

Boosting

Like bagging boosting is also an ensemble method. where multiple weak learners works together to build a strong learners. Bagging was a variance reduction technique whereas boosting is an bias-reduction technique. In boosting multiple weak-learners are trained sequentially. Each of the weak learners focuses more on the data points which were incorrectly guessed by previous weak learner. For example. Let's assume a classifier '0' incorrectly labeled the data point 2, 5, 7, 9. so the next classifier will focus more on the data point 2, 5, 7, 9. For this reason boosting cannot run in parallel. A weak-classifier needs to see the results of previous weak classifier. There are multiple ways on focusing to some specific classifier. For example.

(i) Duplicate the training example that needs to be focused.

(ii) Assign weights to each of the training data. Increase weight that needs to be

focused, decrease otherwise.

AdaBoost

Each of the weak classifiers will also have their individual weight (α). This will be calculated based on their computed weighted error (r).

Training:-

Let,

$G_m \rightarrow$ weak estimator m .

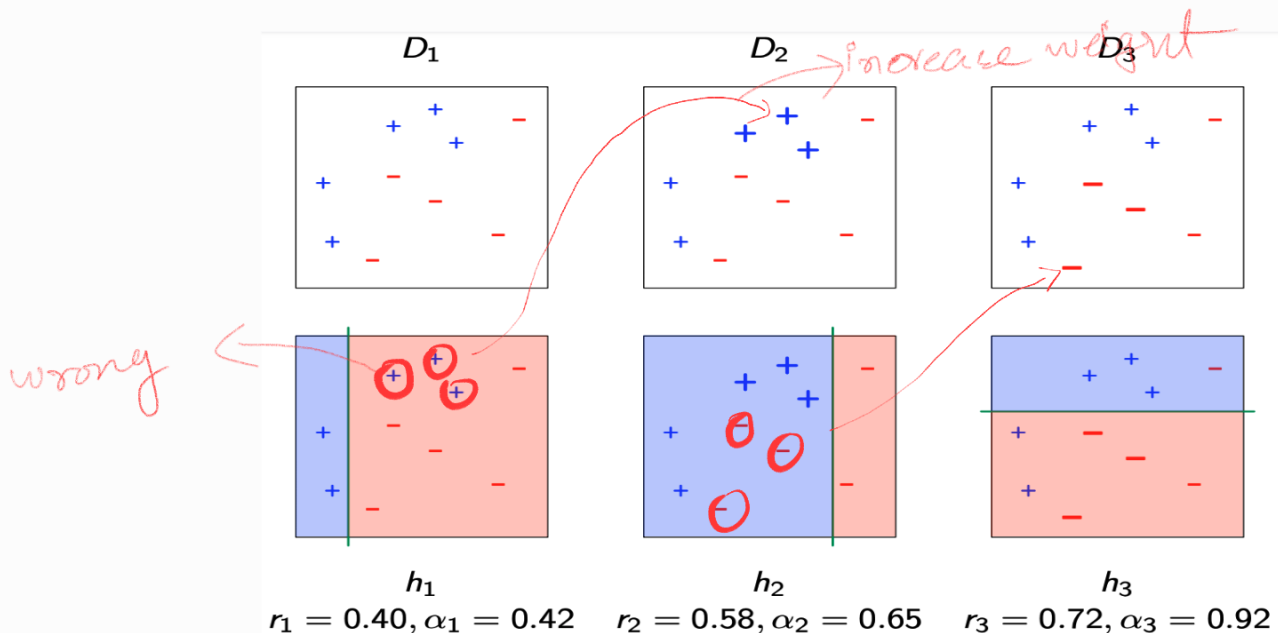
$w_i \rightarrow$ Weight of the training data i .

So,

$$r_m = \frac{\sum_i w_i * 1(y_i \neq G_m(x_i))}{\sum_i w_i}$$

$$\alpha_m = \log\left(\frac{1 - r_m}{r_m}\right)$$

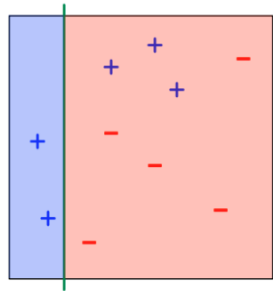
$$w_i \leftarrow w_i * \exp(\alpha_m * 1(y_i \neq G_m(x_i)))$$



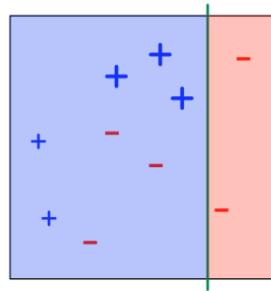
Prediction:-

For making prediction we will combine the result of all the weak estimators And we will use relative weight for each estimator.

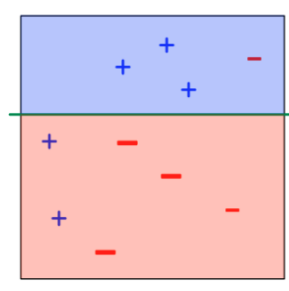
$$f(x) = \text{sign} \left(\sum_m \alpha_m G_m(x) \right)$$



h_1
 $\alpha_1 = 0.42$



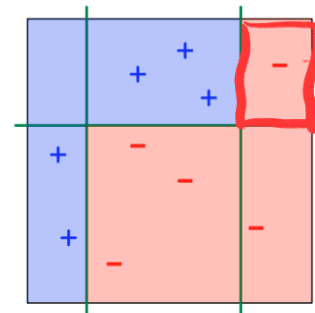
h_2
 $\alpha_2 = 0.65$



h_3
 $\alpha_3 = 0.92$

Final classifier:

$$\text{sign} (0.42h_1(x) + 0.65h_2(x) + 0.92h_3(x))$$



→ h_3 has more weight so it will overrule h_2 's decision.

