Statistics Learning Theory:Logistic Regression

2019.02.23

Logistic Regression:Sigmoid Function

In this week slide, we will introduce the implementation of Logistic regression and related property.
To understand the method of logistic regression, we first introduce the sigmoid function. That is, the function $\sigma_{sig} : R \to [0, 1]$ over the class of linear functions L_d such that

$$\sigma_{sig}(z) = \frac{1}{1 + exp(-z)}$$

The hypothesis class becomes

$$\mathcal{H} = \sigma_{sig} \circ L_d = \{x \to \sigma_{sig}(\langle w, x \rangle)) : w \in R^d\}$$

Logistic Regression:Loss function

Given the classifier $h_w(x)$, we should define how bad it is to predict some $h_w(x) \in [0,1]$ given that the true label is $y \in \{1, -1\}$
Therefore, we would like that $h_w(x)$ would be large if $y = 1$ and that $1 - h_w(x)$ would be large if $y = -1$. Since

$$1 - h_w(x) = \frac{1}{1 + exp(\langle w, x \rangle)}$$

Therefore, any resonable loss function would increase monotonically with $\frac{1}{1+exp(y\langle w,x \rangle)}$

Logistic Regression:Loss function

We can choose the log function, that is the loss function

$$l(h_w, (x, y)) = \log(1 + exp(-y\langle w, x\rangle))$$

The ERM problem associated with logistic regression is

$$arg \min_{w \in R^d} \frac{1}{m} \sum_{i=1}^{m} \log(1 + exp(-y\langle w, x\rangle))$$

Logistic Regression: Remark

1. logistic loss function is convex function, the optimization can be solved efficienyly
2. The ERM problem associated with logistic regress is identical to the problem of finding a maximimum Likelihood Estimator.

Sigmoid function