AdaBoost

Oliver Zhao

1 Description

AdaBoost was developed for two-class classification problems. It is a sequential ensemble method that fits a sequence of weak classifiers to modified version of the training data. The modifications involve giving more weight to misclassified points, hence having future weak classifiers focused more on the misclassified data points.

After looking at the algorithm in the next subsection, it becomes clear that AdaBoost is a form of linear regression in which a linear combination of weak classifiers is used to create the overall AdaBoost classifier.

2 Algorithm

Description: Here we have a dataset $\{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\}$, where $x_i \ \epsilon \ \mathscr{X}$ describes the features and $y_i \ \epsilon \ \{-1, +1\}$ describes the class. $\mathbf{w_t}$ is the weight of the m samples $\{w_t(1), w_t(2), \ldots, w_t(m)\}$ at the t^{th} iteration. h_t is the weak learner that is generated at the t^{th} iteration, which has an associated weight of α_t and error of ϵ_t. $h_t(x_i)$ is the predicted class of the i^{th} datapoint by the weak learning generated at iteration t.

---

1. Initialize the observation weights $w_1(i) = 1/m$, $i = 1, ..., m$.
2. For $t = 1, \ldots, T$, repeat the following steps:
   (a) The the weak learner h_t to the training data, using observation weights $w_1(i)$.
   (b) Compute the misclassification error of the weak learning h_t:

$$\epsilon_t = \frac{\sum_{i=1}^{n} w_t(i) I[y_i \neq h_t(x_i)]}{\sum_{i=1}^{n} w_t(i)}, \tag{2.1}$$

Where

$$I[y_i \neq h_t(x_i)] = 1, \tag{2.2}$$

Only when $y_i \neq h_t(x_i)$.

(c) Compute the weight for the weak learner, or α_t:

$$\alpha_t = ln\left(\frac{1-e_t}{e_t}\right). \tag{2.3}$$

(d) Update the weights for the next iteration, $w_{t+1}(i)$:

$$w_{t+1}(i) = \frac{w_t(i)exp(-\alpha_t y_i h_t(x_i))}{Z_t}, \tag{2.4}$$

Where Z_t is a somewhat arbitrary normalization factor such that $\mathbf{w_{t+1}}$ has a reasonable distribution.

3. Output the final hypothesis of the ensemble of weak learners:

$$H(\mathbf{x}) = sgn\left(\sum_{t=1}^{T} \alpha_t y_i h_t(x_i)\right). \tag{2.5}$$

3 Derivation

After the $(t-1)^{th}$ iteration, the AdaBoost classifier is a linear combination of the weak classifiers where

$$C_{t-1}(x_i) = \alpha_1 h_1(x_i) + \cdots + \alpha_{t-1} h_{t-1}(x_i). \tag{3.1}$$

After the t^{th} iteration, we want to extend the equation above to a better boosted classifier by adding another weak classifier h_t with another weight α_t, where

$$C_t(x_i) = C_{t-1}(x_i) + \alpha_t h_t(x_i). \tag{3.2}$$

We define the total error E of C_t as the sum of its exponential loss on each data point:

$$E = \sum_{i=1}^{N} e^{-y_i C_t(x_i)} = \sum_{i=1}^{N} e^{-y_i C_{t-1}(x_i)} e^{-y_i \alpha_t h_t(x_i)} \tag{3.3}$$

Letting $w_1(i) = 1$ and $w_t(i) = e^{-y_i C_{t-1}(x_i)}$ for $t > 1$, we have:

$$E = \sum_{i=1}^{N} w_t(i) e^{-y_i \alpha_t h_t(x_i)}. \tag{3.4}$$

We can split the summation between data points that are correctly classified by h_t and those that are misclassified by h_t, such that

$$E = \sum_{y_i = k_t(x_i)}^{N} w_t(i) e^{-\alpha_t} + \sum_{y_i \neq k_t(x_i)}^{N} w_t(i) e^{-\alpha_t}. \tag{3.5}$$

To find the desired weight α_t that minimizes E, we just differentiate the error function with respect to the weight of the weak classifier and set it equal to zero:

$$\frac{dE}{d\alpha_t} = \frac{d\left(\sum_{y_i = k_t(x_i)}^{N} w_t(i) e^{-\alpha_t} + \sum_{y_i \neq k_t(x_i)}^{N} w_t(i) e^{-\alpha_t}\right)}{d\alpha_t} = 0. \tag{3.6}$$

This yields the solution:

$$\alpha_t = \frac{1}{2} ln \left(\frac{\sum_{y_i = k_t(x_i)}^{N} w_t(i)}{\sum_{y_i \neq k_t(x_i)}^{N} w_t(i)} \right) \tag{3.7}$$