MLSM exam 20.12.2021 - essay
Name and student number: Tat Hong Duong Le - 894834

Boosting is a framework in esemble learning. It was proposed as an answer to the question of whether weak learners can be combined together to make a strong PAC learner.

Boosting's underlying principle is to combine multiple weak learners (those that have accuracy slightly better than random guessing) with weight according to their errors to make a stronger model. Since we are 'averaging' the models with different weights, diversity (meaning that different base models make error on different training examples) helps improve performance of the combined model. The weight on the examples will also adjusted during the training, with misclassified data points received a higher weight to promote diversity.

AdaBoost is a Boosting algorithm that is used in practical machine learning problems. Its objective is to find a linear combination of base models. Its inputs are labeled training examples $S = \{(x_i, y_i)_{i=1}^m\}$, a hypothesis class H where we will take base hypotheses $h_i \in H$, and a distribution D_t, where the weight $D_t(i)$ for the training example are drawn. Given those inputs, the output of AdaBoost is:

$$f_T(x) = \sum_{t=1}^{T} \alpha_t h_t(x), \quad \alpha_t \geq 0$$

In round t, a new weak learner h_t that minimizes the empirical error of the sample will be added. The error ϵ_t and the weight α_t for the learner is calculated as follows:

$$\begin{aligned} \epsilon_t &= \min_{h \in H} \sum_{i=1}^{m} D_t(i) 1_{h(x_i) \neq y_i} \\ \alpha_t &= \frac{1}{2} \log \frac{1-\epsilon_t}{\epsilon_t} \end{aligned}$$

If the error is less than 0.5 (the weak learner is better than random guessing), then the weight for the model is positive. In the next round, weights on training examples are updated as follow:

$$D_{t+1}(i) = D_t(i) \cdot \frac{e^{-\alpha_t y_i h_t(x_i)}}{Z_t}$$

If previous models misclassify an example, its weight will increase (through the minus sign on the exponent), thus future base learners will focus more on the misclassified data.

AdaBoost has been shown to have an empirical error rate goes exponentially down in the number of classifiers used. Thus, if we have enough base classifiers, then the empirical error can be abitrarily low. It should also be noted that although AdaBoost is suggested to have high overfitting risk when the number of base classifier is high, it has not been observed in practical.

In conclusion, if we can access to base models that are in general better than random guesses, then we can combine them using Boosting methods to achieve a model with significantly lower error rate. And the more diversity the base learners are, the better the performance of the combined model.