Supervised learning II

Liangjie Cao

29 April 2018

Today I learn what is objective function. The objective function1, averaged over all the training examples, can be seen as a kind of hilly landscape in the high-dimensional space of weight values. The negative gradient vector indicates the direction of steepest descent in this landscape, taking it closer to a minimum, where the output error is low on average. [2]It says most practitioners use a procedure called stochastic gradient descent (SGD).

The process is repeated for many small sets of examples from the training set until the average of the objective function stops decreasing. It is called stochastic because each small set of examples gives a noisy estimate of the average gradient over all examples. This simple procedure usually finds a good set of weights surprisingly quickly when compared with far more elaborate optimization techniques. [1] After training, the performance of the system is measured on a different set of examples called a test set. It also ays many of the current practical applications of machine learning use linear classifiers on top of hand-engineered features. A two-class linear classifier computes a weighted sum of the feature vector components. And output can be classified in the following table.1

$$R_{emp}(g) = \frac{1}{N} \sum_{i} L(y_i, g(x_i))$$

Figure 1: **Function**

The next step, I will continue learning how the function really works.

1	Quantitative output is called regression, or continuous variable prediction
2	Qualitative output is called classification, or discrete variable prediction.

Table 1: **Type of output variable**

References

[1] Bottou and L.Bousque. The tradeoffs of large scale learning. *Advances in Neural Information Processing Systems*, 20(161):168, 2007.

[2] Yoshua Bengio& Geoffrey Hinton Yann LeCun. Deep learning. *Nature*, 521(28):9, 2015.