Dual problem in SVM

Ali Gooya

Primal problem

- Recall that the in linear separable SVM, the constrained optimization is specified as:

$$\min_{\mathbf{w},b} \quad f(\mathbf{w}) = \tfrac{1}{2}\mathbf{w}^\top \mathbf{w} \tag{1}$$

$$\text{subject to} \quad t_i\left(\mathbf{w}^\top \mathbf{x}_i + b\right) \geq 1, \quad i = 1, \ldots, n \tag{2}$$

- Applying the KKT conditions, the primal problem is:

$$L(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\mathbf{w}^\top \mathbf{w} + \sum_{i}^{n} \alpha_i (1 - t_i(\mathbf{w}^\top \mathbf{x}_i + b))$$

$$\alpha_i \geq 0 \qquad \boldsymbol{\alpha} = (\alpha_1 \cdots, \alpha_n)^\top$$

$$1 - t_i(\mathbf{w}^\top \mathbf{x}_i + b) \leq 0$$

$$\alpha_i (1 - t_i(\mathbf{w}^\top \mathbf{x}_i + b)) = 0$$

Slackness conditions

From primal to dual problems

- The primal problem is a convex optimization problem that can be solved by solvers such as ALGLIB.
- But our SVM will remain linear ☹
- Moving towards dual problem can be rewarding and can make SVM non-linear ☺ (-- using kernel trick!)
- To define the dual problem, we first eliminate the primal variables.

$$
\begin{aligned}
\frac{\partial L}{\partial b} &= 0 \Rightarrow \sum_{i=1}^{n} \alpha_i t_i = 0 \\
\frac{\partial L}{\partial \mathbf{w}} &= 0 \Rightarrow \mathbf{w} = \sum_{i=1}^{n} \alpha_i t_i \mathbf{x}_i
\end{aligned}
$$

Dual problem

- Substitute the results in the primal:

$$
\begin{aligned}
L(\mathbf{w}, b; \alpha) &= \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j t_i t_j \mathbf{x}_i^\top \mathbf{x}_j \; + \; \sum_{i=1}^{n} \alpha_i \left(1 - t_i \left(\sum_{j=1}^{n} \alpha_j t_j \mathbf{x}_j^\top \mathbf{x}_i + b \right) \right) \\
&= \sum_{i=1}^{n} \alpha_i \; - \; \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j t_i t_j \mathbf{x}_i^\top \mathbf{x}_j
\end{aligned}
$$

Dual problem

- We arrive at the following dual problem

$$g(\boldsymbol{\alpha}) = \sum_i \alpha_i - \frac{1}{2}\sum_{i,j} t_i t_j \alpha_i \alpha_j \mathbf{x}_i^\top \mathbf{x}_j$$

$$\alpha_i \geq 0$$

$$\sum_{i=1}^{n} \alpha_i t_i = 0$$

Complementary slackness condition

- Note that the dual problem is concave (why?) and should be maximized (see the the previous lecture)
- This can be solved by quadratic programming.

Solution for the dual

- Let $\boldsymbol{\alpha}^* = (\alpha_1^*, \cdots, \alpha_n^*)^\top$ be the solution for above then

$$\mathbf{w}^* = \sum_{i=1}^{n} \alpha_i^* t_i \mathbf{x}_i$$

- To determine b^*, we note that for any $\alpha_i^* > 0$

$$t_i({\mathbf{w}^*}^\top \mathbf{x}_i + b^*) = 1$$

Sparsity in predication

- For any test point such as $\mathbf{x}$ the discriminant will be

$$
\begin{aligned}
y(\mathbf{x}) &= \mathbf{w}^{*\top}\mathbf{x} + b^* \\
&= \sum_{i:\alpha_i^*>0} \alpha_i^* t_i \mathbf{x}^\top \mathbf{x}_i + b^*
\end{aligned}
$$

- This shows that prediction uses only a sparse number of $\mathbf{x}_i$'s.
- One interesting points is that only inner products ($\mathbf{x}_i^\top \mathbf{x}_j$) appear in the dual problem and the prediction!
- We will exploit this observation to make SVMs non-linear.