Boosting and Bagging

An ensemble method is a technique that combines the predictions from multiple algorithms together to make more accurate predictions than any individual model.

Bootstrapping is a sampling technique in which we create subsets of observations from the original data with replacement.

Bagging (or bootstrap aggregating) uses these subsets to get a fair idea of the entire distribution.

If the error rate of a classifier is almost equal to or a bit less than 0.5, it is said to be a weak classifier. If it is near 0, it is said to be a strong classifier. In boosting, we can take a bunch of these samples and make them "vote" to make them a strong classifier currently. This mainly works when the chosen hypotheses are wrong and the examples on which they are wrong are isolated from each other.

If these regions are not isolated, we need to test if the region is large enough to give a worse accuracy on the model than an individual test.

In Adaboost, we choose hypotheses from a space of possible hypotheses and make decision tree stumps out of them, each test being a possible classifier. We say that these are a weighted group of experts, where the prediction is given by weighting the predictions of each expert in the group.

$$H(x) = sign\left(\alpha^1 h^1 + \alpha^2 h^2 + \alpha^3 h^3 + \ldots\ldots + \alpha^n h^n \right)$$

The Adaboost algorithm is given as follows

1. Initialise all the weights w to 1/N.
2. Pick a hypothesis h at time t that minimises E such that

$$E = \frac{\sum_{cases\ where\ pred\ is\ wrong} w_i}{N}$$

3. Pick alpha for h as

$$\alpha^t = \frac{1}{2}\ln\left(\frac{(1 - E^t)}{E^t}\right)$$

4. Then update the weights of all the samples by the rules

$$w_i^{t+1} = \frac{w_i^t}{Z} e^{-\alpha^t h^t(x) y(x^t)}$$

This can be simplified as

$$w_i^{t+1} = \frac{w_i^t}{Z} \cdot \begin{cases} \sqrt{\dfrac{E^t}{1 - E^t}} & \text{if prediction is correct} \\ \sqrt{\dfrac{(1 - E^t)}{E^t}} & \text{if prediction is wrong} \end{cases}$$

But now,

$$\sqrt{\frac{E^t}{1 - E^t}} \sum_{correct} w_i + \sqrt{\frac{1 - E^t}{E^t}} \sum_{wrong} w_i = Z$$

Which gives $Z = 2\sqrt{E^t(1 - E^t)}$

This gives a simpler update rule for w as

$$w_i^{t+1} = \begin{cases} \dfrac{\frac{w_i^t}{2}}{1 - E} & \text{if correct} \\ \dfrac{\frac{w_i^t}{2}}{E} & \text{if wrong} \end{cases}$$

This method does not overfit as the learners' errors are exaggerated onto the consequent learners.

Note: We increase the weights if the examples are