AdaBoost

Paul Bouyé

This sheet is based on a video of Joshua Starmer on AdaBoost [1].

Introduction to Random Forests

Decision trees are prone to overfitting

1. BAGGING

   → train B decision trees and get the opinion of the majority

2. RANDOM SUBSPACES

   1 issue: highly correlated trees → solution: at each split, only consider a subset of features

   Find importance by permuting features (shuffle data in a feature and check for accuracy changes)

Basically, 1 ↔ randomly sampling the rows and 2 ↔ randomly sampling the columns.

AdaBoost

→ stump: decision tree with one node and two leaves

⇒ one variable ⇒ "weak learner"

→ In a random forest, each tree has an equal vote on the final classification. In a forest of stumps made with AdaBoost:

- some stumps have more say in the final classification than others
- the order of the stumps matter

EXAMPLE:

Chest Pain	**Blocked Arteries**	**Patient Weight**	**Heart Disease**	**Sample Weight** ①
Y	Y	205	Y	1/8
N	Y	180	Y	1/8
Y	N	210	Y	1/8
Y	Y	167	Y	1/8
N	Y	156	N	1/8
N	Y	125	N	1/8
Y	N	168	N	1/8
Y	Y	172	N	1/8

Table 1: data of patient data with associated attributes and sample weights

We can try to predict whether a patient has heart disease given a single column.

- chest pain $\rightarrow$ 3 incorrect
- blocked arteries $\rightarrow$ 4 incorrect
- weight > 176 (minimizes least squares) $\rightarrow$ $\boxed{\text{1 incorrect}}$

$\Rightarrow$ weight > 176 will make up the first stump

$$\varepsilon_1 = total\ error = \sum_{error} weight = \frac{1}{8}$$

Amount of say: $\alpha_1 = \frac{1}{2}\log\left(\frac{1-\varepsilon_1}{\varepsilon_1}\right) = \frac{1}{2}\log 7 \simeq 0.97$

α_1 being the weight of the weak learner

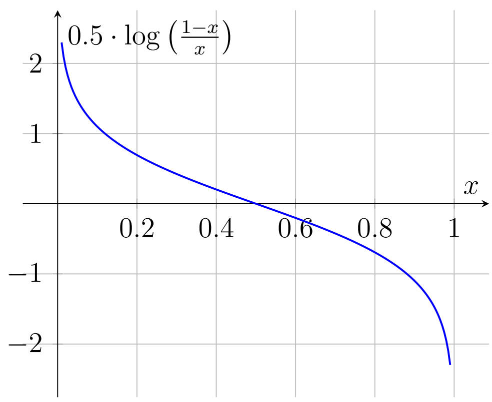


Figure 1: Amount of say over the total error

- **INCREASE** the sample weight for <u>incorrectly</u> classified samples
- **DECREASE** the sample weight for <u>correctly</u> classified samples

- Update weights:

$$D_{t+1}(i) = \frac{D_t(i)e^{-\alpha_t y_i h_t(x_i)}}{Z_t} \qquad (Z_t \text{ being a NC such that } \sum_{i=1}^{m} D_{t+1}(i) = 1)$$

$$\Rightarrow \text{correctly classified: } y_i h_t(x_i) = \begin{cases} 1 \cdot 1 = 1 \\ (-1) \cdot (-1) = 1 \end{cases} \quad \Rightarrow e^{-\alpha_t} < 1 \Rightarrow D \downarrow$$

$$\Rightarrow \text{incorrectly classified: } y_i h_t(x_i) = \begin{cases} (-1) \cdot 1 = -1 \\ 1 \cdot (-1) = -1 \end{cases} \quad \Rightarrow e^{\alpha_t} > 1 \Rightarrow D \uparrow$$

- Here: $D_2(4) = \frac{D_1(4)e^{\alpha_1}}{Z_1} \simeq \frac{0.33}{Z_1}$

  For $i \neq 4$, $D_2(i) = \frac{D_1(i)e^{-\alpha_1}}{Z_1} \simeq \frac{0.05}{Z_1}$. $\sum_{i=1}^{m} D_2(i) = 1 \Rightarrow \frac{0.33}{Z_1} + 7 \cdot \frac{0.05}{Z_1} = 1 \Rightarrow Z_1 = 0.68$

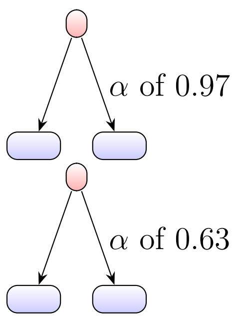
Sample Weight ②
0.07
0.07
0.07
0.49
0.07
0.07
0.07
0.07

$\rightarrow$ In theory, we could use those weights to calculate weighted Gini indexes to determine which variable should split the next stump.

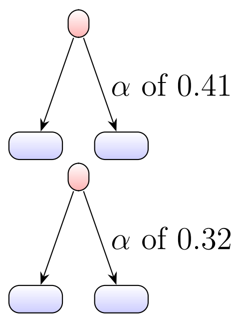
$\rightarrow$ Alternatively, we can make a new collection of samples by sampling according to the samples weights ($\Rightarrow$ there can be duplicates), then we would reset samples weights to $\frac{1}{m}$

- Finally:

Patient has heart disease



Patient does not have heart disease



$$0.97 + 0.63 = 1.60 \text{ and } 0.41 + 0.32 = 0.73 \Rightarrow \text{classified as "Has heart disease"}$$

Combination of weak learners $\rightarrow$ crowd wisdom

References

[1] StatQuest with Josh Starmer. *AdaBoost, Clearly Explained*. https://www.youtube.com/watch?v=LsK-xG1cLYA, 2019. YouTube Video.