Liheng Cao

April 5, 2021

1

(a)

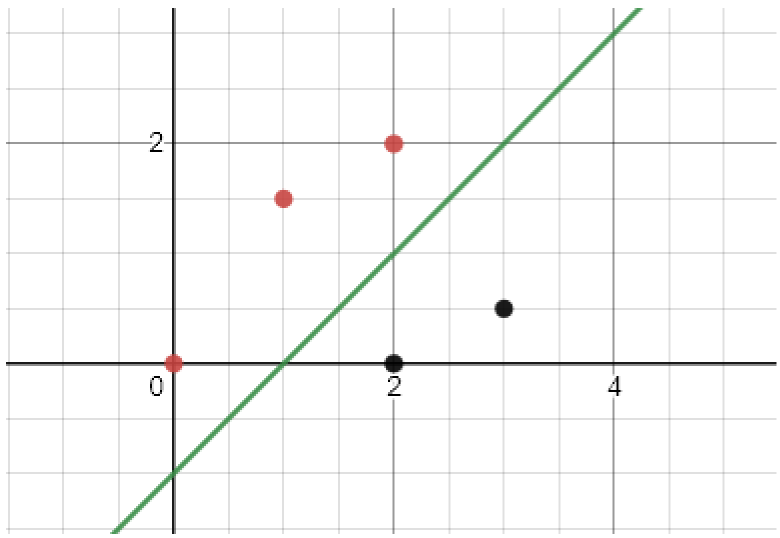

Figure 1: horizontal axis: x_0, vertical axis: x_1

(b) They are linearly separable.

(c) i.

x_1	x_2	y	functional margin
0	0	-1	-3
2	2	-1	-3
2	0	1	3
1	1.5	-1	-4.5
3	0.5	1	4.5

The absolute minimum is 3.

ii. To find the geometric margin, we just divide everything by $\|w\|_2 = \sqrt{3^2 + 3^2} = 3\sqrt{2}$.

x_1	x_2	y	geometric margin
0	0	-1	$-\frac{1}{\sqrt{2}}$
2	2	-1	$-\frac{1}{\sqrt{2}}$
2	0	-1	$\frac{1}{\sqrt{2}}$
1	1.5	-1	$-\frac{3}{2\sqrt{2}}$
3	0.5	1	$\frac{3}{2\sqrt{2}}$

The absolute minimum is $\dfrac{1}{\sqrt{2}}$

iii. We divide w, w_0 by the functional margin to get the canonical weights. They are $[3, -3]/3 = [1, -1], -3/3 = -1$

(d) The first 3 examples are support vectors because they have the same absolute margin (doesn't matter which type of margin since they are just a constant multiple of each other).

(e) $(1, 3)^T(1, -1) - 1 = 1 - 3 - 1 = -3$. $y(x^T w_{canon} - w_{0,canon}) > 1$ meaning it is on the correct side of the hyperplane and it is not a support vector. Since it's the value is greater than 1, the margin does not change.

(f) Without doing any calculations, neither of them change. It isn't a support vector to begin with and there are multiple support vectors, so removing one of them wouldn't change the margins. There is no risk of a data set begin no longer separable if we remove something; this risk is only present when we're adding something.

(g) Without doing any calculations, neither of them change. Even though this point is a support vector, it was not the only one, meaning that that the canonical margins would not change. Likewise, the separating plane would not change because there is no risk when removing a point.

(h) minimize $w_0^2 + w_1^2 + w_2^2$

subject to :

$$
\begin{aligned}
-1 \cdot (w_0 + 0 \cdot w_1 + 0 \cdot w_2) &\geq 1 \\
-1 \cdot (w_0 + 2 \cdot w_1 + 2 \cdot w_2) &\geq 1 \\
1 \cdot (w_0 + 2 \cdot w_1 + 0 \cdot w_2) &\geq 1 \\
-1 \cdot (w_0 + 1 \cdot w_1 + 1.5 \cdot w_2) &\geq 1 \\
1 \cdot (w_0 + 3 \cdot w_1 + 0.5 \cdot w_2) &\geq 1
\end{aligned}
$$

2

Yes.

max γ, subject to :

$$
\begin{aligned}
y^{(i)}(w_0 + \mathbf{w}^T x^{(i)}) &\geq \gamma \\
\|\mathbf{w}\|_2 &= 1
\end{aligned}
$$

is equivalent to

min $\|\mathbf{w}\|_2^2$, subject to :

$$
y^{(i)}(w_0 + \mathbf{w}^T x^{(i)}) \geq 1
$$

Both are trying to find the parameters that make the margin as large as possible. The first tries to maximize the margin directly, and keeps the magnitude of the parameters constant to prevent the "constant multiple" trick. The second takes the opposite approach: it tries to minimize the parameters, while keeping margin a constant value.

3

$$\xi^{(i)} = \begin{cases} 0 & y^{(i)}(\mathbf{w}^T\mathbf{x}^{(i)}) \geq 1 \\ 1 - y^{(i)}(\mathbf{w}^T\mathbf{x}^{(i)}) & \text{otherwise} \end{cases}$$

(a) If $\xi^{(i)} = 0$, the point is correctly classified and outside of the margin.

(b) If $0 < \xi^{(i)} < 1$, the point is correctly classified, but it is very close to the hyperplane and inside the margin.

(c) If $\xi^{(i)} > 1 \rightarrow y^{(i)}(\mathbf{w}^T\mathbf{x}^{(i)}) < 0$, the point is incorrectly classified.

4

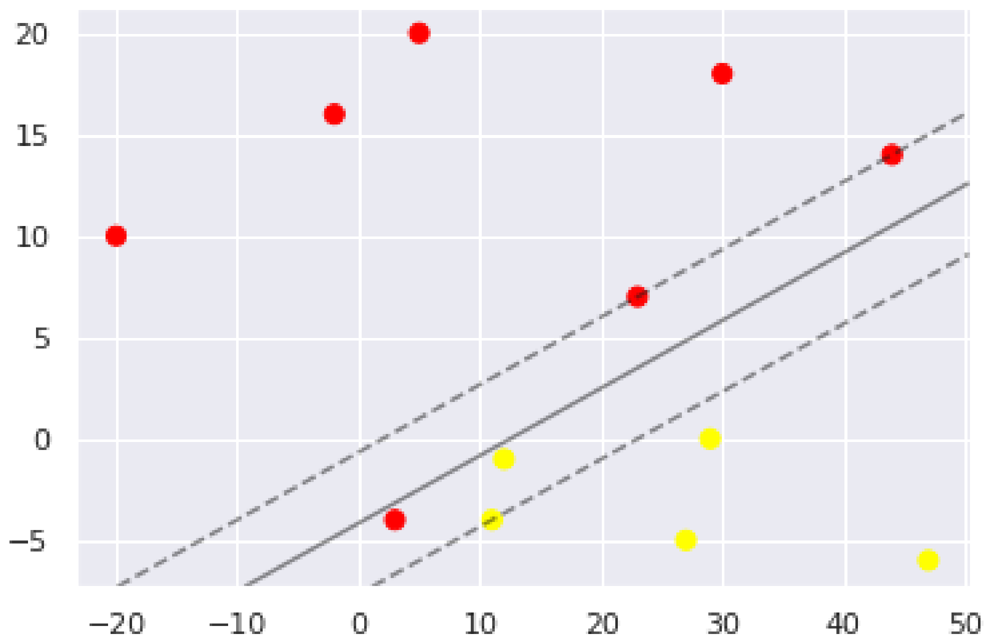

Figure 2: C=0.1

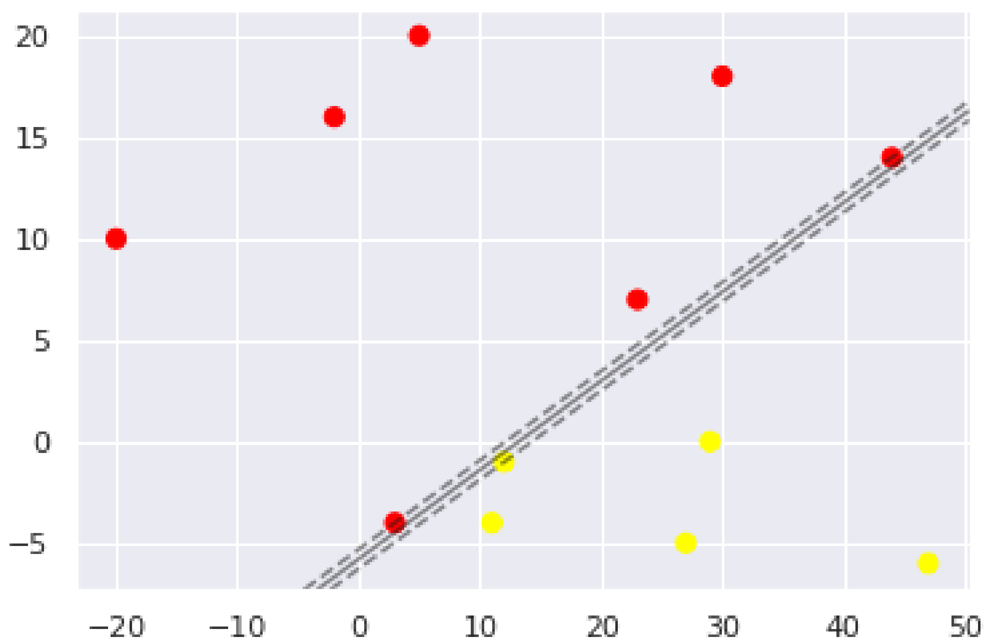

Figure 3: c=10

As expected, the margin is inversely related with the penalty.

5

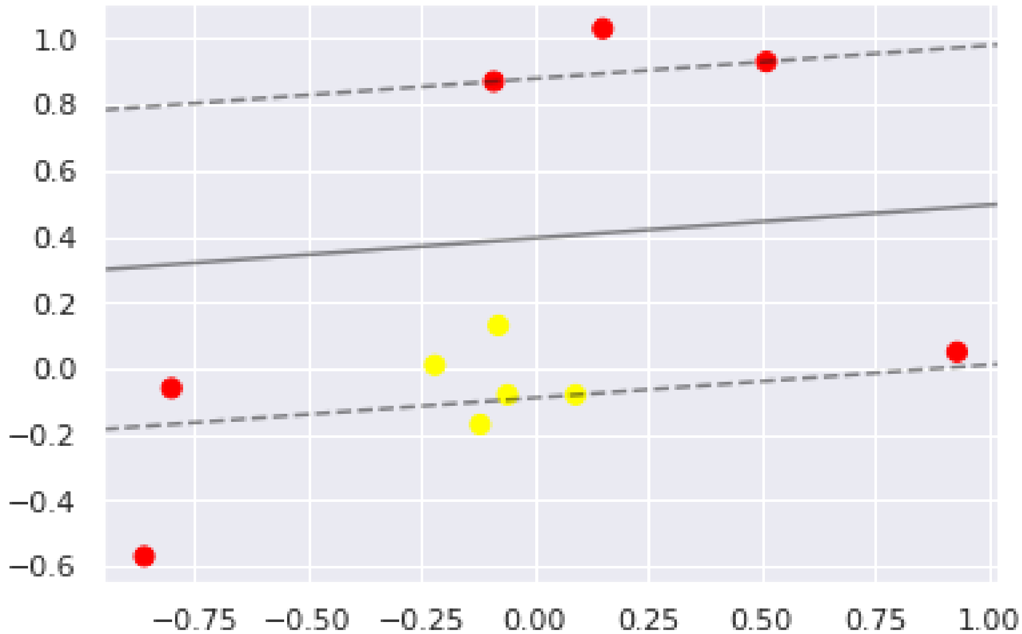

Figure 4: Before transformation

It looks like we can either use RBF or Polynomial (quadratic that faces downward) Kernel.

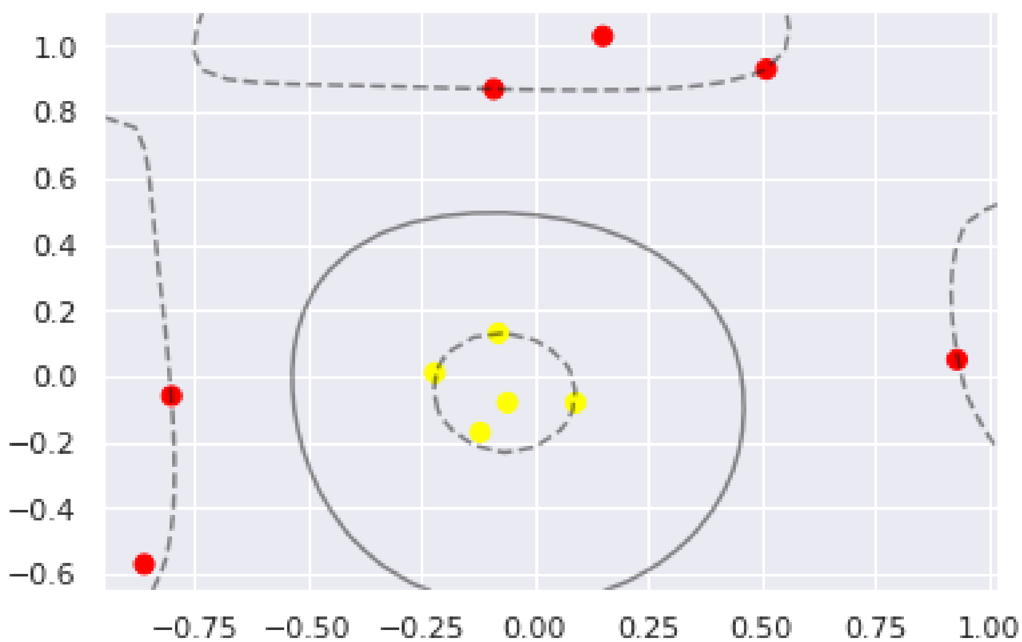

Figure 5: After RBF Kernel

6

Assume typo: $(-0.08, -0.06) \implies (-0.06, -0.08)$.

That point is not a support vector, as it is not on the dotted lines, and the point's removal will not change the margin. If you remove a support vector, there is a possibility that you could find a better margin. However, this is not guaranteed.

7

$$
\begin{aligned}
\Phi(\mathbf{x})^T\Phi(\mathbf{x}') &= (1+\mathbf{x}^T\mathbf{x}')^2 \\
(1, \sqrt{2}x_{i1}, \sqrt{2}x_{i2}, x_{i1}^2, x_{i2}^2, \sqrt{2}x_{i1}x_{i2})^T & \\
\times(1, \sqrt{2}x'_{i1}, \sqrt{2}x'_{i2}, x_{i1}^{2'}, x_{i2}^{2'}, \sqrt{2}x'_{i1}x'_{i2}) &= 1 + 2\mathbf{x}^T\mathbf{x}' + (\mathbf{x}^T\mathbf{x}')^2 \\
\cancel{1} + 2x_{i1}x'_{i1} + 2x_{i2}x'_{i2} + x_{i1}^2x_{i1}^{2'} + x_{i2}^2x_{i2}^{2'} + 2x_{i1}x_{i2}x'_{i1}x'_{i2} &= \cancel{1} + 2[x_{i1}, x_{i2}]^T[x'_{i1}, x'_{i2}] \\
&\quad + ([x_{i1}, x_{i2}]^T[x'_{i1}, x'_{i2}])^2 \\
\cancel{2x_{i1}x'_{i1} + 2x_{i2}x'_{i2}} + x_{i1}^2x_{i1}^{2'} + x_{i2}^2x_{i2}^{2'} + 2x_{i1}x_{i2}x'_{i1}x'_{i2} &= \cancel{2x_{i1}x'_{i1} + 2x_{i2}x'_{i2}} + (x_{i1}x'_{i1} + x_{i2}x'_{i2})^2 \\
\cancel{x_{i1}^2x_{i1}^{2'} + x_{i2}^2x_{i2}^{2'} + 2x_{i1}x_{i2}x'_{i1}x'_{i2}} &= \cancel{x_{i1}^2x_{i1}^{2'} + 2x_{i1}x'_{i1}x_{i2}x'_{i2} + x_{i2}^2x_{i2}^{2'}} \\
&\square
\end{aligned}
$$