Hw-4

a)

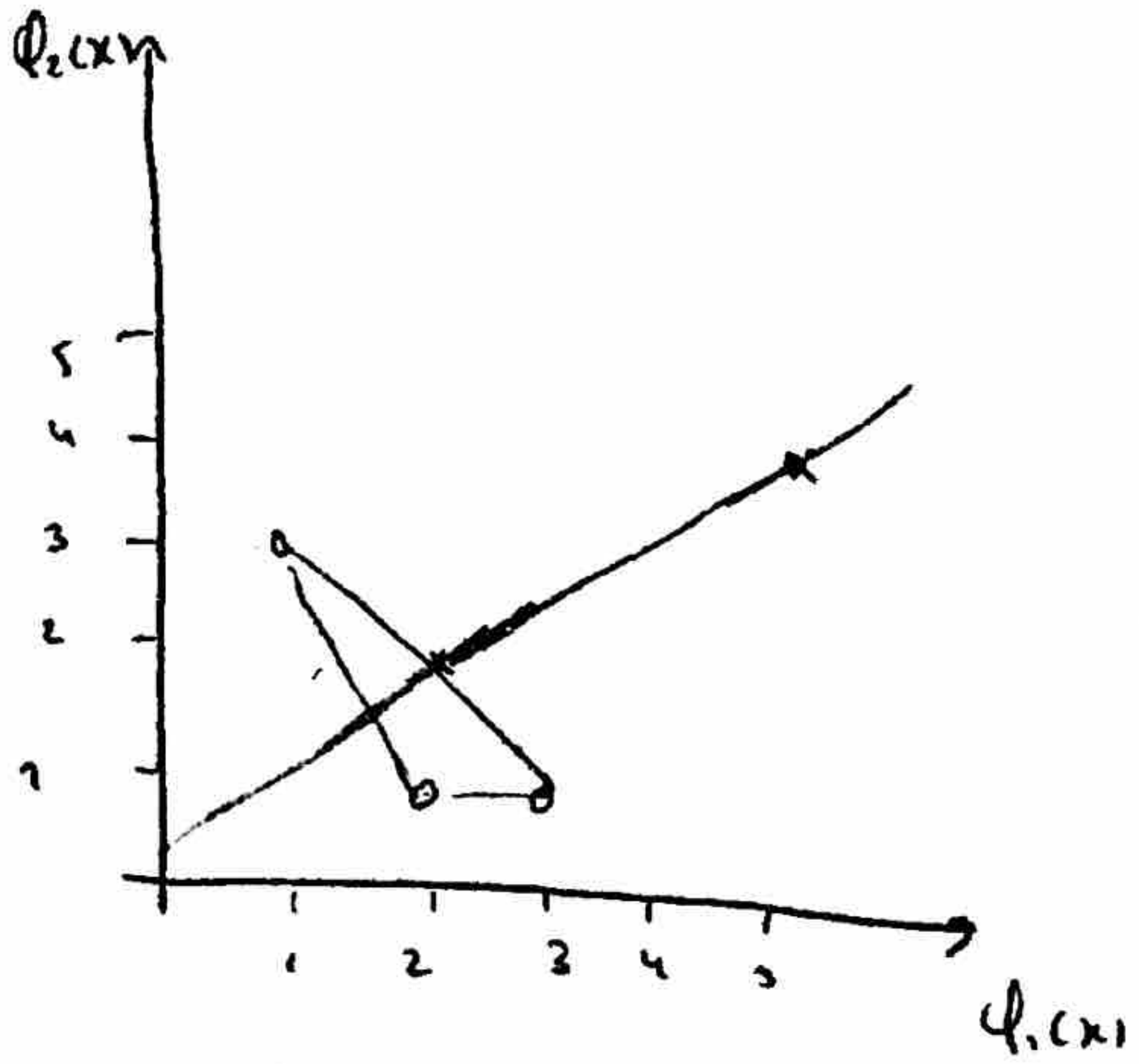
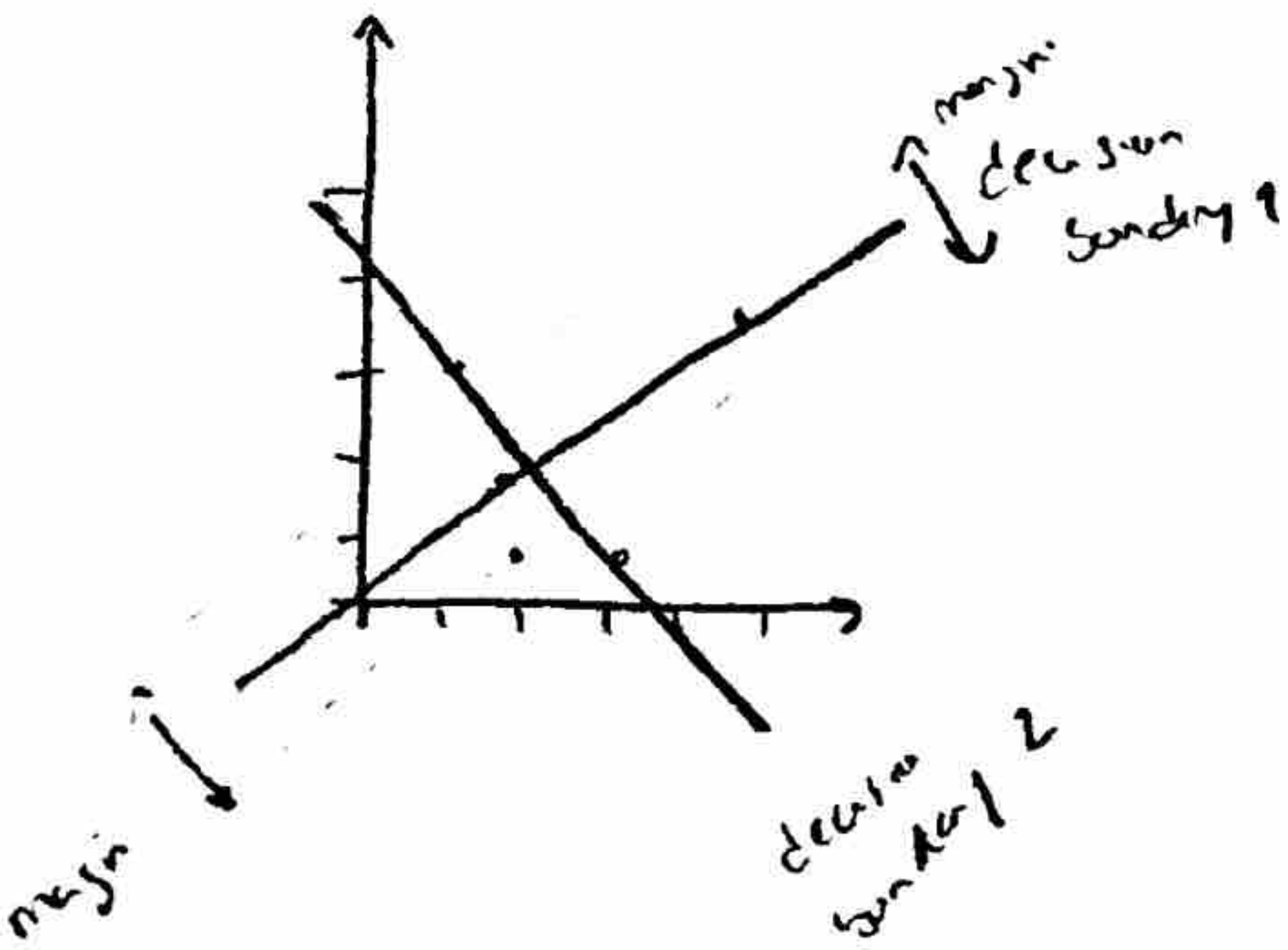


Figure 1

→ There are several ways to tell if a set of data is linearly separable or not. As shown in figure 1, we can simply use a single model (perceptron to check what we want). By just applying the convex hull algorithm to the data and find out if they are overlapping or not.

→ Other way to look at this is by using decision boundary. Take into consideration the following figure.



· As we can tell, there is no margin such that the data set above the decision label is 1 and below otherwise.

· ~~Even if we use~~

→ Since there is no hyperplane that perfectly separates the two classes, the data is not linearly separable.

(b) The aims is to find w and w_0 that minimize the negative log-likelihood function

$$\pi(1|x) = \frac{1}{1+e^{-(w_0 + w_1^T(x))}}$$

In this case, we can compute w_0 and w using the gradient descent, such that

$$w_{new} = w_{old} - \alpha \nabla L(D)$$

$$= w_{old} - \alpha \Delta_w E$$

where α is the gradient descent and $\nabla \alpha(D)$ that tells "the direction" of the deepest ascent.

We can compute the "error" using derivatives for each points:

So we want to minimize the J-function. That being said, we want to maximize the argument of log. Since the derivative of the product is a sum, the derivative of $\pi(a_n|x_n)$ is $\pi(1|n)\cdot a_n$ and the product is over all 5 data set, we still need to normalize. Let's calculate β_0:

$\rightarrow \beta_0 = \frac{\partial \alpha\, \ell(D)}{\partial \beta_0} = \alpha \frac{1}{N} \sum_{n=1}^{5} (\pi(1|n) - a_n)$, a_n is the action

Since the classifier is 0, we get $\frac{1}{1-e^0} = \frac{1}{2}$

Expanding for all points such that $a_n = \{0, 1, 0, 0, 1\}$, we get:

$$= \alpha\left(\frac{1}{5}\right)\left(\frac{1}{2} - 0 + \frac{1}{2} - 1 + \frac{1}{2} - 0 + \frac{1}{2} - 0 + \frac{1}{2} - 1\right) = 0,1 \text{ for } \alpha = 1$$

$$\rightarrow \beta_1 = \frac{\partial \alpha\, \ell(D)}{\partial \beta_1} = \alpha \frac{1}{N} \sum_{n=1}^{5} \underbrace{\Phi_{1,x}}_{x\text{-coordinate}} (\pi(1|n) - a_n) =$$

$$= 1\left(\frac{1}{5}\right)\left(1\left(\frac{1}{2}-0\right) + 2\left(\frac{1}{2}-1\right) + 2\left(\frac{1}{2}-0\right) + 3\left(\frac{1}{2}-0\right) + 5\left(\frac{1}{2}-1\right)\right) =$$

$$= -0,1$$

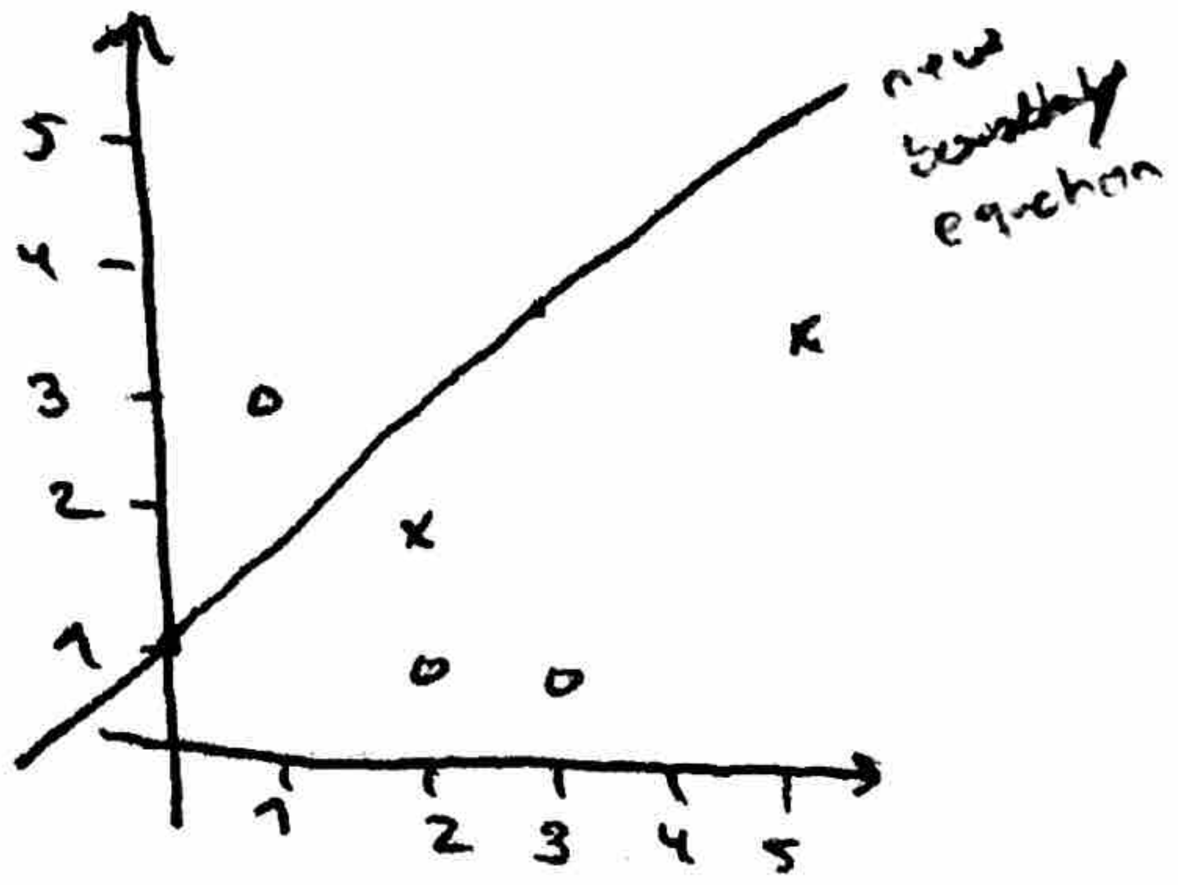
$$\rightarrow \beta_2 = \frac{\partial\, L(D)}{\partial \beta_2} = \alpha \frac{1}{N} \sum_{i=1}^{5} \underbrace{\Phi_{2,x}}_{y\text{-coordinate}} (\pi(1|x) - a_n) =$$

$$= 1\, \frac{1}{5}\left(3\left(\frac{1}{2}-0\right) + 2\left(\frac{1}{2}-1\right) + 1\left(\frac{1}{2}-0\right) + 1\left(\frac{1}{2}-0\right) + 4\left(\frac{1}{2}-1\right)\right) =$$

$$= -0,1$$

(c)

c) the new decision boundary will be $y = w_1\varphi_1 + w_2\varphi_2 + \varphi_3$



y is going to be 0 cause we want to calculate the intersection of the pos, then:

$$0 = \cancel{\beta_0}^{w_1}\varphi_1 + \cancel{\beta_1}^{w}\varphi_2 + \varphi_3 \Rightarrow$$

$$\cancel{\varphi_3} \Rightarrow -w_1\varphi_1 = w_2\varphi_2 + \varphi_3$$

, since $\varphi_1, \varphi_2, \varphi_3$ is going to be $\beta_0, \beta_1, \beta_2$

$$\Leftrightarrow -0{,}1\,w_1 = -0{,}1\,w_2 - 0{,}1$$

$$\Leftrightarrow w_1 = w_2 + 1$$

It doesn't properly classify the points in trans set set the data set follow the equation is from two different classes.