Machine Learning 1 HW 3

Maria Lentini

October 22, 2022

1 The l_2 Support Vector Machine [10pts]

We are given

$$\arg\min_{\boldsymbol{w},b,\xi} \frac{1}{2}||\boldsymbol{w}||^2 + \frac{C}{2}\sum_{i=1}^{n}\xi_i^2 \text{ s.t. } y_i(\boldsymbol{w}^T\boldsymbol{x_i} + b) \geq 1 - \xi_i \;\; \forall i \in [n]$$

The Lagrangian is therefore given by

$$L(\boldsymbol{w}, b, \xi, \alpha) = \frac{1}{2}||\boldsymbol{w}||^2 + \frac{C}{2}\sum_{i=1}^{n}\xi_i^2 - \sum_{i=1}^{n}\alpha_i\left(y_i(\boldsymbol{w}^T\boldsymbol{x_i} + b) - 1 + \xi_i\right)$$

Where $\alpha_i \geq 0$. Differntiating with respect $\boldsymbol{w}$, we have:

$$\frac{\partial L}{\partial \boldsymbol{w}} = \boldsymbol{w} - \sum_{i=1}^{n}\alpha_i y_i \boldsymbol{x_i} = \boldsymbol{0}$$
$$\Rightarrow \boldsymbol{w} = \sum_{i=1}^{n}\alpha_i y_i \boldsymbol{x_i}$$

Differentiating w.r.t. ξ_i we have

$$\frac{\partial L}{\partial \xi_i} = C\xi_i - \alpha_i = 0$$
$$\Rightarrow C\xi_i = \alpha_i$$

Differentiating w.r.r. b:

$$\frac{\partial L}{\partial b} = -\sum_{i=1}^{n}\alpha_i y_i = 0 \Rightarrow \sum_{i=1}^{n}\alpha_i y_i = 0$$

Plugging the derived expression for $\boldsymbol{w}$ into the original Lagrangian eliminates the explicit dependence on $\boldsymbol{w}$, and recalling that $\alpha_i y_i(\boldsymbol{w}^T\boldsymbol{x_i} + b) = 0$:

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

Thus, the dual form is therefore given by:

$$\arg\min_{\alpha}\sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j \boldsymbol{x_i}^T\boldsymbol{x_j} - \frac{1}{2C}\sum_{i=1}^{n}\alpha_i^2$$

2 Domain Adaptation Support Vector Machines [10pts]

We are given the primal form

$$\arg\min_{\boldsymbol{w}_T,\xi}\frac{1}{2}||\boldsymbol{w_T}||^2 + C\sum_{i=1}^{n}\xi_i - B\boldsymbol{w}_T^T\boldsymbol{w}_S \ \text{ s.t. }\ y_i(\boldsymbol{w}_T^T\boldsymbol{x}_i + b) \geq 1-\xi_i, \ \ \xi_i \geq 0 \ \ \forall i \in [n]$$

The Lagrangian is therefore:

$$L(\boldsymbol{w}_T,\xi,\mu,\alpha) = \frac{1}{2}||\boldsymbol{w_T}||^2 + C\sum_{i=1}^{n}\xi_i - B\boldsymbol{w}_T^T\boldsymbol{w}_S - \sum_{i=1}^{n}\mu_i\xi_i - \sum_{i=1}^{n}\alpha_i\left[y_i(\boldsymbol{w}_T^T\boldsymbol{x}_i + b) - 1 + \xi_i\right]$$

Deriving the dual form, we differentiate L w.r.t. $\boldsymbol{w}_T$:

$$\begin{aligned}
\frac{\partial L}{\partial \boldsymbol{w}_T} &= \boldsymbol{w}_T - B\boldsymbol{w}_S - \sum_{i=1}^{n}\alpha_i y_i \boldsymbol{x}_i = 0 \\
&\Rightarrow \boldsymbol{w}_T = B\boldsymbol{w}_S + \sum_{i=1}^{n}\alpha_i y_i \boldsymbol{x}_i
\end{aligned}$$

Differentiating w.r.t. ξ_i we have

$$\begin{aligned}
\frac{\partial L}{\partial \xi_i} &= C - \mu_i - \alpha_i = 0 \\
&\Rightarrow C = \mu_i + \alpha_i
\end{aligned}$$

Differentiating w.r.t. b

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{n} \alpha_i y_i = 0$$

Let's take a separate look at each of the components of the Lagrangian and simplify using the results from the derivatives. First note that

$$\begin{aligned}\frac{1}{2}||\boldsymbol{w_T}||^2 &= \left(B\boldsymbol{w}_S + \sum_{i=1}^{n} \alpha_i y_i \boldsymbol{x}_i\right)^T \left(B\boldsymbol{w}_S + \sum_{i=1}^{n} \alpha_i y_i \boldsymbol{x}_i\right) \\ &= B^2||\boldsymbol{w}_S||^2 + 2B\boldsymbol{w}_S^T \sum_{i=1}^{n} \alpha_i y_i \boldsymbol{x}_i + \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i \boldsymbol{x}_j\end{aligned}$$

Next, let's look at

$$\begin{aligned}-\sum_{i=1}^{n} \alpha_i \left[y_i(\boldsymbol{w}_T^T \boldsymbol{x}_i + b) - 1 + \xi_i\right] &= -\sum_{i=1}^{n} \alpha_i y_i \boldsymbol{w}_T^T \boldsymbol{x}_i - \sum_{i=1}^{n} \alpha_i - \sum_{i=1}^{n} \alpha_i \xi_i \\ &= -\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i \boldsymbol{x}_j - \sum_{i=1}^{n} \alpha_i - \sum_{i=1}^{n} \alpha_i \xi_i\end{aligned}$$

And finally,

$$\begin{aligned}-B\boldsymbol{w}_T^T \boldsymbol{w}_S &= -B(B\boldsymbol{w}_S + \sum_{i=1}^{n} \alpha_i y_i \boldsymbol{x}_i)^T \boldsymbol{w}_S \\ &= -B^2||\boldsymbol{w}_S||^2 - B\boldsymbol{w}_S^T \sum_{i=1}^{n} \alpha_i y_i \boldsymbol{x}_i\end{aligned}$$

Where I have colored like terms the same. The Lagrangian is therefore

$$\begin{aligned}L(\alpha) &= B\boldsymbol{w}_S^T \sum_{i=1}^{n} \alpha_i y_i \boldsymbol{x}_i + C\sum_{i=1}^{n} \xi_i - \sum_{i=1}^{n} \mu_i \xi_i - \sum_{i=1}^{n} \alpha_i - \sum_{i=1}^{n} \alpha_i \xi_i \\ &= B\boldsymbol{w}_S^T \sum_{i=1}^{n} \alpha_i y_i \boldsymbol{x}_i + C\sum_{i=1}^{n} \xi_i - \sum_{i=1}^{n} \alpha_i - \sum_{i=1}^{n} (\alpha_i + \mu_i)\xi_i \\ &= B\boldsymbol{w}_S^T \sum_{i=1}^{n} \alpha_i y_i \boldsymbol{x}_i + C\sum_{i=1}^{n} \xi_i - \sum_{i=1}^{n} \alpha_i - C\sum_{i=1}^{n} \xi_i \\ &= B\sum_{i=1}^{n} \alpha_i y_i \boldsymbol{w}_S^T \boldsymbol{x}_i - \sum_{i=1}^{n} \alpha_i\end{aligned}$$

The dual problem is therefore,

$$\arg\min_{\alpha} B\sum_{i=1}^{n} \alpha_i y_i \boldsymbol{w}_S^T \boldsymbol{x}_i - \sum_{i=1}^{n} \alpha_i \text{ s.t. } 0 \le \alpha \le C, \ \sum_{i=1}^{n} \alpha_i y_i = 0$$