$$\frac{\partial L(x,\mu)}{\partial x}\Big|_{x=x^*} = 0$$

$$\begin{cases} \min\limits_{x} \max\limits_{\mu} L(x,\mu) = \max\limits_{\mu} \min\limits_{x} L(x,\mu) < \min\limits_{x} f(x) = f(x^*) \\ \mu_k g_k(x) = 0 \\ \frac{\partial L(x,\mu)}{\partial x}\Big|_{x=x^*} = 0. \end{cases}$$

回到SVM：

其KKT $\begin{cases} a_i \geq 0 \\ y_i f(x_i) - 1 \geq 0 \\ a_i(y_i f(x_i) - 1) = 0 \end{cases}$

当 $a_i = 0$，该样本将不会在式 $f(x) = \sum_{i=1}^{m} \alpha_i y_i x_i^T x + b$ 的求和中出现，也不会对 $f(x)$ 有任何影响。

当 $a_i > 0$，则必有 $y_i f(x_i) = 1$

SMO (Sequential Minimal Optimization)

① 选取一对需要更新的变量 α_i 和 α_j

② 固定 α_i 和 α_j 以外的参数，求解

$$\max_{\alpha} \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

获得更新后的 α_i 和 α_j

SMO算法之所以高效，是由于在固定其他参数后，仅优化两个参数的过程能做到非常高效

$$\alpha_i y_i + \alpha_j y_j = c, \quad \alpha_i \geq 0, \quad \alpha_j \geq 0$$

$$c = -\sum_{k \neq i,j} \alpha_k y_k$$

使 $\sum_{i=1}^{m} \alpha_i y_i = 0$. 用 $\alpha_i y_i + \alpha_j y_j = c$ 消去 α_j

可求出更新后 α_i 和 α_j.

对任意支持向量 (x_s, y_s)，都有 $y_s f(x_s) = 1$

$$y_s\left(\sum_{i \in S} \alpha_i y_i x_i^T x_s + b\right) = 1$$

用所有支持向量求解的平均值：

$$b = \frac{1}{|S|} \sum_{s \in S} \left(\frac{1}{y_s} - \sum_{i \in S} \alpha_i y_i x_i^T x_s\right).$$