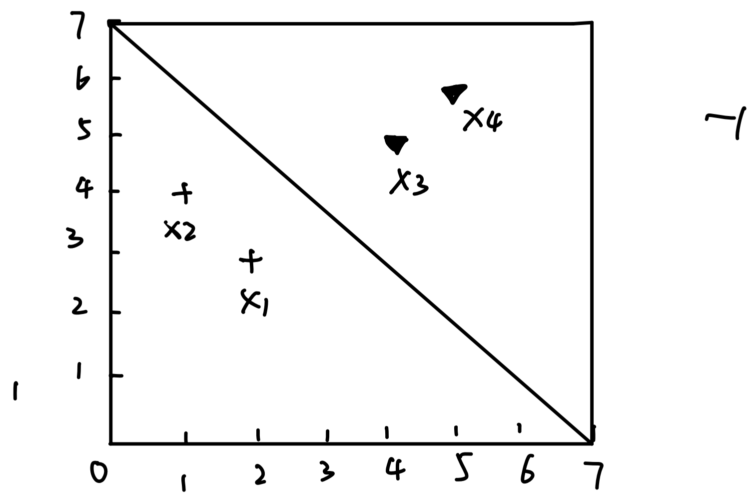
22-S2-Q3

Q: (a) SVM

(i) ① decision boundary



② we denote $x_1(2,3)$ $x_2(1,4)$ $x_3(4,5)$ $x_4(5,6)$

we consider x_1 x_2 x_3 are support vector

we have $\lambda_1, \lambda_2, \lambda_3 \neq 0$

$$\sum_{i=1}^{N} \lambda_i y_i = 0$$

$$\lambda_1 + \lambda_2 - \lambda_3 = 0$$

$$w = \sum_{i=1}^{N} \lambda_i y_i x_i$$

$$\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \lambda_1 \begin{pmatrix} 2 \\ 3 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ 4 \end{pmatrix} - \lambda_3 \begin{pmatrix} 4 \\ 5 \end{pmatrix}$$

$$(w_1\ w_2)\begin{pmatrix} 2 \\ 3 \end{pmatrix} + b = 1 \qquad (1)$$

$$(w_1 w_2)\begin{pmatrix}1\\4\end{pmatrix} + b = 1 \qquad (2)$$

$$(w_1 w_2)\begin{pmatrix}4\\5\end{pmatrix} + b = -1 \qquad (3)$$

$$(1)(2)(3) \rightarrow \quad w = \begin{pmatrix}-\frac{1}{2}\\-\frac{1}{2}\end{pmatrix} \qquad b = \frac{7}{2}$$

$$-\frac{1}{2}x_1 - \frac{1}{2}x_2 + \frac{7}{2} = 0$$

$$x_1 + x_2 = 7$$

$$d = \frac{2}{\|w\|} = \frac{2}{\sqrt{\frac{1}{4}+\frac{1}{4}}} = 4\sqrt{2}$$

③ The linear decision boundary leads to max margin because SVM selects the hyperplane that max the distance between itself and the support vectors.
This is achieved by min $\|w\|$ resulting the widest margin $\frac{2}{\|w\|}$

(ii) Support vector : $x_1(2,3)$ $x_2(1,4)$ $x_3(4,5)$

Reason ① They satisfy the Karush - Kuhn - Tucker condition $\alpha_i > 0$

② Removing or moving any one of them would

change the optimal hyper-plane.
Relocationg x_1 would not.

(b) (i) ① sharing parameters slashing memory
② much lower risk of over-fitting
③ retaining the receptive-field hierarchy

(ii) ① Prevents Overfitting: By randomly omitting neurons, the netuork canaot rely ~~too~~ heavily on specific features, forcing it to learn more robust and generalized representations
② Acts as ne Ensemble Method: During trainning, different sub-networks are trained due to the randomness of dropped neurons

(iii) Data augmentation
① Data augmentation increases the diversity of training data, helping the model generalize better without reducing complexity
② Early stopping and dropout both reduce model capacity, so they trade variance for higher bias