Notes on Support Vector Machines

Tianjian Li

February 2022

1 Question Formulation

Support Vector Machines aim to solve this problem: Given training data

$$D = (x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$$

We want to find a hyperplane which seperates different types of training data.

Any plane can be described with such an equation:

$$w^\top x + b = 0$$

The distance between any sample x and this plane is:

$$r = \frac{|w^\top x + b|}{||w||}$$

Assume that the hyperplane correctly classifies every sample, that is to say

$$w^\top x_i + b \geq +1, \forall y_i = +1$$

and

$$w^\top x_i + b \leq -1, \forall y_i = -1$$

The samples that are the closest to the hyperplane are called **support vectors**. The distance between the positive and negative support vectors are called the **margin**. The margin is

$$\gamma = \frac{2}{||w||}$$

Our goal is to find the hyperplane that maximizes the margin. A mathematical denotation of our problem is

$$\max_{w,b} \frac{2}{||w||}$$

$$s.t. \; y_i(w^\top x_i + b) \geq 1, \forall i$$

Apparently, to maximize the margin, we need to maximize $||w||^{-1}$, this is equivalent to minimizing $||w||^2$, therefore the above equation can be rewritten as

$$\min_{w,b} \frac{1}{2}||w||^2$$

$$s.t. \; y_i(w^\top x_i + b) \geq 1, \forall i$$

2 Derivation

We use the **Lagrange Multiplier Method** to solve this problem. Assigning a multiplier α_i to each contraint, we have

$$L(w, b, \alpha) = \frac{1}{2}||w||^2 + \sum_{i=1}^{m} \alpha_i(1 - y_i(w^T x_i + b))$$

Let the partial derivative of L w.r.t. w and b equal to zero, we have

$$w = \sum_{i=1}^{m} \alpha_i x_i y_i$$

$$0 = \sum \alpha_i y_i$$

Therefore, the problem we need to solve can be reformulated as

$$\max_{\alpha} \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j x_i^\top x_j$$

$$s.t. \sum_{i=1}^{m} \alpha_i y_i = 0$$