Lecture 8 Supplementary

6-Feb-09

Derivation of Dual

of soft-margin SVM using arbitrary kernel k.

Let gram-matrix of training points with k be $K_{m \times m}$. Let Y be the diagonal matrix with entries as labels of training pts. and y be the vector with entries as labels of training pts. Let $Q = YKY$. Let 1 be the vector of ones of appropriate size. Let $\mathcal{H}$ be an RKHS for which kernel is k. Let $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ & $\|\cdot\|_{\mathcal{H}}$ be the inner product & induced norm in $\mathcal{H}$.

SVM:

$$\min_{w, b, \xi} \ \frac{1}{2}\|w\|_{\mathcal{H}}^2 + C\sum_{i=1}^{m} \xi_i$$

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

using representer theorem we have this equivalent formulation:

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

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

$$\mathcal{L} = \frac{1}{2}\alpha^T Q\alpha + C\, 1^T \xi + \lambda^T(1 - \xi + by - Q\alpha) - \beta^T \xi$$

(here λ, β are Lagrange multipliers: both are vectors)

$$\nabla_\alpha \mathcal{L} = 0 \Rightarrow Q\alpha = Q\lambda$$

(Note: Q is symmetric, may not be invertible)

$$\nabla_\xi \mathcal{L} = 0 \Rightarrow C1 = \lambda + \beta$$

$$\frac{\partial \mathcal{L}}{\partial b} = 0 \Rightarrow \lambda^T y = 0$$

eliminating α we get:

$$\mathcal{L} = -\frac{1}{2}\lambda^T Q \lambda + 1^T \lambda \qquad \left(\because \frac{1}{2}\alpha^T Q \alpha = \frac{1}{2}\alpha^T Q \lambda = \frac{1}{2}\lambda^T Q \alpha = \frac{1}{2}\lambda^T Q \lambda\right)$$

Hence dual is:

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

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

vector of zeros

This is usually expressed in the following form where λ is replaced by alpha:

$$\min_{\alpha_i} \quad \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j k(x_i, x_j) - \sum_{i=1}^{m} \alpha_i$$

$$\underline{\underline{\text{s.t.}}} \quad 0 \leq \alpha_i \leq C, \quad \sum_{i=1}^{m} y_i \alpha_i = 0$$