年 月 日 第 次作业

Support Vector Machine (SVM)

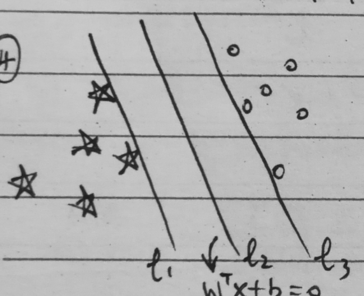
May 2, 2020
Guorui Shen
guorui233@outlook.com

① $\{(x_i, y_i)\}_{i=1}^{N}$, $x_i \in R^n$, $y_i = 1$ or -1

② An affine hyperplane is defined as $w^T x + b = 0$, or $w_1 x_1 + \cdots + w_n x_n + b = 0$, here $\boxed{x_i \in R}$ is different from $x_i \in R^n$.
$x = [x_1, \cdots, x_n]^T \in R^n$ is a n dimensional variable vector.

③ The task of SVM is to find an affine hyperplane $w^T x + b = 0$ such that $w^T x_i + b > 0$ if $y_i = 1$, or $w^T x_i + b < 0$ if $y_i = -1$.

④



$l_1: w^T x + b = k$ $\Leftrightarrow$ $l_1: w_{new}^T x + b_{new} = 1$

$l_2: w^T x + b = 0$ $\Leftrightarrow$ $l_2: w_{new}^T x + b_{new} = 0$

$l_3: w^T x + b = -k$ $\Leftrightarrow$ $l_3: w_{new}^T x + b_{new} = -1$

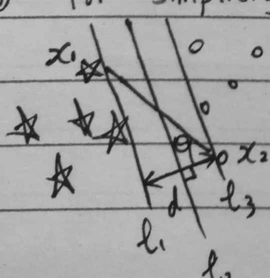
$\boxed{\text{Therefore}}$: Finding $w^T x + b = 0$ is equivalent to finding $w_{new}^T x + b_{new} = 0$

Once we know w_{new} & b_{new}, we can compute $w_{new}^T x_i + b_{new}$ to decide $y_i = 1$ or $y_i = -1$.

⑤ For simplicity, let $l_1: w^T x + b = 1$

$l_2: w^T x + b = 0$

$l_3: w^T x + b = -1$



The goal is to maximize the distance between l_1 & l_2 while satisfying $w^T x_i + b \geq 1$ if $y_i = 1$ & $w^T x_i + b \leq -1$ if $y_i = -1$

⑥ In the figure of step ⑤, $\begin{cases} x_1 \text{ is in } l_1 \text{, then } w^T x_1 + b = 1 \\ x_2 \text{ is in } l_3 \text{, then } w^T x_2 + b = -1 \end{cases}$

hence $w^T(x_1 - x_2) = 2$. Since w & $x_1 - x_2$ represents two n-dimensional vectors, we have $w^T(x_1 - x_2) = \|w\| \cdot \|x_1 - x_2\| \cdot \cos\theta$

therefore $d = \|x_1 - x_2\| \cdot \cos\theta = \frac{w^T(x_1 - x_2)}{\|w\|} = \frac{2}{\|w\|}$

⑦ Finally, we have: $\boxed{\begin{aligned} &\max \frac{2}{\|w\|} \\ &\text{s.t. } w^T x_i + b \geq 1 \quad \text{if } y_i = +1 \\ &\qquad w^T x_i + b \leq -1 \quad \text{if } y_i = -1 \end{aligned}}$

$\Updownarrow$

$\boxed{\begin{aligned} &\max_{w,b} \frac{2}{\|w\|} \\ &\text{s.t. } y_i(w^T x_i + b) \geq 1 \\ &\qquad i = 1, 2, \cdots, N \end{aligned}} \Leftrightarrow \boxed{\begin{aligned} &\min \frac{1}{2} w^T w \\ &\text{s.t. } y_i(w^T x_i + b) \geq 1 \\ &\qquad i = 1, \cdots, N \end{aligned}}$

$\Updownarrow$

convex problem:

$\boxed{\begin{aligned} &\min \frac{1}{2} w^T w \\ &\text{st. } -y_i(w^T x_i + b) + 1 \leq 0 \\ &\qquad i = 1, \cdots, N \end{aligned}}$

⑧ Let $L(w, b, \lambda) = \frac{1}{2} w^T w - \sum_{i=1}^{N} \lambda_i [y_i(w^T x_i + b) - 1]$, $\lambda_i \geq 0$, $i = 1, \cdots, N$

~~Analysis: If $\exists\, i$, $y_i(w^T x_i + b) \geq 1$ is violated, then let $\lambda_i = +\infty \Rightarrow L = +\infty$.~~

~~else for all $i = 1, \cdots, N$, $y_i(w^T x_i + b) \geq 1$ holds, then let $\lambda_i = 0$, $\Rightarrow L = \frac{1}{2} w^T w$.~~

~~Conclusion: Solving~~ $\frac{\partial L}{\partial w} = w - \sum_{i=1}^{N} \lambda_i y_i x_i = 0 \Rightarrow w = \sum_{i=1}^{N} \lambda_i y_i x_i$

$\frac{\partial L}{\partial b} = -\sum_{i=1}^{N} \lambda_i y_i = 0 \Rightarrow \sum_{i=1}^{N} \lambda_i y_i = 0$

Therefore, the dual function of the primal convex problem is

$$g(\lambda) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j x_i^T x_j - \sum_{i=1}^{N} \lambda_i$$

$\boxed{\begin{aligned} &\max g(\lambda) \\ &\sum_{i=1}^{N} \lambda_i y_i = 0 \\ &\lambda_i \geq 0, \ i = 1, \cdots, N \end{aligned}}$ is a ~~unconstrain~~ quadratic programing, $\Rightarrow \hat{\lambda}$, then

$\hat{w} = \sum_{i=1}^{N} \hat{\lambda}_i y_i x_i$. $\exists\, j$, such that $y_j(x_j^T \hat{w} + \hat{b}) = 1$

$\Rightarrow \hat{b} = y_j - \hat{w} x_j$

Sequential Minimal Optimization for SVM

guorui233@outlook.com

May 19, 2023

It's May 18, 2023.

1 Sequential Minimal Optimization for SVM

Define $g(\alpha_1, \cdots, \alpha_N) = \frac{1}{2}\sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j x_i^T x_j - \sum_{i=1}^N \alpha_i$. Then, the linearly constrained quadratic problem related to SVM is

$$
\begin{aligned}
\underset{\alpha_1,\cdots,\alpha_N}{\text{maximize}} \quad & g(\alpha_1, \cdots, \alpha_n), \\
\text{subject to:} \quad & \sum_{i=1}^N \alpha_i y_i = 0, \\
& \alpha_i \geqslant 0, i = 1, \cdots, N.
\end{aligned} \tag{1}
$$

To solve $\alpha_1, \cdots, \alpha_N$, the basic idea of sequential minimal optimization (SMO) is to let $\alpha_3, \cdots, \alpha_N$ be fixed, which enables us to solve α_1, α_2 at first. Once α_1, α_2 are solved, we can repeat the process to solve α_3, α_4 in a same way, and stop untill α_{N-1}, α_N are solved.

Under the condition that $\alpha_3, \cdots, \alpha_N$ are fixed, we have $y_1\alpha_1 + y_2\alpha_2 = \gamma$, where $\gamma = -\sum_{i=3}^N \alpha_i y_i$. Remember that $y_1, y_2 \in \{-1, +1\}$ and thus $y_i^2 = 1, i = 1, 2$. And let $s = y_1 y_2$.

$$
\begin{aligned}
& g(\alpha_1, \alpha_2) = \frac{1}{2}(\alpha_1^2 x_1^T x_1 + \alpha_2^2 x_2^T x_2) + \alpha_1\alpha_2 s x_1^T x_2 + \alpha_1 y_1 v_1 + \alpha_2 y_2 v_2 - (\alpha_1 + \alpha_2) + const, \\
& v_1 = \sum_{j=3}^N \alpha_j y_j x_j^T x_1, \\
& v_2 = \sum_{j=3}^N \alpha_j y_j x_j^T x_2, \\
& const = \frac{1}{2}\sum_{i=3}^N \sum_{j=3}^N \alpha_i \alpha_j y_i y_j x_i^T x_j - \sum_{i=3}^N \alpha_i.
\end{aligned} \tag{2}
$$

Replacing α_1 in Eq. (2) via $\alpha_1 = y_1 \cdot y_1\alpha_1 = y_1 \cdot (\gamma - \alpha_2 y_2) = y_1\gamma - s\alpha_2, s = y_1 y_2$, results in

$$\begin{aligned} g(\alpha_2) = \;& \frac{1}{2}((y_1\gamma - s\alpha_2)^2 x_1^T x_1 + \alpha_2^2 x_2^T x_2) + (y_1\gamma - s\alpha_2)\alpha_2 s x_1^T x_2 \\ & + (y_1\gamma - s\alpha_2)y_1 v_1 + \alpha_2 y_2 v_2 - (y_1\gamma - s\alpha_2 + \alpha_2) + const \\ = \;& \frac{1}{2}((y_1\gamma - s\alpha_2)^2 x_1^T x_1 + \alpha_2^2 x_2^T x_2) + (y_2\gamma\alpha_2 - \alpha_2^2) x_1^T x_2 \\ & + (\gamma - y_2\alpha_2)v_1 + \alpha_2 y_2 v_2 - (y_1\gamma - s\alpha_2 + \alpha_2) + const \end{aligned} \tag{3}$$

which leads to

$$\begin{aligned} \frac{\partial g(\alpha_2)}{\partial \alpha_2} &= (\alpha_2 - y_2\gamma)x_1^T x_1 + \alpha_2 x_2^T x_2 + (y_2\gamma - 2\alpha_2)x_1^T x_2 + y_2(v_2 - v_1) + s - 1 \\ &= (x_1^T x_1 + x_2^T x_2 - 2x_1^T x_2)\alpha_2 + (x_1^T x_2 - x_1^T x_1)y_2\gamma + y_2(v_2 - v_1) + s - 1 \end{aligned} \tag{4}$$

By $\frac{\partial g(\alpha_2)}{\partial \alpha_2} = 0$, we then have

$$(x_1^T x_1 + x_2^T x_2 - 2x_1^T x_2)\alpha_2 + (x_1^T x_2 - x_1^T x_1)y_2\gamma + (v_2 - v_1)y_2 + s - 1 = 0. \tag{5}$$

Up to now, we still can not solve α_2 analytically from the equation above, because $\alpha_3, \cdots, \alpha_N$ are unknown thus γ, v_1, v_2 are unknown. But, we have

$$\gamma = \gamma_{fixed} = \alpha_1^{new} y_1 + \alpha_2^{new} y_2 = \alpha_1^{old} y_1 + \alpha_2^{old} y_2, \tag{6}$$

thus γ is known. We also have $f(x_i) = w^T x_i + b = \sum_{j=1}^N \alpha_j y_j x_j^T x_i + b$. Compared with Eq. (2), v_1, v_2 are now known:

$$\begin{aligned} v_1 &= f(x_1) - b - \alpha_1 y_1 x_1^T x_1 - \alpha_2 y_2 x_2^T x_1, \\ v_2 &= f(x_2) - b - \alpha_1 y_1 x_1^T x_2 - \alpha_2 y_2 x_2^T x_2. \end{aligned} \tag{7}$$

Therefore, Eq. (5) can be re-wrote as

$$\begin{aligned} & (x_1^T x_1 + x_2^T x_2 - 2x_1^T x_2)\alpha_2^{new} \\ = \;& 1 - s + (x_1^T x_1 - x_1^T x_2)y_2\gamma + (v_1 - v_2)y_2 \\ = \;& y_2[y_2 - y_1 + (x_1^T x_1 - x_1^T x_2)\gamma + v_1 - v_2] \\ = \;& y_2\{y_2 - y_1 + (x_1^T x_1 - x_1^T x_2)(\alpha_1 y_1 + \alpha_2 y_2) + [\big(f(x_1) - b - \alpha_1 y_1 x_1^T x_1 - \alpha_2 y_2 x_2^T x_1\big) \\ & - \big(f(x_2) - b - \alpha_1 y_1 x_1^T x_2 - \alpha_2 y_2 x_2^T x_2\big)]\} \\ = \;& y_2\{y_2 - y_1 + \big(\alpha_2 y_2 x_1^T x_1 - \alpha_2 y_2 x_1^T x_2\big) + \big(f(x_1) - \alpha_2 y_2 x_2^T x_1 \\ & - f(x_2) + \alpha_2 y_2 x_2^T x_2\big)\} \\ = \;& y_2\{y_2 - y_1 + \big(x_1^T x_1 - 2x_1^T x_2 + x_2^T x_2\big)y_2\alpha_2 + f(x_1) - f(x_2)\} \\ = \;& \big(x_1^T x_1 - 2x_1^T x_2 + x_2^T x_2\big)\alpha_2^{old} + y_2[\big(f(x_1) - y_1\big) - \big(f(x_2) - y_2\big)] \end{aligned} \tag{8}$$

which allows us to solve α_2 numerically, i.e., $\alpha_2^{new} = \alpha_2^{old} + \frac{y_2(E_1 - E_2)}{x_1^T x_1 + x_2^T x_2 - 2x_1^T x_2}$, where $E_i = f(x_i) - y_i, i = 1, 2$.

2 Bibliography

References

[1] Platt, John. "Sequential minimal optimization: A fast algorithm for training support vector machines." (1998).