23-S1-Q3

Q(a)(i) plot SVM + boundary

(ii) SV

(iii) margin

(b)(i) CNN Ad.

(ii) Tree vs KNN

(c) (i) Neural

(ii) overfit

Solution (a) (i) ① samples plot

we denote $x_1(1,2)$ $x_2(1,9)$, $x_3(5,5)$ as ● +1

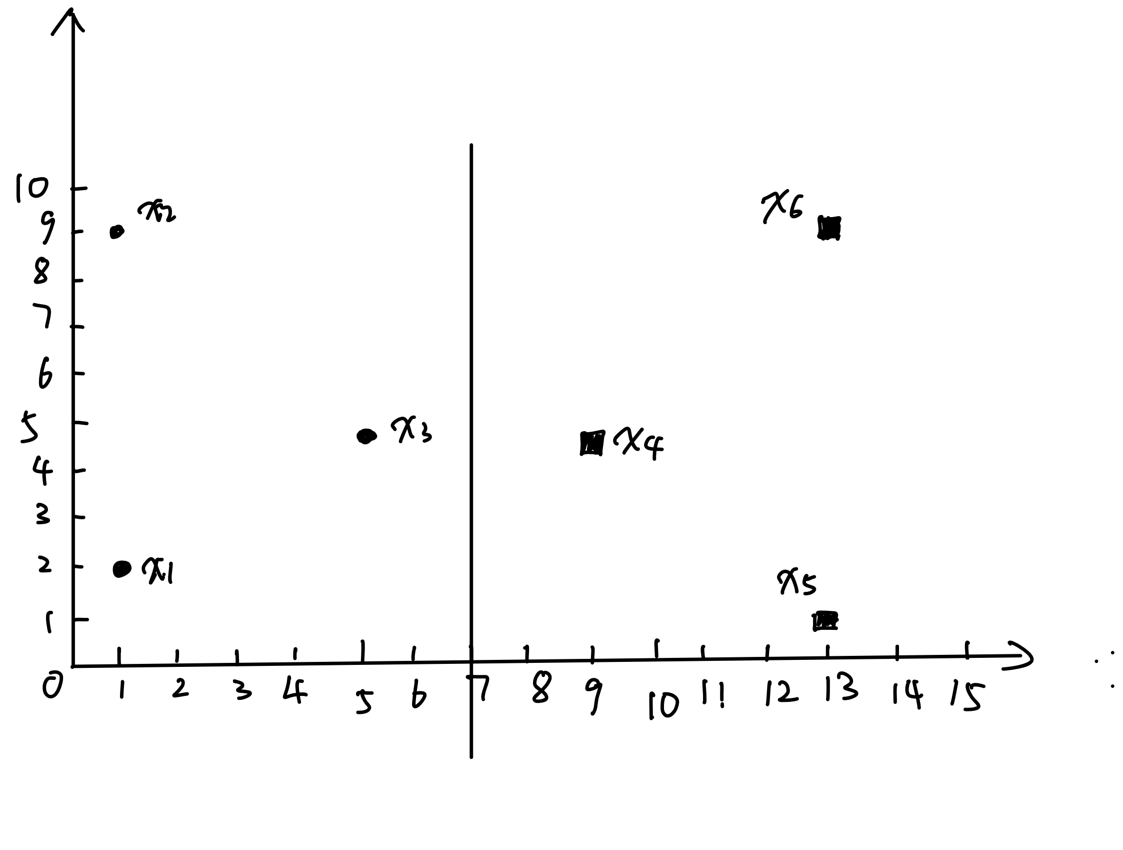
$x_4(9,5)$ $x_5(13,1)$ $x_6(13,9)$ as ■ −1

the decision boundary at $x=7$ ∵

Because every class-1 point has $x \le 5$

class-2 $x \ge 9$

placing the decision boundary half way at $x=7$ yeilds the maximum margin separator



we consider x_3 and x_4 as support vector

$\lambda_3, \lambda_4 \neq 0$

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

$$\lambda_3 - \lambda_4 = 0$$

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

$$\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \lambda_3 \begin{pmatrix} 5 \\ 5 \end{pmatrix} - \lambda_4 \begin{pmatrix} 9 \\ 5 \end{pmatrix}$$

$$y_i (w \cdot x_i + b) \geq 1$$

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

$$(w_1\ w_2)\begin{pmatrix} 9 \\ 5 \end{pmatrix} + b = -1$$

$$5w_1 + 5w_2 + b = 1$$

$$9w_1 + 5w_2 + b = -1$$

$$4w_1 = -2$$

$$w_1 = -\frac{1}{2} \qquad w_2 = 0 \qquad b = \frac{7}{2}$$

$$w = \begin{pmatrix} -\frac{1}{2} \\ 0 \end{pmatrix}$$

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

(ii) ① Support vectors are Class 1 $x_3(5,5)$
and Class 2 $x_4(9,5)$
② They are the closest points to the hyper plane (distant = 2) and therefore define it
③ Moving or removing them would move the boundary.

(iii) ① For a line $wx+b=0$, margin $= \frac{2}{||w||}$
Here $w = [-\frac{1}{2}, 0]^T$ so $||w|| = \frac{1}{2}$
and margin $d = 4$
② Any other admissible hypler-plane would either shrink this distance or misclassif a point, so the SVM solution is the unique one that maximizes the margin

(b)(i) ① Local receptive field & weight sharing
convolution kernels reuse parameter, slashing memory and overfitting risk

② Translation invariance
achieved through convolutions + pooling, so an object is recognised wherever is appears

③ Automatic hierarchical feature learning
early layers learn edges/colours, deeper layers learn shapes and objects, avoiding handcrafted descriptors

(ii) Decision tree VS KNN

① Model Type
Decision tree is eager (model build once)
K-NN is lazy (all computation deferred to query time.

② prediction cost
Decision Tree : $O(\log N)$ after traing
K-NN : $O(N \cdot d)$ per query

③ Interpretability & feature weight
Decision Tree gives explicit IF-THEN rules and can ignore irrelevant attributes

KNN is opaque and sensitive to irrelevant or differently-scaled features

(c) (i) The test-accuracy curve peaks at ≈ 10 neurons (≈ 0.75) and declines afterwards. Select 10 neurons to maximise expected accuracy on future data drawn from the same distribution.

(ii) ① At 10 neurons the gap between training (≈ 0.78) and ~~testing~~ (≈ 0.75) accuracy is about 3 percentage-points, so only mild over-fitting is anticipated.

② When the network grows (≥ 50 neurons) the gap widens to > 10 points, evidencing heavier over-fitting