Optimization task for SVM (non-separable case)

Victor Kitov

Consider the optimization task:

$$\begin{cases} \frac{1}{2}\left\|w\right\|^2 + C\sum_{n=1}^{N}\xi_n \to \min_{w,w_0,\xi} \\ y_n(w_0 + w^T x_n) \ge 1 - \xi_n \\ \xi_n \ge 0 & n = \overline{1,N} \end{cases} \tag{1}$$

Lagrangian becomes

$$L = \frac{1}{2}\left\|w\right\|^2 + C\sum_{n=1}^{N}\xi_n - \sum_{n=1}^{N}\alpha_n[y_n(w_0 + w^T x_n) - 1 + \xi_n] - \sum_{n=1}^{N} r_n\xi_n$$

with dual variables $\alpha_n \ge 0$, $r_n \ge 0$ $n = \overline{1,N}$. Objective function is convex, constraint functions are linear (so both convex and concave) => solution is a saddle point of L. Lagrangian is minimized w.r.t. primal variables w, w_0, ξ_n and minimized w.r.t. dual vaiables α_n, r_n.

Differentiate w.r.t. primal variables to obtain

$$\frac{\partial L}{\partial w} = w - \sum_{n=1}^{N}\alpha_n y_n x_n = 0 \tag{2}$$

$$\frac{\partial L}{\partial w_0} = -\sum_{n=1}^{N}\alpha_n y_n = 0 \tag{3}$$

$$\frac{\partial L}{\partial \xi_n} = C - \alpha_n - r_n = 0 \tag{4}$$

Substituting (2),(3),(4) into Lagrangian, we obtain:

$$\begin{aligned} L &= \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y_i y_j\langle x_i, x_j\rangle + C\sum_{n=1}^{N}\xi_n - \left(\sum_{n=1}^{N}\alpha_n y_n\right)w_0 - \sum_{n=1}^{N}\alpha_n\sum_{j=1}^{N}\alpha_j y_j\langle x_j, x_n\rangle + \sum_{n=1}^{N}\alpha_n(1-\xi_n) - \sum_{n=1}^{N} r_n\xi_n \\ &= -\frac{1}{2}\sum\sum_{i,j}\alpha_i\alpha_j y_i y_j\langle x_i, x_j\rangle - \left(\sum_{n=1}^{N}\alpha_n y_n\right)w_0 + \sum_{n=1}^{N}(C - r_n - \alpha_n)\,\xi_n + \sum_{n=1}^{N}\alpha_n = \\ &= -\frac{1}{2}\sum\sum_{i,j}\alpha_i\alpha_j y_i y_j\langle x_i, x_j\rangle + \sum_{n=1}^{N}\alpha_n \end{aligned}$$

where we used equations (3) and (4).

Since primal task (in primal variables) was minimization, dual task (in dual variables) will be maximization:

$$\begin{cases} -\frac{1}{2}\sum\sum_{i,j}\alpha_i\alpha_j y_i y_j \langle x_i, x_j\rangle + \sum_{n=1}^N \alpha_n \to \max_\alpha \\ \sum_{n=1}^N \alpha_n y_n = 0 \\ 0 \leq \alpha_n \leq C \end{cases}$$

where first constraint comes from (3) and second - from non-negativity of dual variable α_n, (4) and non-negativity of dual variable r_n. Kernel version of this equation is

$$\begin{cases} -\frac{1}{2}\sum\sum_{i,j}\alpha_i\alpha_j y_i y_j K(x_i, x_j) + \sum_{n=1}^N \alpha_n \to \max_\alpha \\ \sum_{n=1}^N \alpha_n y_n = 0 \\ 0 \leq \alpha_n \leq C \end{cases}$$