Statistics Review

junyan.xu*

April 6, 2019

Contents

*junyanxu5513@gmail.com

1 Support Vector Machine and Regression

1.1 Separable

A distance of a point a to a plane $w \cdot x + b = 0$ is $\frac{a \cdot w + b}{\|w\|}$. Define margin ζ to be the value that all (x_i, y_i) are having distance larger that this.

$$\frac{x_i \cdot w + b}{\|w\|} \geq \zeta \tag{1}$$

Then if we fix ζ to be 1, mathematically the margin is $\frac{1}{\|w\|}$. The we are actually minimize $\frac{1}{2}\|w\|^2$

1.2 Non Separable

If the data is not separable, then we add penalty to the optimization function (constant muliply soft margin).

$$\begin{aligned} loss &= \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i \\ s.t. \forall i, y_i(w.x_i + b) &>= 1 - \xi_i \end{aligned} \tag{2}$$

The x_i represent the hinge loss, an "option" style loss function. Which could be rewrite as $max\{0, 1 - y(w \cdot x + b)\}$

1.3 Solver

Currently according to Stanford course this one should use SGD to solve.

1.4 Dual Problem

According to representer theorem the w can be a linear combination of x_i and is $w = \sum_{i=1}^{n} \alpha_i x_i$. Then we have $f(x_i) = y_i \sum_{j=1}^{n} \alpha_j x_j^T x_i + b$ The primal problem of this is minimize:

$$\frac{1}{2}\sum_{jk}\alpha_j\alpha_k y_j y_k (x_j^\top) + \sum_{i=1}^{n} C \cdot max(0, 1 - y_i \sum_{j=1}^{n}\alpha_j x_j^T x_i + b) \tag{3}$$

The dual problem is maximize

$$\begin{aligned} &-\frac{1}{2}\sum_{jk}\alpha_j\alpha_k y_j y_k (x_j^\top) + \sum_{i=1}^{n}\alpha_i \\ &0 \leq \alpha_i \leq C \\ &\sum_i \alpha_i y_i = 0 \end{aligned} \tag{4}$$

1.5 Regression

For regression part. The we assume the regression lies in range ϵ. Also the penalty of upper ϵ is ξ_u and down is ξ_d So the primal problem is

$$loss = \frac{1}{2}\left\|w\right\|^2 + C\sum_{i=1}(max\{f(x_i) - y_i - \epsilon, 0\} + max\{y_i - f(x_i) - \epsilon, 0\}) \tag{5}$$