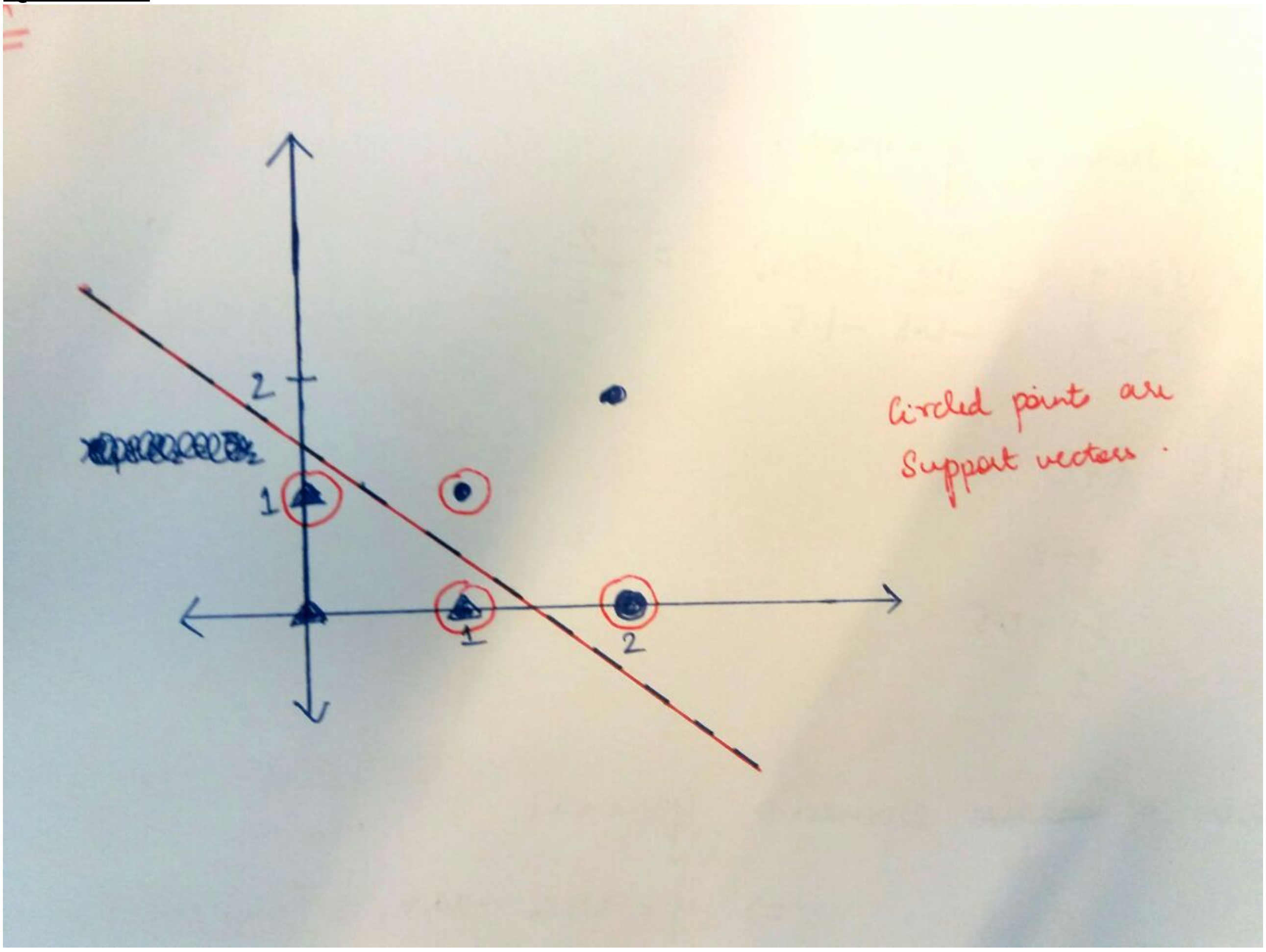
Shikhar Sakhuja : ss9131 : Assignment 3 : Machine Learning

Question1:



a. Yes

b. The maximum margin hyperplane in this case should have a slope of -1 and should satisfy x_1 = 3/2, x_2 = 0. Therefore its equation is $x_1 + x_2 = 3/2$ and weight vector is $(1,1)^T$.

c. In the given data points the optimal margin increases if we remove support vectors (1,0) or (1,1) and stays the same when we remove the other two.

Shikhar Sakhuja : ss9131 : Assignment 3 : Machine Learning

Question2.

a.

Problem 2

(a) Given w, b is optimal.

Geometric Margin $\Rightarrow$ $\delta^i = \dfrac{y^i(w \cdot x^i + b)}{|w|}$

Since w, b are optimal.

we prove $\delta^i = \dfrac{y^i(w \cdot x^i + b)}{|w|} = \dfrac{1}{|w|}$ is optimal.

Our original aim was to maximize δ.

Suppose there exists w^* & b^* that provide a ~~more optimal~~ better solution by maximizing the margin even more

$$\frac{y^i(w^* \cdot x^i + b^*)}{|w^*|} = \frac{y^i(w \cdot x^i + b)}{|w|} + \varepsilon \quad (\varepsilon > 0)$$

for a point on the margin $y^i(w \cdot x^i + b) = 1$

$$\frac{y^i(w^* \cdot x^i + b^*)}{\|w^*\|} = \frac{1}{\|w\|} + \varepsilon.$$

$$\therefore \quad \frac{y^i(w^* \cdot x^i + b^*)}{\|w^*\|} > \frac{1}{\|w\|}$$

However if this is true, w^*, b^* is ~~more~~ optimal as it provides a bigger margin.

But it is given that w, b give us optimal margins.

$\therefore$ w^*, b^* can't exist as that would negate the given claim that w, b are optimal

Thus by proof by contradiction, w, b are indeed optimal.

b.

Problem 2 (b)

$$z' = \frac{z}{\|z\| M} \qquad d' = \frac{d}{\|z\| M} \quad -(1)$$

Given that z, d are any other seperable hyperplane.

So, $y^i (z' \cdot x^i + d')$ $-(2)$.

$= y^i$ ~~since z'~~

Using (1).

$$= y^i \left(\frac{z \cdot x^i}{\|z\| M} - \frac{d}{\|z\| M} \right).$$

$$= y^i \left(\frac{z \cdot x^i - d}{\|z\| M} \right).$$

Given M is margin for dataset then

for each $M^{(i)} = \frac{y^i (x^i \cdot z + d)}{\|z\|} \geq M.$

$$\Rightarrow \frac{y^i (x^i \cdot z + d)}{\|z\|} \geq 1 \quad -(3)$$

from (2) & (3)

$$y^i (z' \cdot x^i + d) = \frac{y^i (z \cdot x^i + d)}{\|z\| M} \geq 1 \quad \text{for } i = 1, 2, 3 \dots n$$

Since w is the optimal solution for hard margin, $\|w\|^2 \leq \|z\|^2$ as it is the minimum solution.

$\therefore \|w\| \leq \|z'\|$

c.

c) Given $z' = \frac{z}{|z| M}$

Magnitude of z'

$$||z'|| = \left|\frac{z}{||z||}\right| \cdot \frac{1}{M}$$

Since $\frac{z}{|z|}$ is a unit vector

$$||z'|| = \frac{1}{M} \Rightarrow M = \frac{1}{||z'||} \quad - (1)$$

from part (b) we have $||w|| \le ||z'||$ — (2)

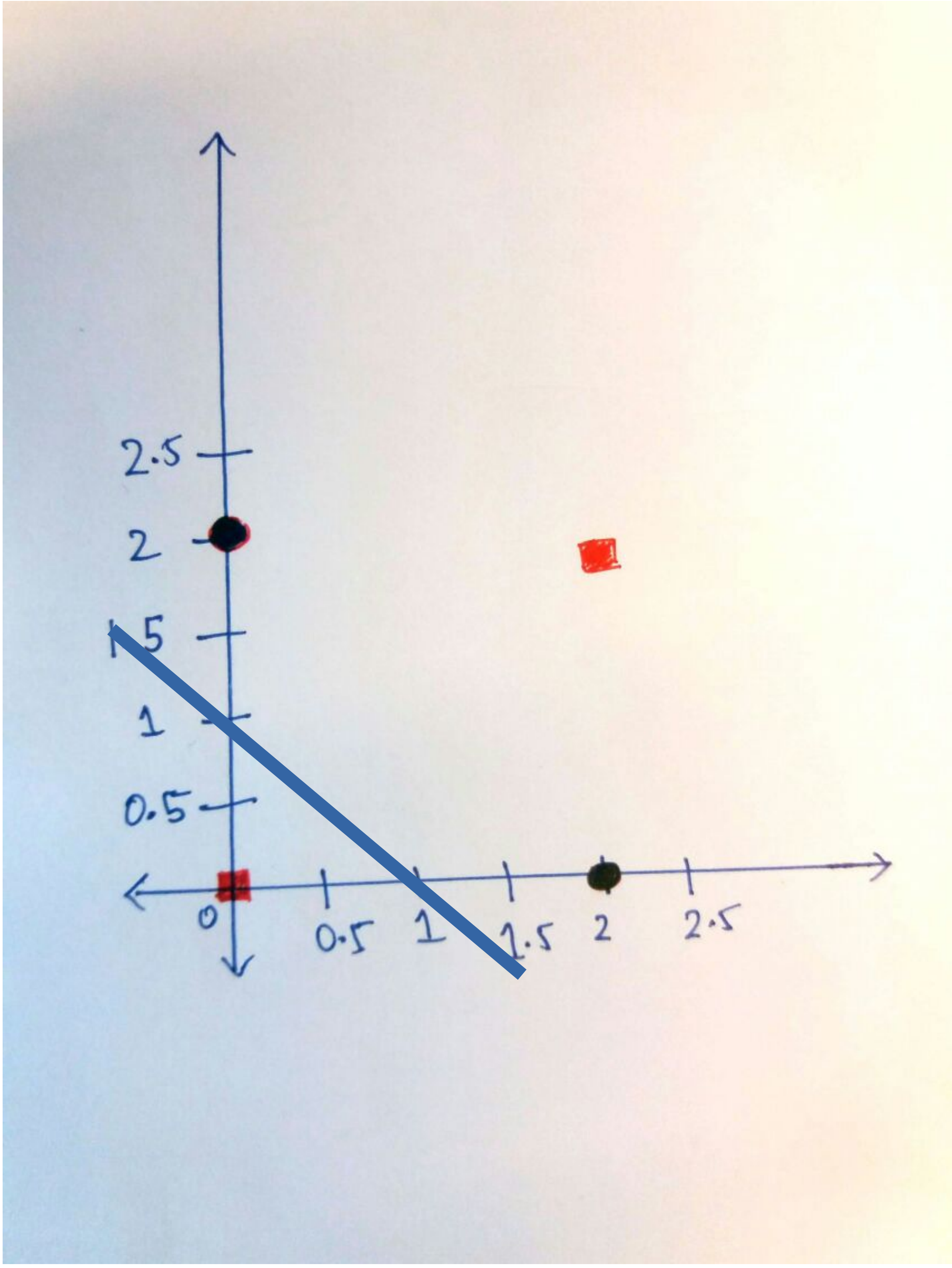
from (1) & (2),

$$\frac{1}{||w||} \geq \frac{1}{||z||}$$

from (1)

$$\boxed{\frac{1}{||w||} \geq M}$$

Question3:



a. Constraint for hard margin SVM: $Y^{(i)} * (W.X^{(i)} + b) >= 1$, i = 1,2,3,…,m
According to the constraint should have a margin of at least 1 unit from the optimal hyperplane. However, since the given dataset is not linearly separable and therefore there doesn't exist a hyperplane that satisfies the constraint for all data points.

b.

Q3 b)

$w_0(0) + w_1(0) + b = 1 \Rightarrow \boxed{b = 1}$

$-(w_0(0) + w_1(2) + b) = 1 \Rightarrow -2w_1 - b = 1 \Rightarrow 2w_1 = -2$; $\boxed{w_1 = -1}$

$-(w_0(2) + w_1(0) + b) = 1 \Rightarrow -2w_0 - 1 = 1 \Rightarrow \boxed{w_0 = -1}$

Since pt (2,2) will be classified wrong because of the hyperplane we introduce slack.

$$\varepsilon^{(2)} = 1 - (w_0(2) + w_1(2) + b)$$
$$= 1 - 2w_0 - 2w_1 - 1.$$
$$= 1 + 2 + 2 - 1$$
$$= 4.$$

$w_1 = -1$; $w_2 = -1$; $b = 1$; $\varepsilon^{(1)} = \varepsilon^{(3)} = \varepsilon^{(4)} = 0$; $\varepsilon^{(2)} = 4$.

Seperating hyperplane:

$$w_0 x_1 + w_1 x_2 + b = 0.$$
$$\boxed{-x_1 - x_2 + 1 = 0}.$$

Shikhar Sakhuja : ss9131 : Assignment 3 : Machine Learning

Question4:

a. Generally, a function k(x,z) is a function is a valid kernel if it is symmetrical, that is k(x,z) = k(z,x). Since k, the function is defined to be the number of unique words between the 2 documents x and z, we can confidently say that the number of unique words in k(x,z) will be equal to k(z,x). Therefore, we can call the function decomposable and hence symmetrical.

b.

$x = [x_1, x_2]$ $\quad z = [z_1, z_2]$ $\quad \beta > 0$.

$n = 2$.

$$k(x,z) = (1 + \beta x \cdot z)^2 - 1 .$$

$$= 1 + \beta^2 (x \cdot z)^2 + 2\beta (x \cdot z) - 1$$

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

$$= \beta^2 \left((x_1 z_1)^2 + (x_2 z_2)^2 + 2 x_1 x_2 z_1 z_2 \right) + 2\beta (x_1 z_1 + x_2 z_2)$$

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

$$\phi(x) = \begin{bmatrix} \beta x_1^2 \\ \beta x_2^2 \\ \sqrt{2\beta}\, x_1 \\ \sqrt{2\beta}\, x_2 \\ \sqrt{2}\,\beta\, x_1 x_2 \end{bmatrix} \qquad \phi(z) = \begin{bmatrix} \beta z_1^2 \\ \beta z_2^2 \\ \sqrt{2\beta}\, z_1 \\ \sqrt{2\beta}\, z_2 \\ \sqrt{2}\,\beta\, z_1 z_2 \end{bmatrix}$$

Shikhar Sakhuja : ss9131 : Assignment 3 : Machine Learning

Question5:

for $k=2$.

Class 1

$$w_+ \cdot x^{(i)} + b_+ \geq w_- \cdot x^{(i)} + b_- \quad \text{for } y^{(i)} = +.$$

Class 2

$$w_- \cdot x^i + b_- \geq w_+ \cdot x^i + b_+ \quad \text{for } y^{(i)} = -$$

let $w = (w_+) - (w_-)$

$b = (b_+) - (b_-)$

$\therefore$ $\text{sign}(w \cdot x^i + b)$ is equivalent to

$$\hat{y} = \arg\max_{k=(+,-)} w_k \cdot x + b_k .$$

for class $= (+)$

$$w \cdot x + b = (w_+ - w_-) \cdot x + (b_+ - b_-) \geq 0 .$$

$$= w_+ \cdot x + b_+ \geq w_- \cdot x + b_-$$

for class $= (-)$

$$w \cdot x + b = (w_+ - w_-) \cdot x + (b_+ - b_-) \leq 0 .$$

$$= w_+ \cdot x + b_+ \leq w_- \cdot x + b_-$$

$\therefore$ for $k=2$ with $k = \{+, -\}$

$$w = w_+ - w_-$$

$$b = b_+ - b_- .$$

Shikhar Sakhuja : ss9131 : Assignment 3 : Machine Learning

Question6:
Python file included in the folder.

b. Test Error o default settings : 7.33%
c. With C = 3 and gamma = 1/150 →
- Test Error : 4.9% ;
- Cross Validation Error : 4.83 % ;
- Cross Validation Score : [0.95702479 0.94029851 0.96333333 0.95477387 0.94285714]

Question 7:

Problem 7

$$\max_{\alpha\geq 0} \min_{w,b} \frac{1}{2}\|w\|^2 - \sum_i \alpha_i \left[(w\cdot x^{(i)} + b) y^{(i)} - 1 \right] \quad -\ (1)$$

$$L(w,\alpha) = \frac{1}{2} w\cdot w - \sum_i \alpha_i \left[(w\cdot x^{(i)} + b) y^{i} - 1 \right] \quad -(1)\ ;\ \alpha_i \geq 0 \text{ for all } i$$

$$\frac{\partial L}{\partial w} = \frac{1}{2}\times 2w - \sum \alpha_i x^{(i)} y^{(i)} = 0$$

$$w = \sum_i \alpha_i x^{(i)} y^{(i)} \quad -(2)$$

$$\frac{\partial L}{\partial B} = -\sum_i \alpha_i y^{(i)} = 0$$

$$\Rightarrow \sum \alpha_i y^{(i)} = 0. \quad -(3)$$

Plugging (2) & (3) into (1)

$$L(w,\alpha) = \frac{1}{2}\left(\sum_i \alpha_i \bar{x}^{(i)} y^{(i)}\right) \cdot \left(\sum_j \alpha_j \bar{x}^{(j)} y^{(j)}\right)$$
$$- \left(\sum_i \alpha_i y^{(i)} \bar{x}^{(i)}\right) \cdot \left(\sum_j \alpha_j y^{(j)} \bar{x}^{(j)}\right) - \sum \alpha_i y^{(i)} b + \sum \alpha_j$$

(Used i & j to differentiate b/w similar but different data points & expanded the multiplication with lagrange multiplier)

Since $\sum \alpha_i y^{(i)} = 0$

$$L = -\frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y^{(i)} y^{(j)} (\vec{x}^{(i)} \cdot \vec{x}^{(j)}) + \sum_{i=1}^{m} \alpha_i$$

$$= \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j y^{(i)} y^{(j)} (\vec{x}^{(i)} \cdot \vec{x}^{(j)}).$$

$$\therefore L(w,b,\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j y^{(i)} y^{(j)} (\vec{x}^{(i)} \cdot \vec{x}^{(j)})$$

Since we reached this eq^n by minimizing L w.r.t w & b.. Putting it together with $\alpha_i \geq 0$ & $\sum_{i=1}^{m} \alpha_i y^{(i)} = 0$ constraints we obtain

$$\max_\alpha W(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} y^{(i)} y^{(j)} \alpha_i \alpha_j \langle x^{(i)}, x^{(j)} \rangle$$

$$\text{st. } \alpha_i \geq 0 \, , \, i = 1, \ldots n$$

$$\sum_{i=1}^{m} \alpha_i y^{(i)} = 0 \, .$$