CSE 446 Winter 2020 - Section 5 - SVM

February 13, 2020

1 Linearly separable case

Recall from the lecture that the SVM problem is given as:

$$\begin{aligned} \min_{w,t} \quad & \tfrac{1}{2}\textstyle\sum_{j=1}^d w_j^2 \\ \text{s.t.} \quad & y_i(w^T x_i - t) \geq 1 \quad \forall i = 1, \dots, n \end{aligned} \tag{1}$$

This lead to defining the Lagrangian of problem (1) as

$$L(w,t,\alpha) = \frac{1}{2}\sum_{j=1}^d w_j^2 - \sum_{i=1}^n \alpha_i (y_i(w^T x_i - t) - 1) \ .$$

Exercise: the optimization problem

$$\begin{aligned} \min_{w,t} \max_{\alpha} \quad & L(w,t,\alpha) \\ \text{s.t.} \quad & \alpha_i \geq 0 \quad \forall i = 1, \dots, n \end{aligned}$$

is the same as problem (1).

Now, the dual representation of problem (1) is the following:

$$\begin{aligned} \max_{\alpha} \min_{w,t} \quad & L(w,t,\alpha) \\ \text{s.t.} \quad & \alpha_i \geq 0 \quad \forall i = 1, \dots, n \end{aligned} \tag{2}$$

Let us solve 2:

$$\begin{aligned} L(w,t,\alpha) &= \frac{1}{2}\sum_{j=1}^d w_j^2 - \sum_{i=1}^n \alpha_i (y_i(w^T x_i - t) - 1) \\ &= \frac{1}{2}\sum_{j=1}^d w_j^2 - w^T \left(\sum_{i=1}^n \alpha_i y_i x_i\right) + t \left(\sum_{i=1}^n \alpha_i y_i\right) + \sum_{i=1}^n \alpha_i \end{aligned}$$

If $\sum_{i=1}^n \alpha_i y_i \neq 0$, then taking $t \to \pm\infty$ (the sign being the opposite of $\sum_{i=1}^n \alpha_i y_i$) leads to $L(w,t,\alpha) \to -\infty$. Therefore $\sum_{i=1}^n \alpha_i y_i = 0$ should hold, as a constraint for α. Thus (2) is equivalent to

$$\begin{aligned} \max_{\alpha} \min_{w,t} \quad & L(w,t,\alpha) \\ \text{s.t.} \quad & \alpha_i \geq 0 \quad \forall i = 1, \ldots, n \\ & \textstyle\sum_{i=1}^n \alpha_i y_i = 0 \end{aligned}$$

When $\sum_{i=1}^n \alpha_i y_i = 0$, we have

$$\begin{aligned} L(w,t,\alpha) &= \frac{1}{2} \sum_{j=1}^d w_j^2 - w^T \left(\sum_{i=1}^n \alpha_i y_i x_i \right) + \sum_{i=1}^n \alpha_i \\ &= \frac{1}{2} w^T w - w^T \left(\sum_{i=1}^n \alpha_i y_i x_i \right) + \sum_{i=1}^n \alpha_i \end{aligned}$$

so $L(w,t,\alpha)$ is a convex and differentiable function of w. We can therefore minimize over w by taking the derivative over w and setting it to 0. Namely,

$$\begin{aligned} \nabla_w L(w,t,\alpha) &= \nabla_w \left(\frac{1}{2} w^T w \right) - \nabla_w \left(w^T \left(\sum_{i=1}^n \alpha_i y_i x_i \right) \right) + \nabla_w \left(\sum_{i=1}^n \alpha_i \right) \\ &= w - \sum_{i=1}^n \alpha_i y_i x_i \ . \end{aligned}$$

So $\nabla_w L(\widehat{w}, t, \alpha) = 0$ if and only if $\widehat{w} = \sum_{i=1}^n \alpha_i y_i x_i$. We finally found the values of w and t that solve

$$\begin{aligned} \max_{\alpha} \min_{w,t} \quad & L(w,t,\alpha) \\ \text{s.t.} \quad & \alpha_i \geq 0 \quad \forall i = 1, \ldots, n \\ & \textstyle\sum_{i=1}^n \alpha_i y_i = 0 \end{aligned}$$

Let us plug these values into $L(w,t,\alpha)$: we obtain

$$\begin{aligned} L(\widehat{w},t,\alpha) &= \frac{1}{2} \left(\sum_{i=1}^n \alpha_i y_i x_i \right)^T \left(\sum_{i=1}^n \alpha_i y_i x_i \right) - \left(\sum_{i=1}^n \alpha_i y_i x_i \right)^T \left(\sum_{i=1}^n \alpha_i y_i x_i \right) + \sum_{i=1}^n \alpha_i \\ &= -\frac{1}{2} \left(\sum_{i=1}^n \alpha_i y_i x_i \right)^T \left(\sum_{i=1}^n \alpha_i y_i x_i \right) + \sum_{i=1}^n \alpha_i \\ &= -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j + \sum_{i=1}^n \alpha_i \end{aligned}$$

so the optimization problem 2 becomes

$$\begin{aligned} \max_{\alpha} \quad & \textstyle\sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j \\ \text{s.t.} \quad & \alpha_i \geq 0 \quad \forall i = 1, \ldots, n \\ & \textstyle\sum_{i=1}^n \alpha_i y_i = 0 \end{aligned}$$

This last problem is the dual representation of SVM.

2 Non linearly separable case

The SVM problem becomes then

$$\begin{array}{rl} \min\limits_{w,t,\xi} & \frac{1}{2}\sum_{j=1}^{d} w_j^2 + C\sum_{i=1}^{n}\xi_i \\ \text{s.t.} & y_i(w^T x_i - t) \geq 1 - \xi_i \quad \forall i = 1,\ldots,n \\ & \xi_i \geq 0 \quad \forall i = 1,\ldots,n \end{array} \tag{3}$$

The Lagrangian of problem (3) is

$$L_2(w,t,\xi,\alpha,\beta) = \frac{1}{2}\sum_{j=1}^{d} w_j^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i(y_i(w^T x_i - t) - (1-\xi_i)) - \sum_{i-1}^{n}\beta_i\xi_i$$

and the dual representation of problem (3) is the following:

$$\begin{array}{rl} \max\limits_{\alpha,\beta}\min\limits_{w,t,\xi} & L_2(w,t,\xi,\alpha,\beta) \\ \text{s.t.} & \alpha_i \geq 0 \quad \forall i = 1,\ldots,n \\ & \beta_i \geq 0 \quad \forall i = 1,\ldots,n \end{array} \tag{4}$$

We remark that

$$L_2(w,t,\xi,\alpha,\beta) = L(w,t,\alpha) + \sum_{i=1}^{n}(C - \alpha_i - \beta_i)\xi_i$$

A similar argument as the one leading to $\sum_{i=1}^{n}\alpha_i y_i = 0$, implies that $C - \alpha_i - \beta_i = 0$. Since $\beta_i \geq 0$, $C - \alpha_i - \beta_i = 0$ holds $\alpha_i \leq C$. Thus (4) becomes

$$\begin{array}{rl} \max\limits_{\alpha} & \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j x_i^T x_j \\ \text{s.t.} & 0 \leq \alpha_i \leq C \quad \forall i = 1,\ldots,n \\ & \sum_{i=1}^{n}\alpha_i y_i = 0 \end{array}$$