Study Guide: Boosting

Student Name

October 9, 2024

1 Introduction

Boosting is a general technique for improving the performance of weak learners, which are learners that perform slightly better than random guessing. Boosting algorithms iteratively apply a weak learner to reweighted versions of the data and combine their predictions into a strong learner. The most prominent boosting algorithm is **AdaBoost**.

2 Weak Learnability (Section 10.1)

Weak Learnability refers to the ability of a learning algorithm to produce a hypothesis with an error rate slightly better than random guessing. Formally, a class H is weakly learnable if, for some $\gamma > 0$, there exists a learning algorithm such that for any distribution D over the domain, the learner outputs a hypothesis $h \in H$ with error:

$$L_D(h) \leq \frac{1}{2} - \gamma.$$

The central question in boosting is whether weak learners can be "boosted" to achieve arbitrarily low error.

3 AdaBoost Algorithm (Section 10.2)

AdaBoost (Adaptive Boosting) is a popular boosting algorithm that combines multiple weak learners to form a strong learner. The key idea is to iteratively reweight the training examples, giving more emphasis to those that were misclassified by previous weak learners.

3.1 AdaBoost Procedure

1. Initialize a uniform distribution over the training data.
2. For each round $t = 1, 2, \ldots, T$:
   - Train a weak learner h_t on the weighted data.
   - Compute the error rate ϵ_t of h_t on the training data.
   - Compute the weight α_t of the weak learner:

     $$\alpha_t = \frac{1}{2} \log\left(\frac{1 - \epsilon_t}{\epsilon_t}\right).$$

   - Update the weights of the examples, increasing the weights of the misclassified examples and decreasing the weights of the correctly classified ones:

     $$w_i^{(t+1)} = w_i^{(t)} \exp(\alpha_t \mathbb{I}[h_t(x_i) \neq y_i]).$$
3. The final hypothesis is the weighted combination of all the weak learners:

   $$H(x) = \text{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right).$$

3.2 Training Error of AdaBoost

The training error of AdaBoost decreases exponentially as the number of boosting rounds T increases. Specifically, the training error after T rounds is bounded by:

$$L_S(H) \leq \exp\left(-2\sum_{t=1}^{T} \gamma_t^2\right),$$

where $\gamma_t = \frac{1}{2} - \epsilon_t$.

4 VC Dimension of Boosted Classifiers (Section 10.3)

When applying boosting to a hypothesis class H, the VC dimension of the resulting boosted classifier is related to the VC dimension of H. Specifically, the VC dimension of the final hypothesis is at most:

$$VC(L(H,T)) \leq T \cdot VC(H),$$

where $L(H,T)$ is the class of boosted classifiers obtained by combining T weak learners from H.

5 AdaBoost for Face Recognition (Section 10.4)

AdaBoost has been successfully applied to real-world tasks, such as face detection. Viola and Jones developed a face detection algorithm that uses AdaBoost to select important features from a large set of possible features. Each weak learner is based on simple features of the image, such as the intensity difference between rectangular regions.

6 Summary (Section 10.5)

- Boosting is a powerful method for improving the accuracy of weak learners.
- **AdaBoost** is one of the most widely used boosting algorithms, combining weak hypotheses to form a strong learner.
- The training error of AdaBoost decreases exponentially, and the VC dimension of the resulting hypothesis class grows linearly with the number of boosting rounds.
- Boosting has practical applications in tasks like face recognition.