Machine Learning 1 Homework 12 Theory Part

25.01.2016

Ex1. The Dual SVM

a)

$$\Lambda(\mathbf{w}, \theta, \alpha) = \frac{1}{2}||\mathbf{w}||^2 - \sum_{i=1}^{N} \alpha_i [y_i(\mathbf{w}^\mathbf{T}\mathbf{x_i} + \theta) - 1]$$

b)Notice the second part of Λ contains the constraint $-\big(y_i(\mathbf{w}^\mathbf{T}\mathbf{x_i} + \theta) - 1\big) \leq 0$. Suppose we get the optimal value $\mathbf{w}^*$ and θ^*. Also notice that if we set Lagrangian multipliers positive, we can get a dual problem from primal problem. The optimal value is a saddle point.

$$\min_{\mathbf{w},\theta} \Lambda(\mathbf{w}, \theta, \alpha^*) = \Lambda(\mathbf{w}^*, \theta^*, \alpha^*) = \max_{\alpha \geq 0} \Lambda(\mathbf{w}^*, \theta^*, \alpha)$$

Take derivatives of Λ and set them to 0:

$$\frac{\partial}{\partial \theta}\Lambda(\mathbf{w}, \theta, \alpha) = -\sum_{i=1}^{N} \alpha_i y_i = 0$$

$$\frac{\partial}{\partial \mathbf{w}}\Lambda(\mathbf{w}, \theta, \alpha) = \mathbf{w} - \sum_{i=1}^{N} \alpha_i y_i \mathbf{x_i} = 0$$

$$\Rightarrow \qquad \mathbf{w} = \sum_{i=1}^{N} \alpha_i y_i \mathbf{x_i}$$

$$\sum_{i=1}^{N} \alpha_i y_i = 0$$

Insert the results into orginal formula, we get:

$$\begin{aligned}
\Lambda &= \frac{1}{2}||\mathbf{w}||^2 - \sum_{i=1}^{N} \alpha_i [y_i(\mathbf{w}^\mathbf{T}\mathbf{x_i} + \theta) - 1] \\
&= \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \mathbf{x_i}^\mathbf{T}\mathbf{x_j} - \sum_{i=1}^{N} \alpha_i y_i \left(\sum_{j=1}^{N} \alpha_j y_j \mathbf{x_j}^\mathbf{T}\right) \mathbf{x_i} + \theta \sum_{i=1}^{N} \alpha_i y_i + \sum_{i=1}^{N} \alpha_i \\
&= -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \mathbf{x_i}^\mathbf{T}\mathbf{x_j} + \sum_{i=1}^{N} \alpha_i
\end{aligned}$$

Now we convert the primal problem to this dual problem:

$$\max_{\alpha \geq 0} \Lambda = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \mathbf{x_i}^\mathbf{T}\mathbf{x_j} + \sum_{i=1}^{N} \alpha_i$$

$$\sum_{i=1}^{N} \alpha_i y_i = 0, \quad \alpha_i \geq 0$$

c)Apply Kernel trick on the primal and dual program, we get:

1)Primal Problem:

$$\min_{\mathbf{w},\theta} \Lambda = \frac{1}{2}||\mathbf{w}||^2 - \sum_{i=1}^{N} \alpha_i [y_i(\mathbf{w}^\mathbf{T}\boldsymbol{\Phi}(\mathbf{x_i}) + \theta) - 1]$$

2)Dual Problem:

$$\max_{\alpha\geq 0} \Lambda = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j \mathrm{k}(\mathbf{x_i},\mathbf{x_j}) + \sum_{i=1}^{N}\alpha_i$$

Ex2. SVMs and Quadratic Programming

a)As the dual Kernel problem is a maximization, we should add a minus in front of it and convert it to a minimization problem, such that this can correspond to the given quadratic optimization problem.

1)P corresponds to the kernel matrix multiply with vector **y**

$$P_{ij} = y_i y_j K(\mathbf{x_i},\mathbf{x_j})$$

2)set $\mathbf{q^T x} = -\sum_{i=1}^{N}\alpha_i$, q corresponds to a vector with length N and all the elements in q is -1.

$$q = (-1,\cdots,-1)^T$$

3)As the inequality constraint in Ex1. is $\alpha_i \geq 0$

$$G = -\mathbf{I}, \quad \mathbf{h} = \mathbf{0}$$

4)The equality constraint in Ex1. is $\sum_{i=1}^{N}\alpha_i y_i = 0$

$$A = diag\{y1,\cdots,y_i,\cdots,y_N\}$$
$$\mathbf{b} = \mathbf{0}$$