Jake Schinto

# Homework 5

## General SVM

1. $\min_{w,b} \max_{\substack{\alpha_i \geq 0 \\ i=1,\dots,m}} \left( \frac{1}{2}\|w\|^2 + \alpha_i\left(1 - y^{(i)} z^{(i)}\right)\right)$

$= \frac{1}{2}\|w^*\|^2 + \alpha^*(1 - y^i z^i)$

in order to maximize the Lagrange it must satisfy KKT conditions

$\alpha \geq 0$

$f(x) \geq 0$

$\alpha f(x) = 0$

Therefore: $\alpha^*(1 - y^i z^i) = 0$

So, $\frac{1}{2}\|w^*\|^2 + 0 = \frac{1}{2}\|w^*\|^2$

2.

$\arg\max_{w,b} \left\{ \frac{1}{\|w\|} \min_i \left[ y^{(i)} \left( w^T x^{(i)} + b \right) \right] \right.$

when $y^{(i)} = -1 \rightarrow w^T x^{(i)} + b \leq -1$

when $y^{(i)} = 1 \rightarrow w^T x^{(i)} + b \geq 1$

So, to minimize $y^{(i)}(w^T x^{(i)} + b) \geq 1$ with $i$ would be to equal 1

Therefore, $\frac{1}{\|w^*\|} \cdot 1 = 1/\|w^*\|$

3. $w = \sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)}$

$w = \sum_{i=1}^{2} \alpha_i y^{(i)} x^{(i)} = \alpha_1 \cdot y^{(1)} \cdot x^{(1)} + \alpha_2 \cdot y^{(2)} x^{(2)}$

$1 \cdot 1 \cdot \langle 1, 1 \rangle + 1 \cdot (-1) \cdot \langle -1, -1 \rangle = \langle 1, 1 \rangle + \langle 1, 1 \rangle = \langle 2, 2 \rangle$

$w^T x^{(0)} + b = \langle 2, 2 \rangle \langle 1, -1 \rangle + 1 = 2 - 2 + 1 = 1$

$y_0 = 1$ if $t_0 = 1$ hinge $= 1 - (1) \cdot (1) = 0$

if $t_0 = -1$ hinge $= 1 - (1) \cdot (-1) = 2$

4. $K(x,x') = (1+\langle x,x'\rangle)'$ $x_1 = \langle 1,1\rangle$ $x_2 = \langle -1,-1\rangle$

$$K = \begin{array}{c c} & \begin{array}{cc} x_1 & x_2 \end{array} \\ \begin{array}{c} x_1 \\ x_2 \end{array} & \begin{bmatrix} 2 & -2 \\ -2 & 2 \end{bmatrix} \end{array}$$ ← using linear kernal

$$K(x,x') = \exp\left\{-\frac{1}{\sigma^2}\|x-x'\|^2\right\}$$

$$K = \begin{bmatrix} \exp\{-1(0+0)\} & \exp\{-1(4+4)\} \\ \exp\{-1(4+4)\} & \exp\{-1(0+0)\} \end{bmatrix} = \begin{bmatrix} 1 & e^{-8} \\ e^{-8} & 1 \end{bmatrix}$$

$$(1+1) - \frac{1}{2}[1,-1]\begin{bmatrix} 2 & -2 \\ -2 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ -1 \end{bmatrix} = 2 - \frac{1}{2}[4\ -4]\begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$= 2 - \frac{1}{2}(8) = -2$$

5. $$K = \begin{bmatrix} x\cdot x & x\cdot x' \\ x'\cdot x & x'\cdot x' \end{bmatrix} \qquad y = \langle y_1, y_2\rangle$$

$$y\cdot K\cdot y^T \qquad \begin{bmatrix} y_1\cdot x\cdot x + y_2\cdot x'\cdot x & y_1\cdot x\cdot x' + y_2 x'\cdot x' \end{bmatrix}\cdot y^T$$

$$y_1^2\cdot x\cdot x + y_1 y_2 x'\cdot x + y_1 y_2\cdot x\cdot x' + y_2^2\cdot x'\cdot x' = y_1^2\cdot x^2 + 2y_1 y_2 x\cdot x' + y_2^2 x'^2$$
$$= (y_1 x + y_2 x')^2 \geq 0 \text{ for all } y$$

Therefore the linear kernal is positive semi-definate

$$K = \begin{bmatrix} f(x)\cdot x\cdot f(x)\cdot x & f(x)\cdot x\cdot f(x')\cdot x' \\ f(x')\cdot x'\cdot f(x)\cdot x & f(x')\cdot x'\cdot f(x')\cdot x' \end{bmatrix}$$ let $x_1 = f(x)\cdot x$ let $x_2 = f(x')\cdot x$

$$K = \begin{bmatrix} x_1\cdot x_1 & x_1\cdot x_2 \\ x_2\cdot x_1 & x_2\cdot x_2 \end{bmatrix} \qquad y = \langle y_1, y_2\rangle$$

$y \cdot K \cdot y^T = y_1^2 \cdot x_1^2 + 2 y_1 y_2 x_1 x_2 + y_2^2 \cdot x_2^2$

$= (y_1 x_1 + y_2 x_2)^2 = (y_1 f(x) \cdot x + y_2 f(x') x')^2 \geq 0$ for all $y$

which is positive semi-definite for any $f: X \to \mathbb{R}$

6. $a_n \geq 0$

$t_n y(x_n) - 1 + \varepsilon_n \geq 0$

$a_n (t_n y(x_n) - 1 + \varepsilon_n) = 0$

$\mu_n \geq 0$

$\varepsilon_n \geq 0$.

$\mu_n \varepsilon_n = 0$

$\rightarrow$ $y(x_n) = w^T x^{(i)} + b$ $\quad t_n = y^i$

$\tilde{L}(a) = \max \sum_{i=1}^{m} a_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} a_i a_j y_i y_j \langle x^{(i)}, x^{(j)} \rangle$

- when $a_i = 0$

$0 \cdot (y^{(i)}(w^T x^{(i)} + b) - 1 + \varepsilon_n) = 0$

which gives no new predictive information

so, $y^{(i)}(w^T x^{(i)} + b) - 1 + \varepsilon_n \geq 0$

$y^{(i)}(w^T x^{(i)} + b) \geq 1 - \varepsilon_n$

or $y^{(i)}(w^T x^{(i)} + b) \geq 1$ when $\varepsilon_n = 0$

- when $a_i > 0$

it must satisfy $(y^{(i)}(w^T x^{(i)} + b) - 1 + \varepsilon_n) = 0$

or $y^{(i)}(w^T x^{(i)} + b) = 1 - \varepsilon_n$

- if $a_i < C$ the $\varepsilon_n$ must equal 0

so, $y^{(i)}(w^T x^{(i)} + b) = 1$

- if $a_i = C$ points will be classified correctly

since $\varepsilon_n \geq 0$ -

and $\varepsilon_n = 1 - y^{(i)}(w^T x^{(i)} + b)$

then $1 - y^{(i)}(w^T x^{(i)} + b) \geq 0$ or $y^{(i)}(w^T x^{(i)} + b) \leq 1$

SMO algorithm for SVM

1. $E_i = w^T x^{(i)} + b - y^{(i)}$

$y^{(i)} E_i = y^{(i)} \left( w^T x^{(i)} + b - y^{(i)} \right)$

$= y^{(i)} (w^T x^{(i)} + b) - y^{(i)2}$ where $y^{(i)} = 1, -1$

so, $y^{(i)2} = 1$

$y^{(i)} E_i = y^{(i)} (w^T x^{(i)} + b) - 1$

2. $y^{(i)} E_i < -tol$ when $a_i < C$

$y^{(i)} E_i = y^{(i)} (w^T x^{(i)} + b) - 1 < -tol$

$y^{(i)} (w^T x^{(i)} + b) < 1 - tol$

from #6: if $a_i < C \rightarrow y^{(i)} (w^T x^{(i)} + b) \geq 1$

therefore this part of the if will catch things outside the constraints because they are opposites

$y^{(i)} E_i > tol$ when $a_i > 0$

$y^{(i)} E_i = y^{(i)} (w^T x^{(i)} + b) - 1 > tol$

$y^{(i)} (w^T x^{(i)} + b) > 1 + tol$

from #6: if $a_i > 0 \rightarrow y^{(i)} (w^T x^{(i)} + b) \leq 1$

therefore this part of the condition will also find pairs that violate KKT

3. $a_j = a_j^{old} - \dfrac{y^{(j)} (E_i - E_j)}{\eta}$

$L = \max \{ 0, a_j^{old} - a_i^{old} \}$ $H = \min \{ C, C + a_j^{old} - a_i^{old} \}$

if $a_i^{old} \geq a_j^{old}$ then $L = 0$ $H = C + a_j^{old} - a_i^{old}$

if $a_j^{old} \geq a_i^{old}$ then $L = a_j^{old} - a_i^{old}$ $H = C$

$y^{(j)} E_i - y^{(j)} E_j = y^{(j)} (w^T x^{(i)} + b) + 1 - y^{(j)} (w^T x^{(j)} + b) + 1$

$$\underbrace{a_j}_{\text{includes } [0, C]} = a_j^{old} - \underbrace{\frac{y^{(j)} \left( w^T \left( x^{(i)} - x^{(j)} \right) \right) + 2}{\eta}}_{a_i^{old}}$$

therefore, $[L, H]$ is a better bound for $a_j$ than $[0, C]$ because of the algorithm

4. $K(x^{(i)}, x^{(j)})$ is a kernal function and is equivalent to $\langle \Phi(x^{(i)}), \Phi(x^{(j)}) \rangle$
$= \|\Phi(x^{(i)})\| \cdot \|\Phi(x^{(j)})\|$
So, $K(x^{(i)}, x^{(i)}) + K(x^{(j)}, x^{(j)}) - 2K(x^{(i)}, x^{(j)})$
$= \langle \Phi(x^i), \Phi(x^i) \rangle + \langle \Phi(x^j), \Phi(x^j) \rangle - 2\langle \Phi(x^i), \Phi(x^j) \rangle$
$= \|\Phi(x^i)\| \cdot \|\Phi(x^i)\| + \|\Phi(x^j)\| \cdot \|\Phi(x^j)\| - 2\|\Phi(x^i)\| \cdot \|\Phi(x^j)\|$
$= \|\Phi(x^i)\|^2 + \|\Phi(x^j)\|^2 - 2\|\Phi(x^i)\| \cdot \|\Phi(x^j)\|$
$= \|\Phi(x^{(i)}) - \Phi(x^{(j)})\|^2$ by quadratic form

5.

6. if $a_i = 0$ and $y^{(i)} \neq y^{(j)}$
then the equation would take the form: $a_i = 0 + (-1)(a_j^{old} - a_j)$
which can be negative if $a_j^{old} > a_j$
this would set $a_i$ to a negative number which is outside the bounds of $[0, C]$
on the other end if $a_i = C$ and $y^{(i)} = y^{(j)}$
$a_i = C + (1)(a_j^{old} - a_j)$ again if $a_j^{old} > a_j$ it is out of bounds

7.

8.