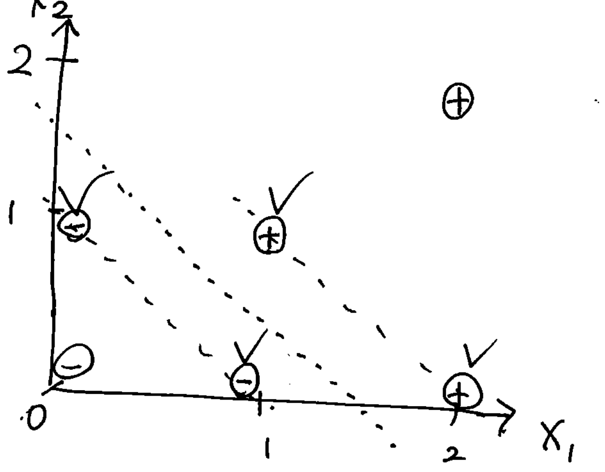
problem 1.

1)

a)



$\oplus$, $\ominus$ are linearly separable.

b) $(X_2 - 1.5) = -(X_1 - 1.5)$

$\vec{w} \cdot \vec{x} + b = X_1 + X_2 - 3 = 0$

$\rightarrow \vec{w} = (1, 1)$

$b = -3.$

c) Removing one support vectors will not change the margin since the other support vectors are still there.

Problem 2.

2a) ~~distance from~~ pg.13.

From the hard margin optimization problem,

$y^{(i)} (\vec{w} \cdot \vec{x}^{(i)} + b) \geq 1$ for all i.

$$\Downarrow$$

$$\frac{y^{(i)} (\vec{w} \cdot \vec{x}^{(i)} + b)}{\|\vec{w}\|} \geq \frac{1}{\|\vec{w}\|}$$

$$\parallel$$

$$\text{distance } |\delta^{(i)}| = \left| \frac{\vec{w} \cdot \vec{x}^{(i)} + b}{\|\vec{w}\|} \right| \geq \frac{1}{\|\vec{w}\|}$$

where the equality holds for the datapoint that is on the margin.

$\frac{1}{\|\vec{w}\|}$ is exactly the maximum margin because there is none other that can be greater than $\frac{1}{\|\vec{w}\|}$.

proof by contradiction: if we assume there exists a margin greater than $\frac{1}{\|\vec{w}\|}$, then there exists any $\varepsilon > 0$ s.t. $y^{(i)} (\vec{w} \cdot \vec{x}^{(i)} + b) \geq 1 + \varepsilon$.

Then $\exists$ $\vec{w}' = \frac{\vec{w}}{1+\varepsilon}$ and $b' = \frac{b}{1+\varepsilon}$ s.t. $\cancel{y^{(i)} (\vec{w}' \cdot \vec{x}^{(i)} + b') \geq 1}$

and $|\vec{w}' \cdot \vec{x}^{(i)} + b'| < |\vec{w} \cdot \vec{x}^{(i)} + b|$

which contradicts the condition that w and b are the optimal solutions.

margin $= \frac{1}{\|w\|}$

of w from pg.13

2b) For a ~~solution to the hard margin optimization problem on page 11,~~ $\vec{z}, d$, a separating hyperplane,

$$\frac{y^{(i)}(\vec{z}\cdot\vec{x}^{(i)}+d)}{|\vec{z}|} \geq M \quad \text{for all } i.$$

$$\Rightarrow \frac{y^{(i)}(\vec{z}\cdot\vec{x}^{(i)}+d)}{|\vec{z}|M} \geq 1$$

$$\Rightarrow y^{(i)}\left(\frac{\vec{z}}{|\vec{z}|M}\cdot\vec{x}^{(i)} + \frac{d}{|\vec{z}|M}\right)$$

$$= y^{(i)}(\vec{z}'\cdot\vec{x}^{(i)} + d') \geq 1 \quad \text{for all } i.$$

~~This satisfies the constraints of the optimization~~ problem.

thus $\vec{z}'$ and d' linearly separates all data points, and ~~there~~ is a feasible solution. for the hard margin problem ~~on pg~~.

Since $\vec{w}$ and b is the optimal solution, for pg. 13,

$|\vec{w}|^2 \leq |\vec{z}'|^2$ and $|\vec{w}| \leq |\vec{z}'|$.

2c) $$|\vec{z}'| = \sqrt{\vec{z}'\cdot\vec{z}'} = \sqrt{\frac{\vec{z}}{|\vec{z}|M}\cdot\frac{\vec{z}}{|\vec{z}|M}} = \frac{|\vec{z}|}{|\vec{z}|M} = \frac{1}{M}.$$

since $|\vec{w}| \leq |\vec{z}'|$ $\quad \frac{1}{|\vec{w}|} \geq \frac{1}{|\vec{z}'|} = M.$

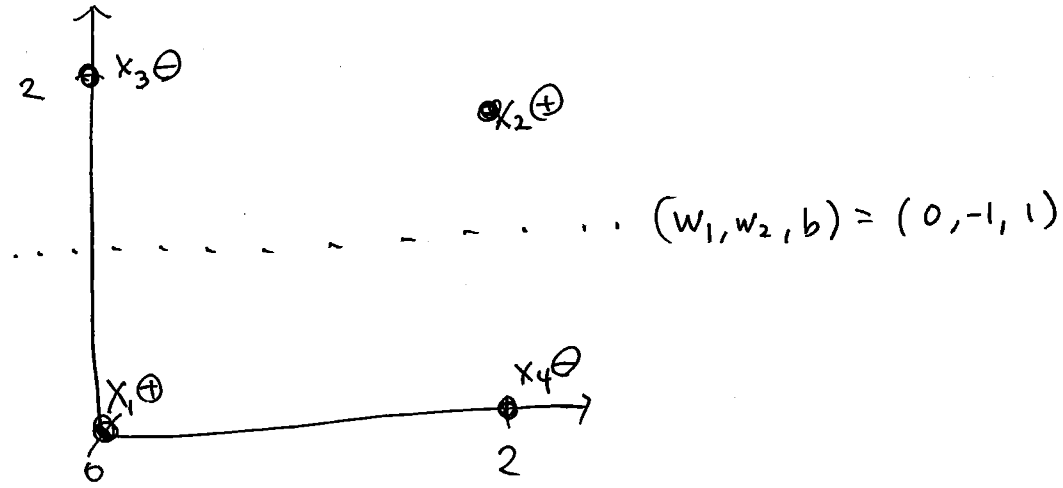
w from pg. 13 problem.

margin of any separating hyperplane.

$\therefore \frac{1}{|\vec{w}|} \geq M$ for any solution $\vec{z}$.

This completes the proof that the hyperplane $\vec{w}\cdot\vec{x}+b=0$ (given from the optimization problem on page. 13) provides the maximum margin.

problem 3.



a) For the data points to have a hard margin, they need to be linearly separable.
For them to be linearly separable, the convex hull of the $\oplus$ class and the convex hull of the $\ominus$ class must not intersect, but they do. Thus $x_1 \sim x_4$ do not have a hard margin.

b) pick the hyperplane with $(w_1, w_2, b) = (0, -1, 1)$.

$\varepsilon^{(i)} = \max(0, 1 - y^{(i)}(\vec{w} \cdot \vec{x}^{(i)} + b))$

$\varepsilon^1 = \max(0, 1 - (+1)([0,-1]\cdot[0,0] + 1)) = 0$

$\varepsilon^2 = \max(0, 1 - (+1)([0,-1]\cdot[2,2] + 1)) = 2$

$\varepsilon^3 = \max(0, 1 - (-1)([0,-1]\cdot[0,2] + 1)) = 0$

$\varepsilon^4 = \max(0, 1 - (-1)([0,-1]\cdot[2,0] + 1)) = 2.$

$\therefore (w_1, w_2, b, \varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4)$

$= (0, -1, 1, 0, 2, 0, 2)$

problem 4.

a) $K(x,z) = \#$ of unique words in both x & z.

$K(x,z)$ is a kernel if $\exists\ \phi(x)$ s.t. $K(x,z) = \phi(x)\cdot\phi(z)$.

define $\phi(x) :=$ feature vector of x.

~~Then~~ If $\phi(x)_j = \phi(z)_j = 1$, $\phi(x)_j \cdot \phi(z)_j = 1$

else $\phi(x)_j\, \phi(z)_j = 0$.

Thus $\phi(x)\cdot\phi(z)$ yields # of unique words in both x & z.

b) $K(\vec{x},\vec{z}) = (1+\beta\vec{x}\cdot\vec{z})^2 - 1$

$= \cancel{1} + \beta^2(\vec{x}\cdot\vec{z})^2 + 2\beta\vec{x}\cdot\vec{z}\ \cancel{-1}$

$= \beta^2(x_1z_1 + x_2z_2)^2 + 2\beta(x_1z_1 + x_2z_2)$

$= \beta^2(x_1^2z_1^2 + 2x_1x_2z_1z_2 + x_2^2z_2^2) + 2\beta(x_1z_1 + x_2z_2)$

$$= \begin{pmatrix} \beta x_1^2 \\ \beta x_2^2 \\ \sqrt{2}\beta x_1x_2 \\ \sqrt{2\beta}\, x_1 \\ \sqrt{2\beta}\, x_2 \end{pmatrix} \cdot \begin{pmatrix} \beta z_1^2 \\ \beta z_2^2 \\ \sqrt{2}\beta z_1z_2 \\ \sqrt{2\beta}\, z_1 \\ \sqrt{2\beta}\, z_2 \end{pmatrix}$$

$= \phi(x)\cdot\phi(z)$.

$$\phi(x) = \begin{pmatrix} \beta x_1^2 \\ \beta x_2^2 \\ \sqrt{2}\beta x_1x_2 \\ \sqrt{2\beta}\, x_1 \\ \sqrt{2\beta}\, x_2 \end{pmatrix}$$

∎

problem 5.

In the binary prediction rule,

$\hat{y} = \text{sign}(\vec{w} \cdot \vec{x} + b) > 0 \longrightarrow$ class $k=1$ $(+)$

$\hat{y} = \text{sign}(\vec{w} \cdot \vec{x} + b) < 0 \longrightarrow$ class $k=2$ $(-)$.

In the multi-class SVM decision rule, $(K=2)$

$$\hat{y} = \underset{k}{\text{argmax}}\ \vec{w}_k \cdot \vec{x} + b_k$$

$\vec{w}_1 \cdot \vec{x} + b_1 > \vec{w}_2 \cdot \vec{x} + b_2 \longrightarrow$ class $k=1$

$\Leftrightarrow (\vec{w}_1 - \vec{w}_2) \cdot \vec{x} + (b_1 - b_2) > 0$

$(\vec{w}_1 - \vec{w}_2) \cdot \vec{x} + (b_1 - b_2) < 0 \longrightarrow$ class $k=2$

The binary prediction rule and multi-class SVM decision rule are equivalent if

$$\vec{w} = \vec{w}_1 - \vec{w}_2 , \quad b = b_1 - b_2 .$$

Problem 6.

For optimal values $C = 1.5$, $\gamma = 0.005$,

test error = 5.0%

cross-validation error (avg.) = 5.9%