1.

(1) The decision boundary is not unique.

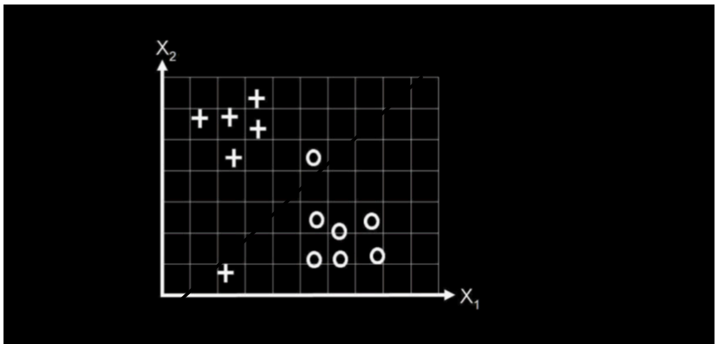


Figure 1: Data for logistic regression question

2 error

(2)

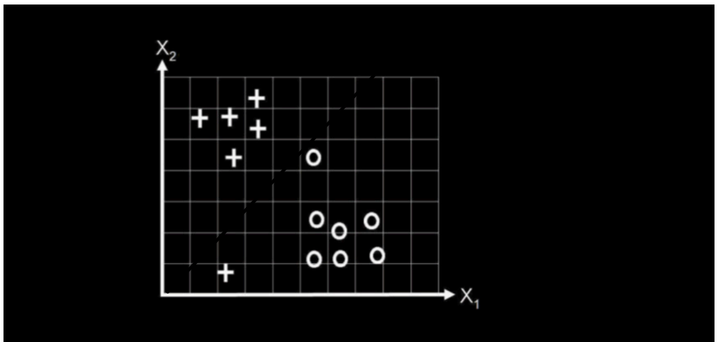


Figure 1: Data for logistic regression question

1 error.

(3)

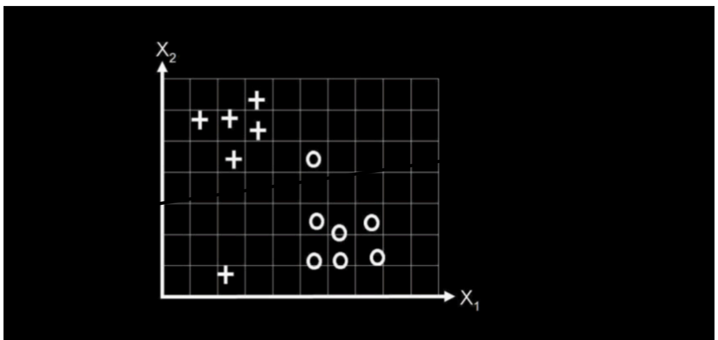


Figure 1: Data for logistic regression question

1 error.

(4)

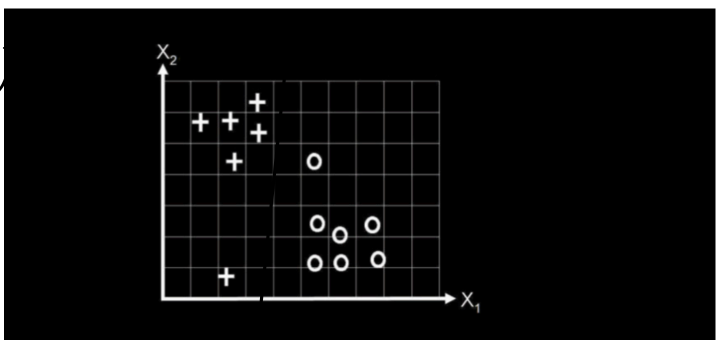


Figure 1: Data for logistic regression question

0 error.

2.

$\varphi(x_1) = [1, -1, 1]$

$\varphi(x_2) = [1, 2, 4]$

① Define Primal problem.

$L(w, w_0, \alpha) = \frac{1}{2}\|w\|_2^2 - \sum_{i=1}^{2} \alpha_i (y_i (w^T \varphi(x_i) + w_0) - 1)$

$w \in \mathbb{R}^3$

② Compute the partial derivatives of L w.r.t. its primal

$\frac{\partial L}{\partial w} = 0 \rightarrow w^* = \alpha_1 y_1 x^1 + \alpha_2 y_2 x^2 = -\alpha_1 \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} + \alpha_2 \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$

$\frac{\partial L}{\partial w_0} = 0 \rightarrow y_1 \alpha_1 + y_2 \alpha_2 = 0$

$\Rightarrow -\alpha_1 + \alpha_2 = 0.$

so $w^* = \alpha_1 \begin{bmatrix} 0 \\ 3 \\ 3 \end{bmatrix}$

Therefore, a vector $\perp w^*$ could be. $\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$

(2) point $\varphi(x_1)$ to the hyperplane

suppose $w = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$

$y_1 w^T \varphi(x_1) + w_0 = -[0, 1, 1] \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} - w_0 = -w_0$

$y_2 w^T \varphi(x_2) + w_0 = [0, 1, 1] \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} + w_0 = 6 + w_0$

$\max_{w_0} \min(|w_0|, |6 + w_0|)$

$\Rightarrow w_0 = -3$

So the margin is $\frac{3}{\|w\|} = \frac{3}{\sqrt{2}} = \frac{3\sqrt{2}}{2}$

(3) $\frac{1}{\|w^*\|} = \frac{3\sqrt{2}}{2}$

$\|w^*\| = \frac{\sqrt{2}}{3}$

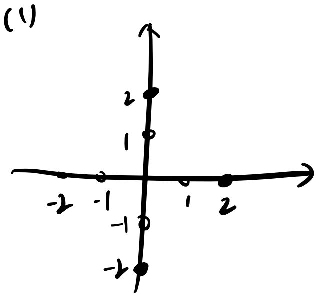
$w^* = \alpha_1 \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$

$\alpha_1 = \frac{1}{3}$

$w^* = \begin{bmatrix} 0 \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}$

3.

(1)



it is not linear separable.

we can increase the dimension.

it might be linear separable in a higher dimension.

(2). RBF kernel $K(x^i, x^j) = \exp\left(-\frac{\|x^i - x^j\|_2^2}{2\sigma^2}\right)$

x^i and x^j are two points in the input space, $\|x^i - x^j\|^2$ is the squared Euclidean distance between the two points. and σ controls the width of the Gaussian.

It allows non-linear separation by creating a boundary that is a smooth curve

Compare to linear SVM, it creates non-linear boundary

can use fewer hyperparameters to tune than the polynomial kernel

In relation to this specific dataset, we found that the data points are not linearly separable, so using RBF kernel is a good choice.

4.

(1) r is the margin

$\therefore r = \frac{1}{\|w\|}$

In the dual problem. $w = \sum_{n=1}^{N} \alpha_n y_n \varphi(x_n)$

$$\|w\|^2 = \sum_{n=1}^{N}\sum_{m=1}^{N} \alpha_n \alpha_m y_n y_m \varphi(x_n)^T \varphi(x_m)$$

$$r^2 = \frac{1}{\sum_{n=1}^{N}\sum_{m=1}^{N} \alpha_n \alpha_m y_n y_m \varphi(x_n)^T \varphi(x_m)}$$

The transformation φ affects the geometry of the decision boundary by mapping the input vectors to a higher-dimensional space, where they can be linearly separated.

This is reflected in the inner product $\varphi(x_n)^T \varphi(x_m)$

(2) Proof : $\frac{1}{r^2} = \sum_{n=1}^{N} a_n$.

Given.

$$r^2 = \frac{1}{\sum_{n=1}^{N}\sum_{m=1}^{N} \alpha_n \alpha_m y_n y_m \varphi(x_n)^T \varphi(x_m)} \quad ①$$

$$\frac{1}{r^2} = \sum_{n=1}^{N}\sum_{m=1}^{N} \alpha_n \alpha_m y_n y_m \varphi(X_n)^T \varphi(X_m)$$

kernel funcition increase the dimension of X

Since Dual problem = Primal problem.

$$\frac{1}{2}\cdot\frac{1}{r^2} = \frac{1}{2}\|w^*\|^2 = \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \varphi(X_i)^T \varphi(X_j)$$

$$= \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\cdot\frac{1}{r^2}$$

$\therefore \frac{1}{r^2} = \sum_{i=1}^{N} \alpha_i$ (this hold true whether kernel function is used or not).