2018211670

郝健乔

清華大學

Machine Learning Problem Set 4

1. Solution: The primal form of linear SVM for cases when the samples are not linearly separable is

$$\min_{w,b,\xi} \ \frac{1}{2}(w\cdot w)+C\sum_{i=1}^{N}\xi_i$$

$$\text{s.t.} \quad y_i(w\cdot x_i+b)\geq 1-\xi_i,\ i=1,2,\cdots,N.$$

$$\xi_i\geq 0\ ,\quad i=1,2,\cdots,N.$$

Then we have the Lagarian

$$L(w,b,\xi,\alpha,\mu)=\frac{1}{2}(w\cdot w)+C\cdot\sum_{i=1}^{N}\xi_i-\sum_{i=1}^{N}\alpha_i\big(y_i(w\cdot x_i+b)-1+\xi_i\big)-\sum_{i=1}^{N}\mu_i\xi_i. \quad ①$$

At the saddlepoints we have

$$\frac{\partial L(w_0,b_0,\xi_i^0,\alpha^0,\mu^0)}{\partial w_0}=w_0-\sum_{i=1}^{N}\alpha_i^0 y_i x_i=0 \Longrightarrow w_0=\sum_{i=1}^{N}\alpha_i^0 y_i x_i. \quad ②$$

$$\frac{\partial L(w_0,b_0,\xi_i^0,\alpha^0,\mu^0)}{\partial b}=\sum_{i=1}^{N}\alpha_i^0 y_i=0. \quad ③$$

$$\frac{\partial L(w_0,b_0,\xi_0,\alpha^0,\mu^0)}{\partial \xi_i}=C-\alpha_i-\mu_i=0 \quad ④$$

Substitute ②.③ ④. into ① we have

$$\begin{aligned}
① = L(w,b,\xi,\alpha,\mu) &= \frac{1}{2}(w\cdot w)-\sum_{i=1}^{N}\alpha_i\big(y_i(w\cdot x_i+b)-1+\xi_i\big)-\sum_{i=1}^{N}\mu_i\xi_i+C\cdot\sum_{i=1}^{N}\xi_i\\
&= \frac{1}{2}(w\cdot w)-\sum_{i=1}^{N}\alpha_i\big(y_i(w\cdot x_i+b)-1\big)+\sum_{i=1}^{N}\xi_i(C-\alpha_i-\mu_i)\\
&= \frac{1}{2}\Big(\sum_{i=1}^{N}\alpha_i^0 y_i x_i\Big)\cdot\Big(\sum_{i=1}^{N}\alpha_i^0 y_i x_i\Big)-\sum_{i=1}^{N}\alpha_i y_i\Big(\sum_{j=1}^{N}\alpha_j^0 y_j x_j\cdot x_i\Big)+\sum_{i=1}^{N}\alpha_i\\
&\qquad +\underbrace{b\sum_{i=1}^{N}\alpha_i y_i}_{=0}+\sum_{i=1}^{N}\xi_i\underbrace{(C-\alpha_i-\mu_i)}_{=0}\\
&= \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j(x_i\cdot x_j)-\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j(x_i\cdot x_j)+\sum_{i=1}^{N}\alpha_i\\
&= -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j(x_i\cdot x_j)+\sum_{i=1}^{N}\alpha_i.
\end{aligned}$$

Since the dual problem is $\max\limits_{\alpha,\mu,\ \alpha_i\geq 0,\ \mu_i\geq 0}\ \min\limits_{w,b,\xi} L(w,b,\xi,\alpha,\mu)$. then we have the dual problem

$$\max_{\alpha,\mu}\ -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j(x_i\cdot x_j)+\sum_{i=1}^{N}\alpha_i.$$

$$\text{s.t.}\ \begin{cases}\sum_{i=1}^{N}\alpha_i y_i=0.\\ C-\alpha_i-\mu_i=0.\ i=1,2,\cdots,N & ⑤\\ \alpha_i\geq 0\quad i=1,2,\cdots,N. & ⑥\\ \mu_i\geq 0\quad i=1,2,\cdots,N. & ⑦\end{cases}$$

From ⑤ ⑥. ⑦. We have write the constraints as one constraint $\longrightarrow 0\leq\alpha_i\leq C$

$\xrightarrow{\text{Dual problem}}$

$$\max_{\alpha}\ -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j(x_i\cdot x_j)+\sum_{i=1}^{N}\alpha_i$$

$$\text{s.t.}\ \begin{cases}\sum_{i=1}^{N}\alpha_i y_i=0.\\ 0\leq\alpha_i\leq C.\ i=1,2,\cdots,N.\end{cases}$$

Since the primal problem is a quadratic convex optimazation problem. ~~et~~ Its solution satisfies the KKT conditions:

$$\begin{cases} \nabla_w L(w^*, b^*, \xi^*, \alpha^*, \mu^*) = w^* - \sum_{i=1}^{N} \alpha_i^* y_i x_i = 0 \\ \nabla_b L(w^*, b^*, \xi^*, \alpha^*, \mu^*) = -\sum_{i=1}^{N} \alpha_i^* y_i = 0. \\ \nabla_\xi L(w^*, b^*, \xi^*, \alpha^*, \mu^*) = C - \alpha^* - \mu^* = 0. \quad (\text{KKT.①}) \\ \alpha_i^* (y_i (w^* \cdot x + b^*) - 1 + \xi_i^*) = 0. \quad (\text{KKT.②}) \\ \mu_i^* \xi_i^* = 0. \quad (\text{KKT.③}) \end{cases}$$

$\therefore$ At the solution ~~α~~ $\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_N^*)$. we have

$$|w^*|^2 = w^* \cdot w^* = \left(\sum_{i=1}^{N} \alpha_i^* y_i x_i\right)^T \cdot \left(\sum_{i=1}^{N} \alpha_i^* y_i x_i\right)$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i^* \alpha_j^* \cdot y_i y_j \cdot (x_i \cdot x_j).$$

where $(x_i \cdot x_j)$ means the inner product of x_i and x_j.

From (KKT.③). We have $\mu_i^* \xi_i^* = 0$, which means that for the wrongly classified samples. we have $\xi_i^* > 0$. then we must have their $\mu_i^* = 0$.

Then from (KKT.①). ~~we~~ We can know that for wrongly classified samples. $\alpha_i^* = C - \mu_i^* = C - 0 = C$.

which means that for wrongly classified samples. ~~we~~ they are includeed in SVs and their $\alpha_i = C$.