_i = \frac{1}{n} \quad \forall i$$

Learner which can work with weighted data points to give

$$h(x) \in \{-1, +1\}$$

Ada Boost

1. Initialize $\{w_i\}$ as $w_i^{(1)} = \frac{1}{n}$ for $i = 1, 2, \ldots n$
2. For $t = 1, \ldots T$

   a) Fit $h_t(x)$ to training data by minimizing

   $$J_t = \sum_{i=1}^{n} w_i^{(t)} \underbrace{I(h_t(x_i) \neq y_i)}_{0/1 \text{ loss}} \quad \text{— A}$$

   b) Find

   $$\epsilon_t = \frac{\sum_{i=1}^{n} w_i^{(t)} I(h_t(x_i) \neq y_i)}{\sum_{i=1}^{n} w_i^{(t)}}$$

   set $\alpha_t = \ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right)$ — B

   c) Update weights

   $$w_i^{(t+1)} = w_i^{(t)} \exp\left(\alpha_t I(h_t(x_i) \neq y_i)\right) \quad \text{— C}$$

3. Final Model

$$H_T(x) = \text{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right)$$

Boosting as sequential minimization of exponential error function

Consider the error

$$E = \sum_{i=1}^{n} \exp(-y_i f_t(x_i))$$

where

$$f_t(x) = \frac{1}{2} \sum_{l=1}^{t} \alpha_l h_l(x)$$

Goal: Minimize E w.r.t both α_l and $h_l(x)$ (parameters)

Assume $h_1(x), \ldots h_{t-1}(x)$ and $\alpha_1, \ldots \alpha_{t-1}$ are fixed

Minimize only w.r.t $\alpha_t, h_t(x)$

$$E = \sum_{i=1}^{n} \exp\left\{-y_i f_{t-1}(x_i) - \frac{1}{2} y_i \alpha_t h_t(x_i)\right\}$$

$$= \sum_{i=1}^{n} w_i^{(t)} \exp\left\{-\frac{1}{2} y_i \alpha_t h_t(x_i)\right\}$$

$w_i^{(t)} = \exp(-y_i f_{t-1}(x_i))$ can be considered constants.

Let C_t = set of data points correctly classified by $h_t(x)$

M_t = set of remaining misclassified points

$$E = e^{-\alpha_t/2} \sum_{i \in C_t} w_i^{(t)} + e^{\alpha_t/2} \sum_{i \in M_t} w_i^{(t)}$$

$$= \left(e^{\alpha_t/2} - e^{-\alpha_t/2}\right) \sum_{i=1}^{n} w_i^{(t)} I(h_t(x_i) \neq y_i) + \underline{e^{-\alpha_t/2} \sum_{i=1}^{n} w_i^{(t)}}$$

→ Minimizing w.r.t $h_t(x)$ second term is constant

Equivalent to minimizing A

→ Similarly minimizing w.r.t α_t we get B

Now $w_i^{(t+1)} = w_i^{(t)} \exp\left\{-\frac{1}{2} y_i \alpha_t h_t(x_i)\right\}$

Also $y_i h_t(x_i) = 1 - 2 I(h_t(x_i) \neq y_i)$

in next iter

$$w_i^{(t+1)} = w_i^{(t)} \exp(-\alpha_t/2) \exp\left(\alpha_t I(h_t(x_i) \neq y_i)\right)$$

Now $\exp(-\alpha_t/2)$ is independent of i and it can be discarded giving us C

Finally, new test points are classified using sign of combined classifier. As $1/2$ does not impact sign it can be removed.