Lecture-9 Supplementary

09-Feb-10

Recovery of ω, b, in particular $f(x)$ from dual solution:

Primal:

$$\min_{\omega, b, \xi_i} \frac{1}{2}\|\omega\|_{\mathcal{H}}^2 + C\sum_i \xi_i$$

$$\text{s.t.} \quad y_i(\langle \omega, \phi_{x_i}\rangle - b) \geq 1 - \xi_i, \quad \xi_i \geq 0$$

$$\omega = \sum_i \alpha_i y_i k(x_i, \cdot) \quad \text{I}$$

Intermediate:

$$\min_{\alpha, b, \xi} \frac{1}{2}\alpha^T Q \alpha + C\, 1^T \xi$$

$$\text{s.t.} \quad Q\alpha - by \geq 1 - \xi, \quad \xi \geq 0$$

II $\quad Q\alpha = Q\lambda, \; C1 = \lambda + \beta, \; \lambda^T y = 0$

$\lambda_i(1 - \xi - Q\alpha + by)_i = 0, \; \beta_i \xi_i = 0$ KKT conds.

Dual:

$$\max_{\lambda} 1^T\lambda - \frac{1}{2}\lambda^T Q\lambda$$

$$\text{s.t.} \quad 0 \leq \lambda \leq C1, \; y^T\lambda = 0.$$

Now by solving the dual one gets λ. In case Q is pd (e.g. always true with a gaussian kernel), we have $\alpha = \lambda$. Hence by I, ω can be got easily.

To get b:

Take any $\lambda_i \in (0, C)$ then $\beta_i \neq 0 \Rightarrow \xi_i = 0$

then $1 - \xi_i - (Q\alpha)_i + b y_i = 0$

$$\Rightarrow b = \frac{(Q\alpha)_i - 1}{y_i}$$

But what if $\nexists$ $\lambda_i \in (0,C)$??

In such a case look at $\overset{\text{all}}{\lambda_i} = C$ for them $\xi_i = 1-(Q\alpha)_i + by_i \geq 0$

Let S_+ be set of all $\lambda_i = C$ & $y_i = 1$ for them $\rightarrow$ $b \geq (Q\alpha)_i - 1$

Let S_- be set of all $\lambda_i = C$ & $y_i = -1$ for them $\rightarrow$ $b \leq 1-(Q\alpha)_i$

$\therefore$ we can select <u>any</u> b satisfying:

$$\max_{i \in S_+} (Q\alpha)_i - 1 \leq b \leq \min_{i \in S_-} 1-(Q\alpha)_i$$

Now in case Q is singular, α need not be λ. However since KKT are necessary & sufficient in this case, it is OK if we take any $\alpha, \lambda, b, \xi_i, \beta$ satisfying (II). In particular we can always choose $\alpha = \lambda$!

Now, we can do better: in fact suppose we choose α such that:

$$\min_{\alpha} \|\alpha\|_0$$
$$\text{s.t.} \quad Q\alpha = Q\lambda$$

λ at optimality

then we can compute $f(x)$ using as less no. of training datapoints as possible. Hence from computational view-point, this is better. One can also solve $\min_{\alpha} \|\alpha\|_1$ s.t. $Q\alpha = Q\lambda$ which is convex relaxation of the above!