CSCI 467 Problem Set 6

Teresa Liu

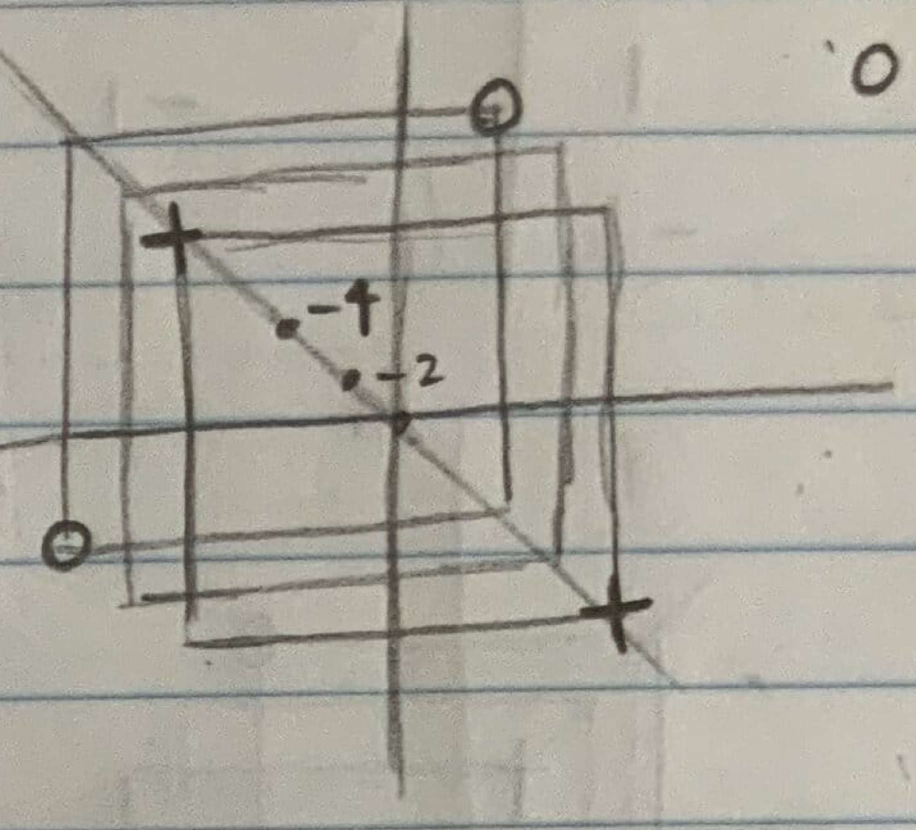
1) a) XOR Gate : $(A+B)\cdot(\bar{A}+\bar{B})$

$z = f(x_1, x_2) = (x_1 + x_2)(-x_1 - x_2)$

y	x_1	x_2	z
-1	-1	-1	$-2\cdot2=-4$
1	-1	1	$0\cdot0=0$
1	1	-1	$0\cdot0=0$
-1	1	1	$-2\cdot2=-4$

$+ \Rightarrow y=1$

$0 \Rightarrow y=-1$

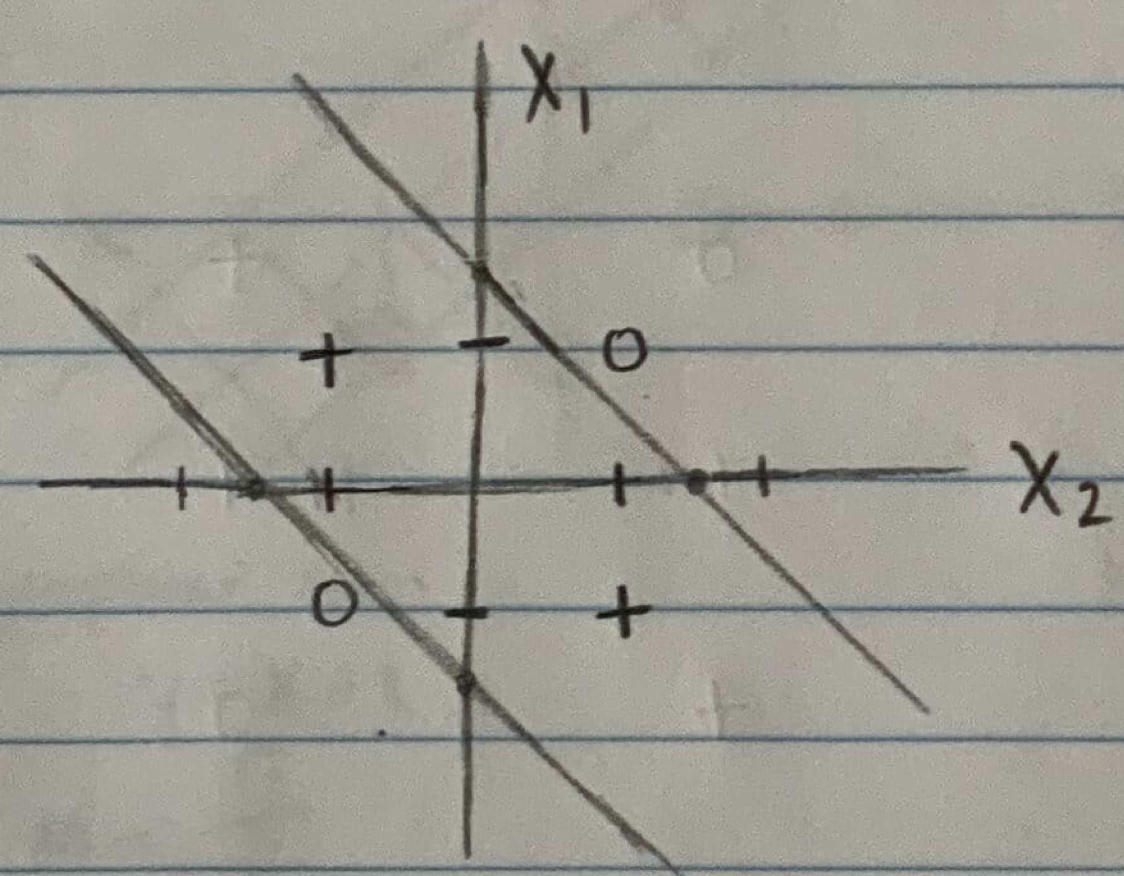


if $z+2>0$ then $y=1$

$z+2<0$ then $y=-1$

where $z=(x_1+x_2)(-x_1-x_2)$

hyperplane at $z=-2$



$f(x) = (x_1+x_2)(-x_1-x_2)+2$

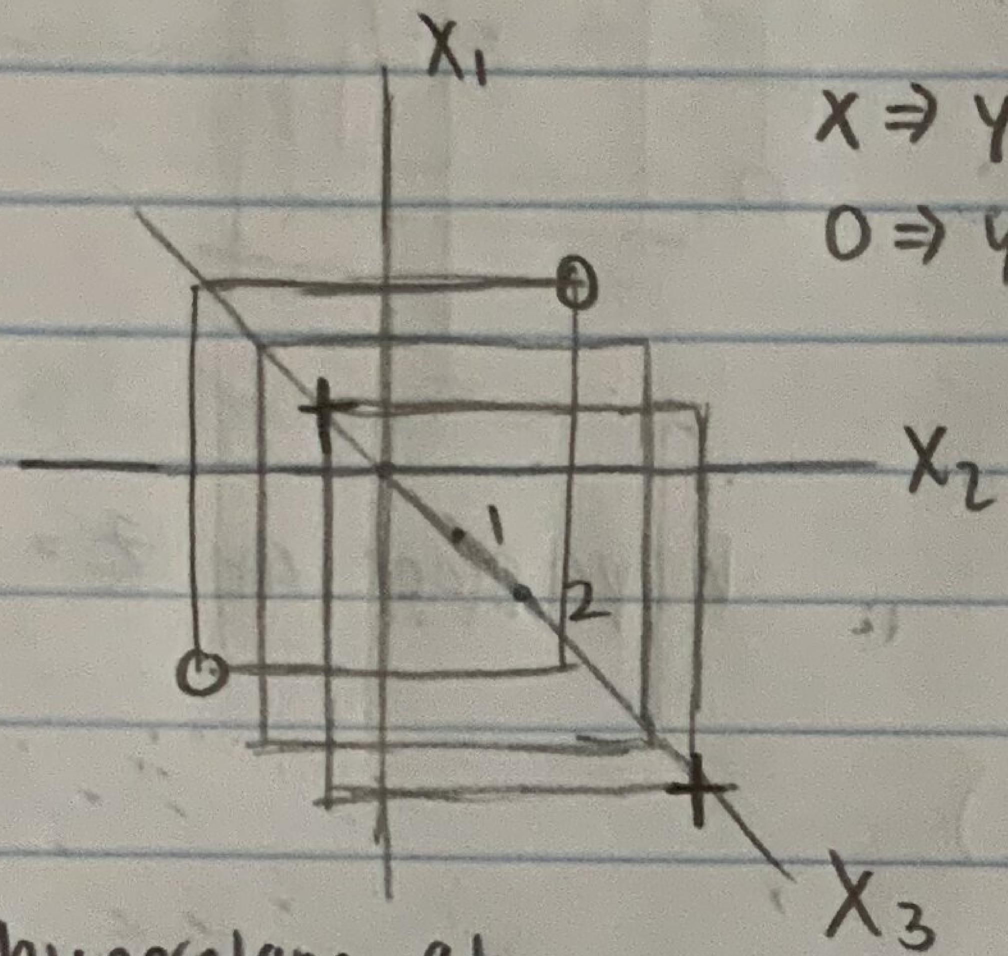
$0 = -x_1^2 - 2x_1x_2 - x_2^2 + 2$

decision boundary

b)

Y	X_1	X_2	X_3
-1	-1	-1	0
1	-1	1	2
1	1	-1	2
-1	1	1	0

$X_3 = |X_1 - X_2|$



$x \Rightarrow y = 1$

$0 \Rightarrow y = -1$

hyperplane at $X_3 = 1$

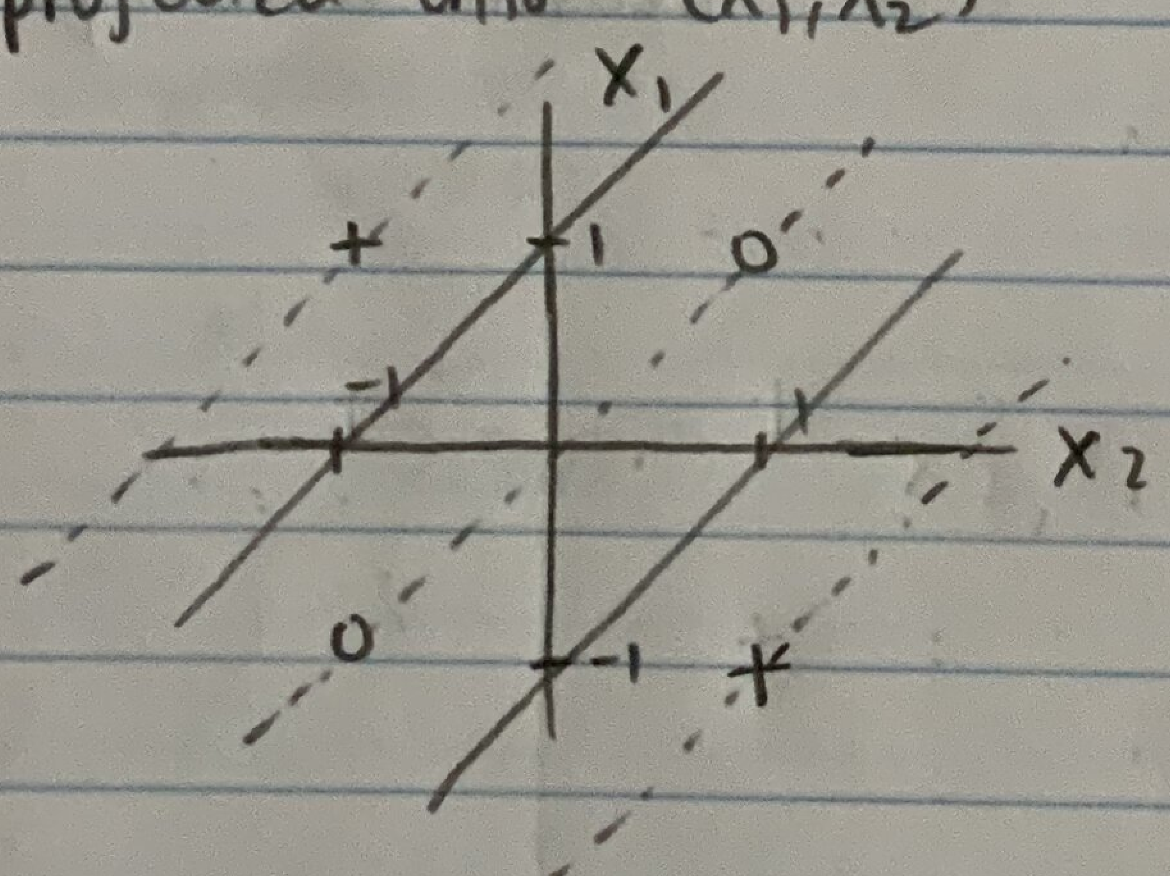
decision boundary at

$f(x) = X_3 - 1 = |X_1 - X_2| - 1$

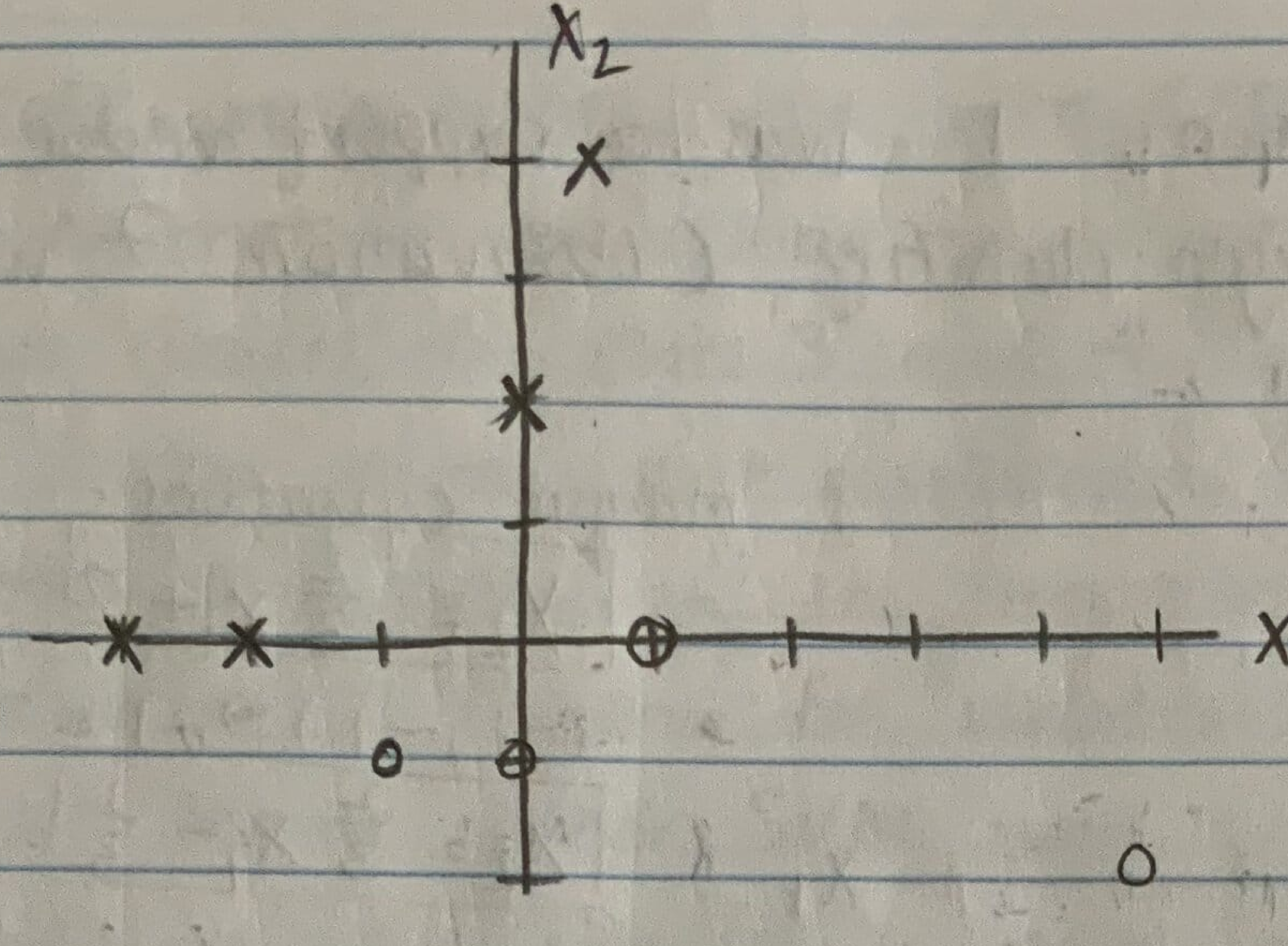
if $X_3 - 1 > 0$ then $y = 1$

if $X_3 - 1 < 0$ then $y = -1$

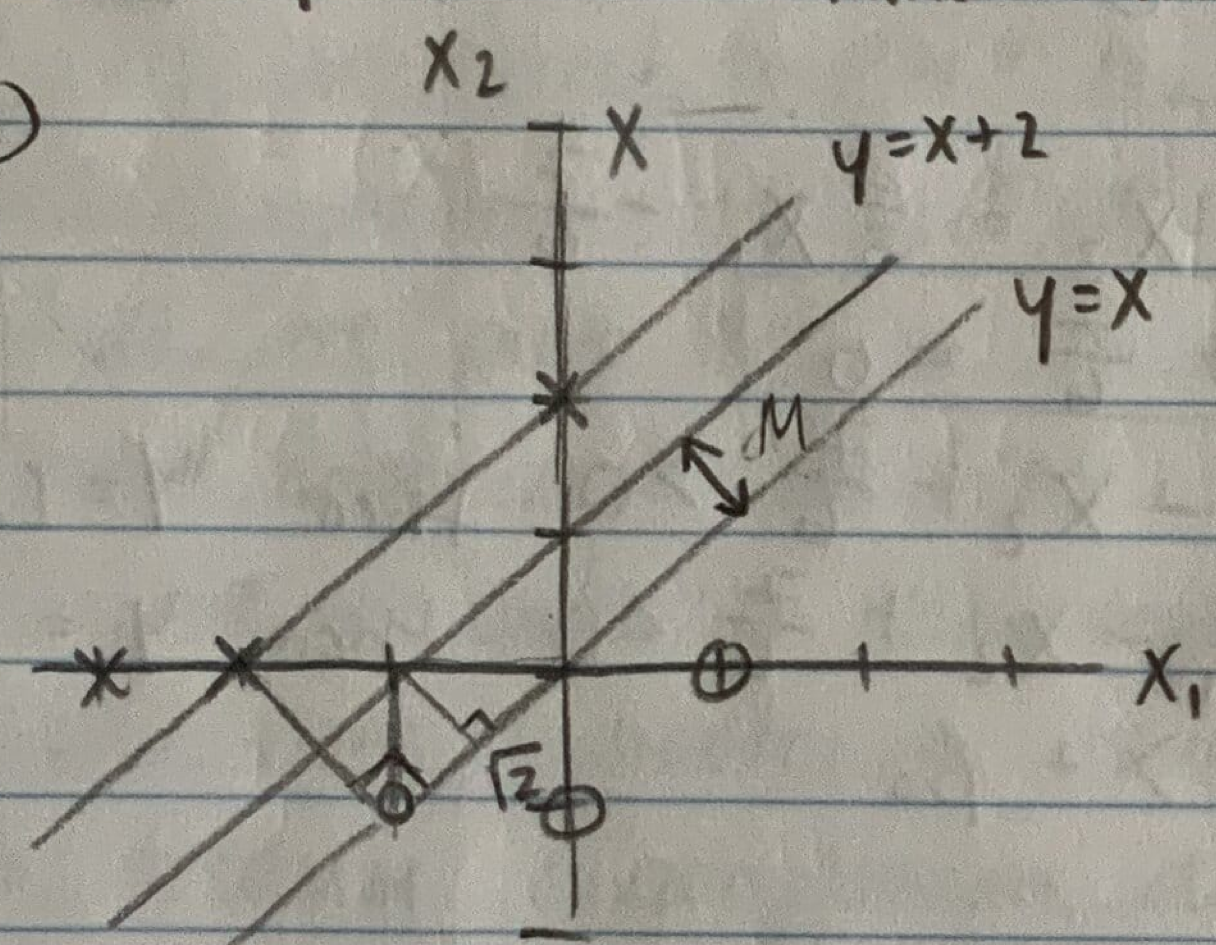
projected onto (X_1, X_2)

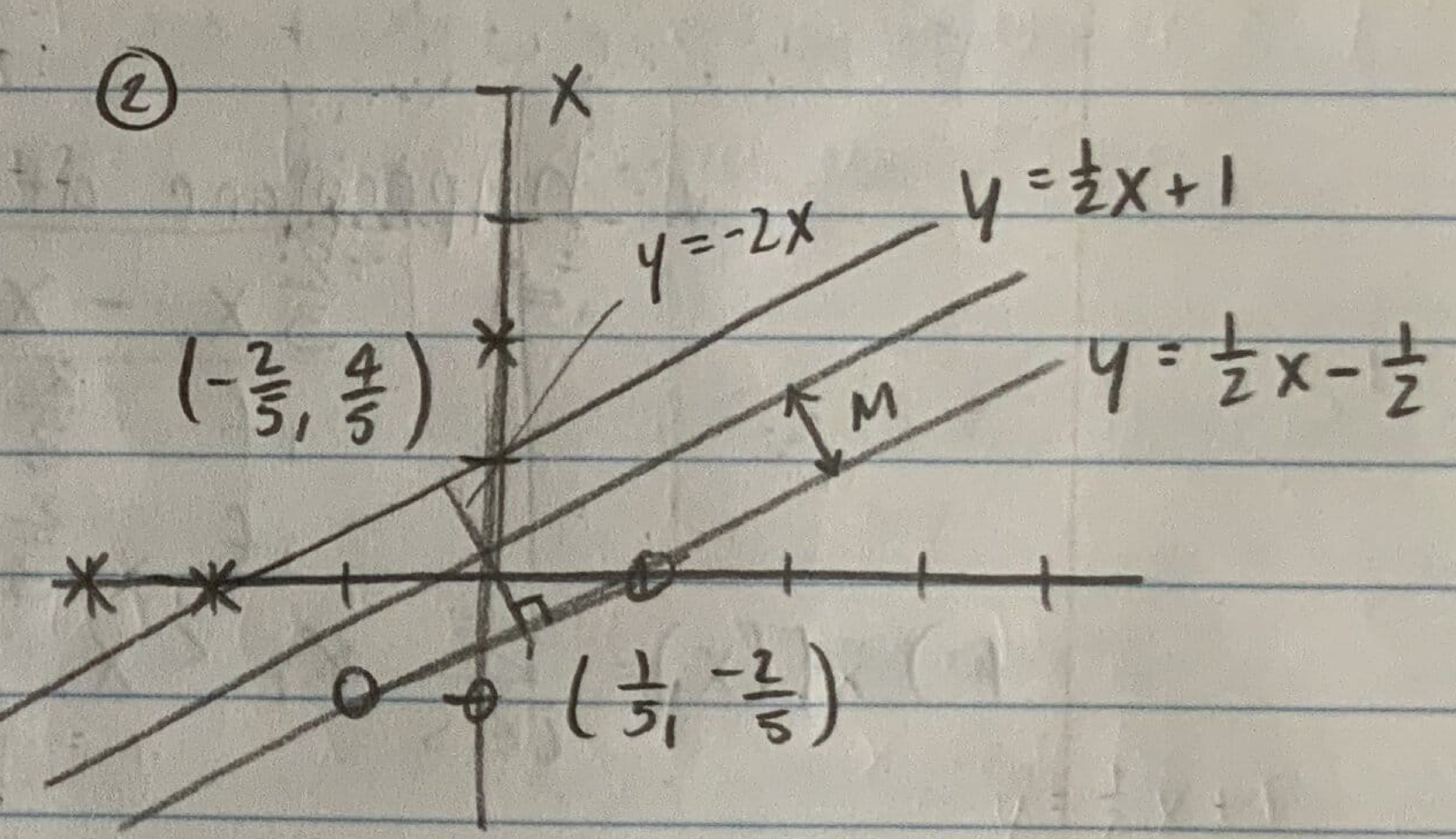


2) a)



b) two options of maximal margin classifiers

① 

② 

hypaplane at
$y = x+1$
$M = \frac{\sqrt{2}}{2}$
$= 0.707$

hyperplane at
$y = \frac{1}{2}x + \frac{1}{4}$

calculating margin M :

$y = \frac{1}{2}x+1$
$-2x = \frac{1}{2}x+1$
$-\frac{5}{2}x = 1$
$x = -\frac{2}{5}$
$y = \frac{4}{5}$

$y = \frac{1}{2}x - \frac{1}{2}$
$-2x = \frac{1}{2}x - \frac{1}{2}$
$\frac{2}{5} \cdot -\frac{5}{2}x = \frac{-1}{2} \cdot \frac{2}{5}$
$x = \frac{1}{5}$
$y = \frac{-2}{5}$

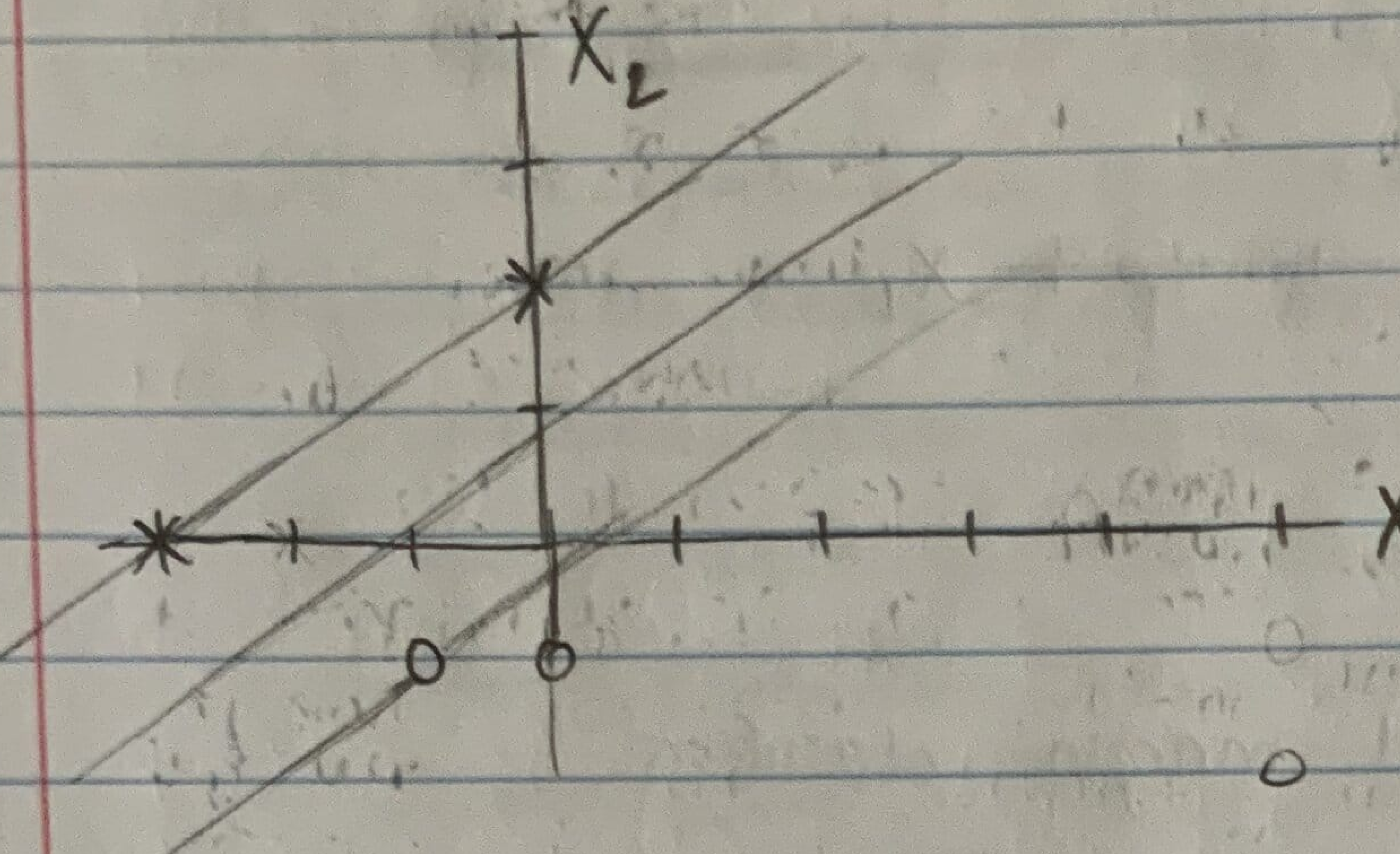
$M = \frac{1}{2}\sqrt{(\frac{1}{5}+\frac{2}{5})^2 + (-\frac{2}{5}-\frac{4}{5})^2} = 0.6708$

c)
Since margin at ① > margin at ②,
we choose hyperplane at $y = x+1$ where $y = X_2$ and $x = X_1$
$\Rightarrow \boxed{X_1 - X_2 + 1 = 0}$

d) if $X_1 - X_2 + 1 > 0$ then Y = cross
if $X_1 - X_2 + 1 < 0$ then Y = circle

margin equations: $\boxed{\begin{array}{l} X_1 - X_2 + 2 = 0 \\ X_1 - X_2 = 0 \end{array}}$

e) Yes, since observation 1 was a support vector for our maximal margin classifier (observation 7 was not)



margin equations:

$\boxed{X_2 = \frac{2}{3}X_1 + 2}$

$(-1,-1) \rightarrow -1 = -\frac{2}{3} + b \quad b = -\frac{1}{3}$

$\boxed{X_2 = \frac{2}{3}X_1 - \frac{1}{3}}$

$\frac{2 + \frac{1}{3}}{2} = \frac{7}{6}$

hyperplane at $X_2 = \frac{2}{3}X_1 + \frac{7}{6}$

$\Rightarrow \boxed{\frac{2}{3}X_1 - X_2 + \frac{7}{6} = 0}$

if $\frac{2}{3}X_1 - X_2 + \frac{7}{6} > 0$ then Y = cross

if $\frac{2}{3}X_1 - X_2 + \frac{7}{6} < 0$ then Y = circle

f) $f(X) = \sum_{i \in S} \alpha_i \langle X, X_i \rangle + \beta_0$

$S = \{(-3,0), (0,2), (-1,-1)\}$

$f(x) = \alpha_1 (X_1, X_2)(-3,0)^T + \alpha_2 (X_1, X_2)(0,2)^T$
$+ \alpha_3 (X_1, X_2)(-1,-1)^T + \beta_0$
$= \alpha_1(-3X_1) + \alpha_2(2X_2) + \alpha_3(-X_1 - X_2) + \beta_0$
$= 3X_1\alpha_1 + 2X_2\alpha_2 - X_1\alpha_3 - X_2\alpha_3 + \beta_0 = \frac{2}{3}X_1 - X_2 + \frac{7}{6}$
$X_1(3\alpha_1 - \alpha_3) + X_2(2\alpha_2 - \alpha_3) + \beta_0 = \frac{2}{3}X_1 - X_2 + \frac{7}{6}$

$3\alpha_1 - \alpha_3 = \frac{2}{3}$

$2\alpha_2 - \alpha_3 = -1$

$\beta = \frac{7}{6}$

$$\boxed{\begin{aligned} \alpha_1 &= 1 \\ \alpha_2 &= \frac{3\alpha_1}{2} - \frac{5}{6} \\ \alpha_3 &= 3\alpha_1 - \frac{2}{3} \end{aligned}}$$

3) a) soft margin linear SVM with $C=0.02 \Rightarrow \boxed{\text{plot 3}}$
Given that we are looking for a linear SVM, we narrow down our choices to plots 3 and 4. Since the tuning parameter C is low in this part compared to part b, we need to choose the plot with the thinner margins, since a lower C indicates less tolerance for violated margins. Hence we chose #3.

b) soft margin linear SVM with $C=20 \Rightarrow \boxed{\text{plot 4}}$
Again, we are looking for a linear SVM, so we narrow down to plots 3 and 4. Since we now have a high tuning parameter C as compared to part a, we choose the plot with the wider margins that indicate more tolerance for violations. hence we pick #4.

c) hard margin kernel SVM, $K(x_i\, x_j) = x_i^T x_j + (x_i^T x_j)^2 \Rightarrow \boxed{\text{plot 5}}$
Since we are given a polynomial kernel, we will look for a plot that indicates a transformation to the data points to produce a non-linear decision boundary. Since the kernel is quadratic, the 3D feature space projected back onto the original plot will most likely look like plot 5, as the curves of the quadratic hyperplane will intersect at two places.

d) hard margin kernel SVM with $K(x_i x_j) = \exp(-5\|x_i - x_j\|^2)$
$\Rightarrow \boxed{\text{plot 6}}$. Since the kernel is in the form of a radial kernel, we must be looking for a circular decision boundary, so plots 1 and 6 match this criteria. Next, the γ value is 5, which is higher than the γ in part e, and since γ is the tuning parameter, and a larger γ indicates a more fluctuating and "wiggly" decision boundary, we choose plot 6.

e) hard margin kernel SVM with $K(x_i x_j) = \exp(-\frac{1}{5}\|x_i - x_j\|^2)$
$\Rightarrow \boxed{\text{plot 1}}$ Since now $\gamma = \frac{1}{5}$, which is lower than γ in part d, we look for a tighter, less fluctuating decision boundary, found in plot 1 with the radial kernel.

4) The parameter γ defines how much a single training observation influences the resulting decision boundary, indicating that the inverse of the radius of influence of each support vector. If γ is too large, then that radius will be very small and leads to overfitting. If γ is small, the model may have a large bias and fail to fit the true shape of the data.
We use γ as a tuning parameter and the optimal γ that minimizes the test error indicates how fine grained the data are.

SVM score

5) $f(x) = \sum_{i=1}^{N} \alpha_i k(x, x_i) + \beta_0$, $k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$

training data $\{(x_1, y_1), \ldots (x_N, y_N)\}$, seperated by distance of at least $\epsilon \Rightarrow \|x_i - x_j\| \geq \epsilon > 0 \ \forall i \neq j$, $y_i \in \{-1, 1\}$

a) reformulate w $\alpha_i = y_i \lambda_i$

$$f(x) = \sum_{i=1}^{N} y_i \lambda_i k(x, x_i) + \beta_0 =$$

$$= \sum_{i=1}^{N} y_i \lambda_i \exp(-\gamma \|x_i - x_j\|^2) + \beta_0$$

b) Let $\lambda_i = 1 \ \forall i$, $\beta_0 = 0$

$y_i \in \{-1, 1\} \Rightarrow x_i$ will be correct if $|f(x_i) - y_i| < 1$

Hint: $|\Sigma_i a_i| \leq \Sigma_i |a_i|$

$$|f(x_j) - y_j| = \left|\sum_{i=1}^{N} y_i \exp(-\gamma \|x_j - x_i\|^2) - y_j\right|$$

if $i = j$, then $y_i \exp(-\gamma \|\cancel{x_i - x_i}\|^2) = y_i = y_j$

$$|f(x_j) - y_j| = \left|\cancel{y_j} + \sum_{i \neq j}^{N} y_i \exp(-\gamma \|x_j - x_i\|^2) - \cancel{y_j}\right|$$

$$|f(x_j) - y_j| = \left|\sum_{i \neq j}^{N} y_i \exp(-\gamma \|x_j - x_i\|^2)\right|$$

$$\leq \sum_{i \neq j}^{N} |y_i \exp(-\gamma \|x_j - x_i\|^2|$$

since $|y_i| = 1$, $e^x \geq 1$

$$= \sum_{i \neq j}^{N} \exp(-\gamma \|x_j - x_i\|^2)$$

$$\leq \sum_{i \neq j} \exp(-\epsilon^2 \gamma) = (N-1) \exp(-\epsilon^2 \gamma)$$

Solve for γ

$$(N-1) \exp(-\epsilon^2 \gamma) < 1$$

$$\exp(-\epsilon^2 \gamma) < \frac{1}{N-1}$$

$$-\epsilon^2 \gamma < \ln\left(\frac{1}{N-1}\right)$$

$$\boxed{\gamma > \frac{-\ln\left(\frac{1}{N-1}\right)}{\epsilon^2}}$$

$\lambda_i = 1$

$\beta_0 = 0$

b) No, because by setting $\lambda_i = 1$ and $\beta_0 = 0$, you are making an assumption about the shape of the data and introducing more bias, which might not necessarily fit the true shape of the data. We can try solving it this way, but as usual, we need to assess the results and calculate the test error of the model to determine if the model is a good fit. However, we cannot just simply look at this problem with $\lambda_i = 1$ and $\beta_0 = 0$, since these parameters help us better fine tune the model to our data a produce better fits.