

Chapter 8 Ensemble Learning

Jinghao.Zhao

Abstract

This is the notes for ensemble learning, section 8.1 to 8.5, including formula and key points.

1 Individual and Ensemble

The structure of ensemble learning is that first generating a group of individual learner(base learner), then combining them based on some strategy. Individual learner should not bad than random picking($\geq 50\%$). There are two types of ensemble learning

1. 个体学习器之间有强依赖关系 \Rightarrow 串行 \Rightarrow Boosting or Random Forest.
2. 个体学习器之间不存在强依赖关系 \Rightarrow 并行 \Rightarrow Bagging.

2 Boosting

Boosting will train a base learner first, then change the distribution of sample which makes wrong classified samples draw more attention in next base learner. Boost will repeat this step until the number of base learner reach a predetermined quantity T.

Adaboost is one of the Boost algorithm based on additive model(加性模型). which is the linear addition

$$H(\mathbf{x}) = \sum_{t=1}^T \alpha_t h_t(\mathbf{x}) \quad (1)$$

Then minimize the loss by it. The algorithm is as following, support we have m samples for dataset D

1. Initial the sample weight distribution(\mathcal{D}_1), always be $\frac{1}{m}$;
2. Training a base learner h_t based on the weight distribution(\mathcal{D});
3. Calculate the error ϵ_t of h_t , if $\epsilon_t > 0.5$ this base learner will be ignore;
4. Calculate the weight α_t of classifier h_t ;

5. Update sample weight distribution \mathcal{D} , return step 2 and train next learner until the number of base learner is T.

Step 4 will be calculated by

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

Step 5 will be calculated by

$$\begin{aligned} \mathcal{D}_{t+1}(x) &= \frac{\mathcal{D}_t}{Z_t} \times \begin{cases} \exp(-\alpha_t), & \text{if } h_t(x) = f(x) \\ \exp(-\alpha_t), & \text{if } h_t(x) \neq f(x) \end{cases} \\ &= \frac{\mathcal{D}_t \cdot e^{(-\alpha_t f(x) h_t(x))}}{Z_t} \end{aligned} \quad (3)$$

The loss function of $h_t(x)$ will be (8.9) and α_t can be calculated by (8.10)