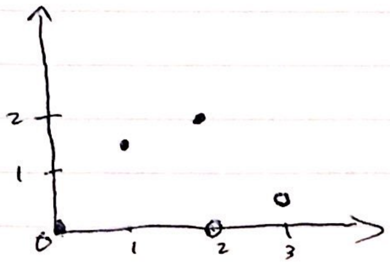
Michael Kwan
Machine Learning
Homework 5

1a.



b. yes, linearly separable.

c. i. $\gamma = \min_{i=1..N} \gamma^{(i)} = \min_{i=1..N} y^{(i)}(w^T x^{(i)} + w_0) = \min_{i=1..N} y^{(i)} \cancel{(w^T x^{(i)} + w_0)} ([3, -3] x^{(i)} + -3)$

$(0,0,-1) \to -1(-3) = 3$
$(2,2,-1) \to -1(-3) = 3$
$(2,0,1) \to 1(3) = 3$
$(1,1.5,-1) \to -1(3+1.5(-3)-3) = 4.5$
$(3,.5,1) \to 1(3(3)+(-3)(.5)-3) = 4.5$

min = 3
$\boxed{\text{functional margin} = 3}$

ii. geometric margin $= \dfrac{\text{functional margin}}{\|w\|} = \dfrac{\text{functional margin}}{\|(3,-3)\|} = \dfrac{\text{functional margin}}{3\sqrt{2}}$

$(0,0,-1) \to \frac{3}{3\sqrt{2}} = \frac{\sqrt{2}}{2}$
$(2,2,-1) \to \frac{3}{3\sqrt{2}} = \frac{\sqrt{2}}{2}$
$(2,0,1) \to \frac{3}{3\sqrt{2}} = \frac{\sqrt{2}}{2}$
$(1,1.5,-1) \to \frac{4.5}{3\sqrt{2}} = \frac{3\sqrt{2}}{4}$
$(3,.5,1) \to \frac{4.5}{3\sqrt{2}} = \frac{3\sqrt{2}}{4}$

min $= \frac{\sqrt{2}}{2}$
$\boxed{\text{geometric margin} = \frac{\sqrt{2}}{2}}$

iii. $\grave{w} = \frac{w}{3} = \frac{(3,-3)}{3} = \boxed{(1,-1)}$
$\grave{w}_0 = \frac{w_0}{3} = \frac{-3}{3} = \boxed{-1}$

d. (0,0), (2,2), and (2,0) are support vectors.

e. No, because the point is far from the boundary. The hyperplane and the support vectors also won't change becuse the point doesn't change the margin.

f. Margin doesn't change becuse there are support vectors still there and (1,1.5) is not a support vector. The hyperplane doesn't change.

g. Since (0,0) is a support vector, the margin and the hyperplane could change.

h. $\min \|w\|_2^2 = \min w_1^2 + w_2^2 + \dots + w_d^2$

subject to $y^{(i)}(w_0 + w^T x^{(i)}) \geq 1$ for all $i = 1, \dots, N$

$-1(w_0 + w^T \begin{bmatrix} 0 \\ 0 \end{bmatrix} \geq 1$

$-1(w_0 + w^T \begin{bmatrix} 2 \\ 2 \end{bmatrix} \geq 1$

$+1(w_0 + w^T \begin{bmatrix} 2 \\ 0 \end{bmatrix} \geq 1$

$-1(w_0 + w^T \begin{bmatrix} 1 \\ 1.5 \end{bmatrix} \geq 1$

$+1(w_0 + w^T \begin{bmatrix} 3 \\ 0.5 \end{bmatrix} \geq 1$

2. Yes, the second constraint optimization is just a different way of writing the first one by dividing by γ. They produce the same decision boundary.

3a. If $\varepsilon^{(i)} = 0$, $y^{(i)}(w^T x^{(i)} + w_0) \geq 1$, and the $x^{(i)}$ has a margin of 1 from the hyperplane and is correctly classified.

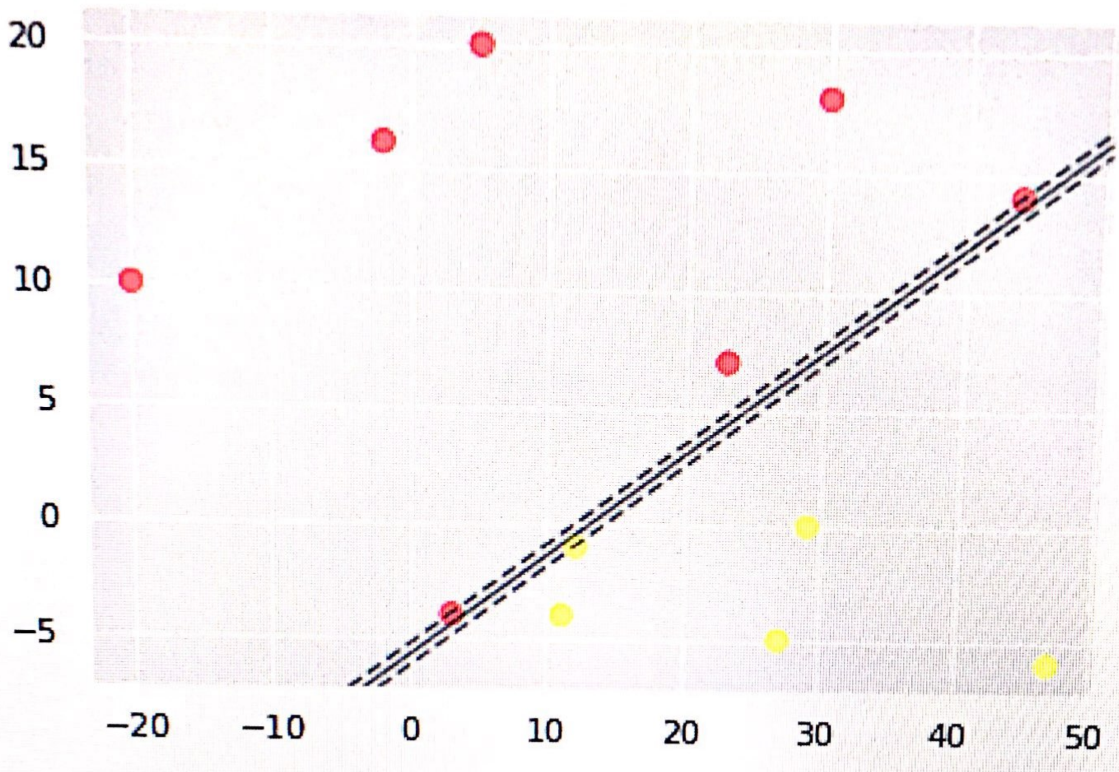
If $0 < \varepsilon^{(i)} \leq 1$, $0 < y^{(i)}(w^T x^{(i)} + w_0) < 1$, and the $x^{(i)}$ has a margin between 0 and 1. So, it is inside the margin and correctly classified.

If $\varepsilon^{(i)} > 1$, $x^{(i)}$ can be anywhere in the margin, or outside. You also cannot tell if $x^{(i)}$ is classified correctly or not.

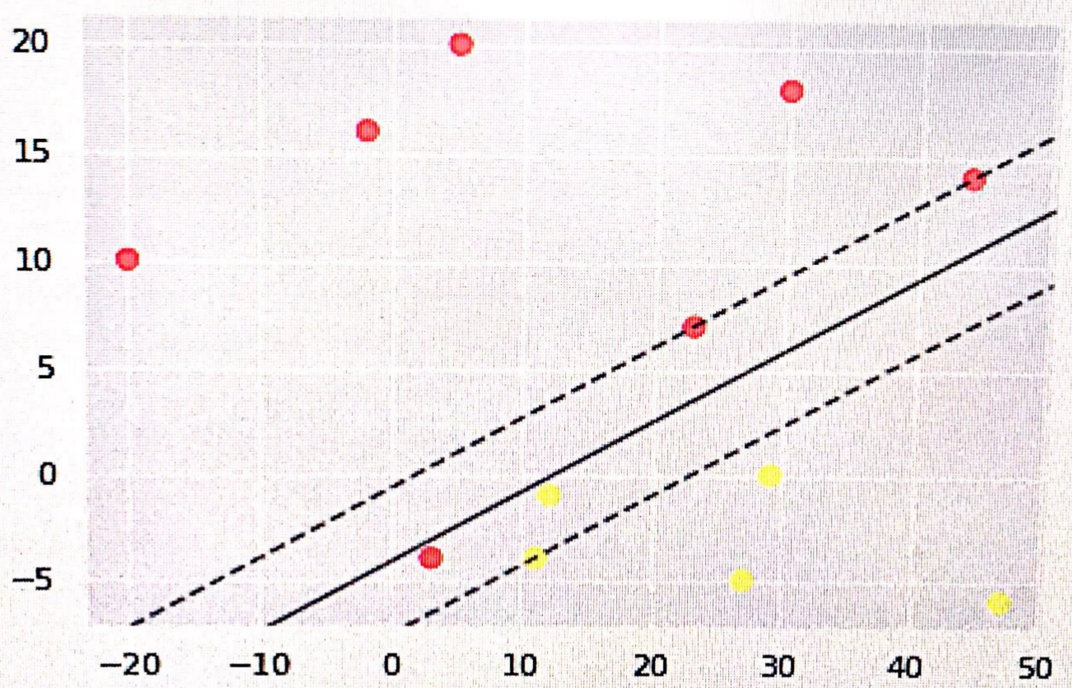
b.

Q4.

Plotted with decision boundary when C = 0.1.



Plotted with decision boundary when C = 10.



5. $e^{\frac{\|x - l\|_2^2}{2\sigma^2}} \rightarrow$ use $(0,0)$ for l $\rightarrow e^{\frac{\|x\|_2^2}{2\sigma^2}}$

$\sigma_1^2 = .376$

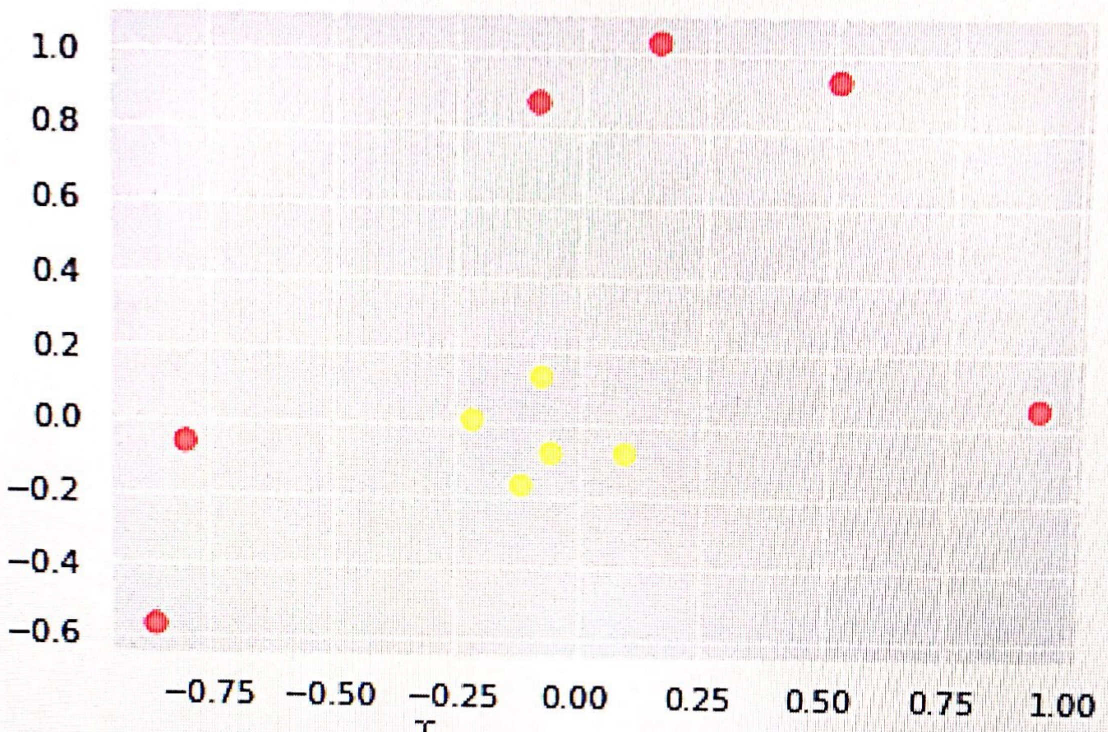
$\sigma_2^2 = .4536$

$$\boxed{(x_1, x_2) = \left(e^{\frac{x_1^2}{2\sigma_1^2}}, e^{\frac{x_2^2}{2\sigma_2^2}}\right)}$$

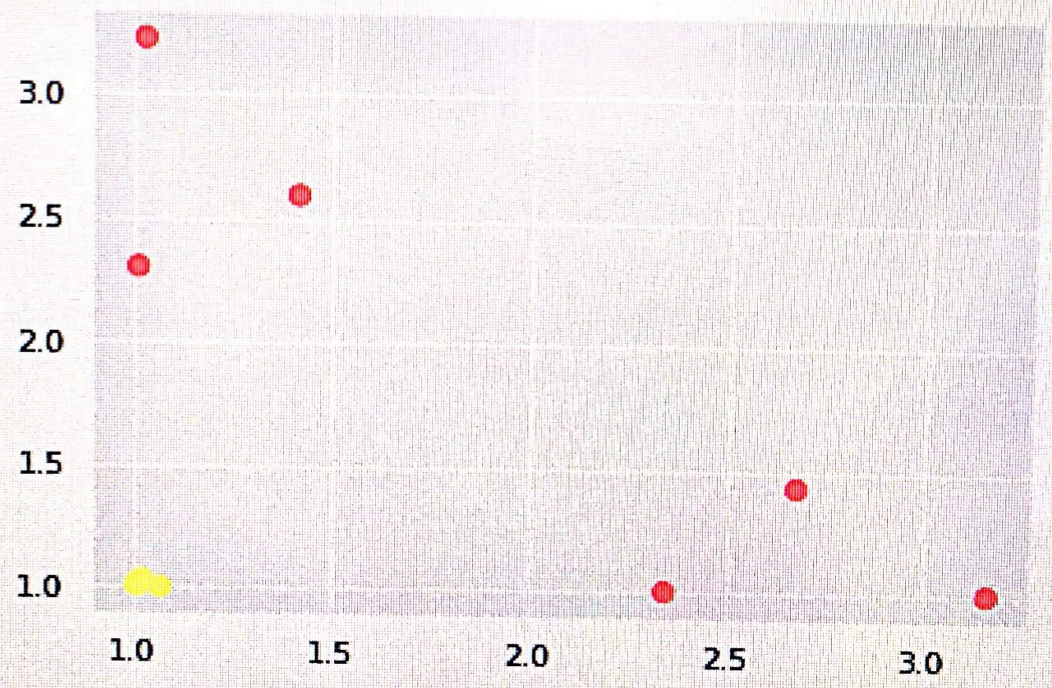
Scanned with CamScanner

Q5.

Plotted before transformation.



Plotted after transformation.



6. No, it will change the margin because it is the point closest to $l(0,0)$ that is classified as -1. So, it is a support vector and changing it will affect the margin. Another point is $(.22, .01)$ since it is the point furthest from $l(0,0)$ classified as 1. Other than these two, the removal of any other points will not immediately change the margin.

7. $\Phi(x) = (1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, x_2^2, \sqrt{2}x_1x_2)$

$\Phi(x') = (1, \sqrt{2}x_1', \sqrt{2}x_2', x_1'^2, x_2'^2, \sqrt{2}x_1'x_2')$

$\Phi(x)\cdot\Phi(x') = (\underline{1 + 2x_1x_1' + 2x_2x_2' + x_1^2x_1'^2 + x_2^2x_2'^2 + 2x_1x_2x_1'x_2'})$

$$
\begin{aligned}
\Phi(x)\cdot\Phi(x') &= (1 + x^T x')^2 \\
&= (1 + x_1x_1' + x_2x_2')^2 \\
&= (1 + x_1x_1' + x_2x_2')(1 + x_1x_1' + x_2x_2') \\
&= 1 + x_1x_1' + x_2x_2' + x_1x_1' + x_1^2x_1'^2 + x_1x_1'x_2x_2' + x_2x_2' + x_2x_2'x_1x_1' + x_2^2x_2'^2 \\
&= \underline{1 + 2x_1x_1' + 2x_2x_2' + x_1^2x_1'^2 + x_2^2x_2'^2 + 2x_1x_2x_1'x_2'}
\end{aligned}
$$

Same

$$\boxed{\Phi(x)\cdot\Phi(x') = (1 + x^Tx')^2}$$