

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

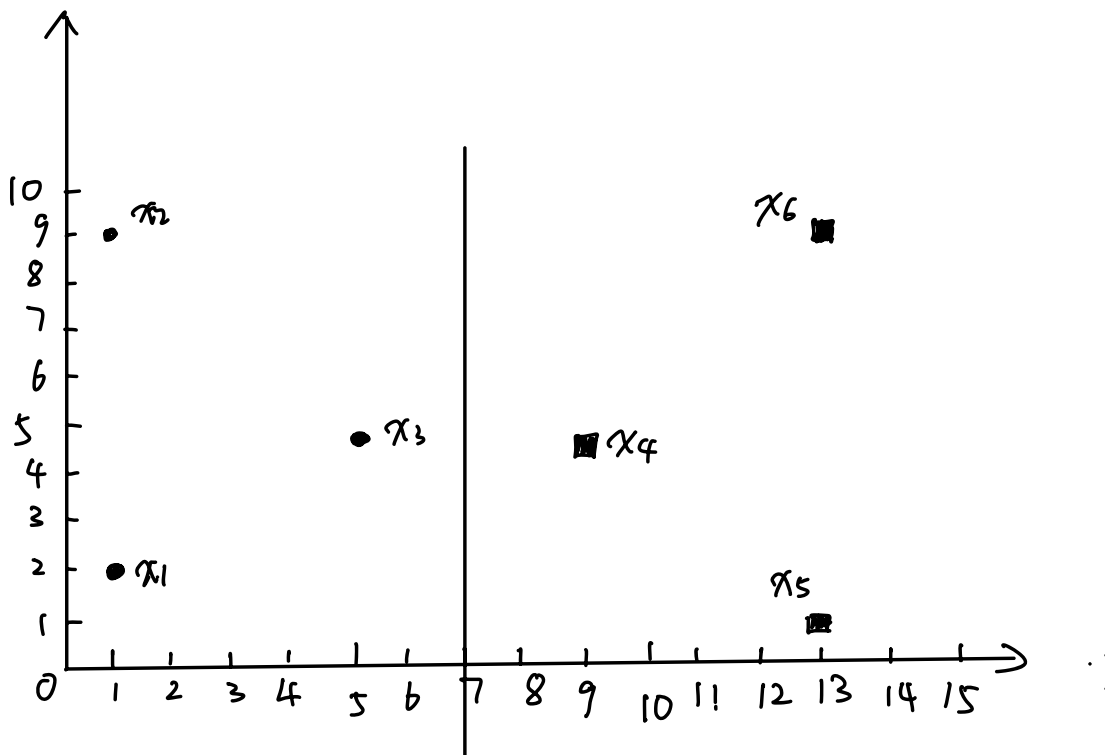
Solution (a) (i) ① samples plot

we denote $x_1(1,2)$ $x_2(1,9)$ $x_3(5,5)$ as \bullet
 $x_4(9,5)$ $x_5(13,1)$ $x_6(13,9)$ as \blacksquare

the decision boundary at $x=7$..

Because every class-1 point has $x \leq 5$
class-2 $x \geq 9$

Placing the decision boundary half way
at $x=7$ yields the maximum margin
separator



(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 (distance = 2) and therefore define it

③ Moving or removing them would move the boundary.

(iii) ① For a line $w \cdot x + b = 0$, margin = $\frac{2}{\|w\|}$

Here $w = [1, 0]^T$ so $\|w\| = 1$

and margin $d = 4$

② Any other admissible hyper-plane would either shrink this distance or misclassify 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 parameters, slashing
memory and overfitting risk

② Translation invariance

achieved through convolutions + pooling, so an
object is recognised wherever it 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 built once)

K-NN is lazy (all computation deferred to
query time).

② Prediction cost

Decision Tree: $O(\log N)$ after training

K-NN: $O(N \cdot d)$ per query

③ Interpretability & feature weights

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

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