

24-51-Q3

Q(a) (i) $X = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 2 \\ 1 & 3 & 1 \end{bmatrix}$ $H = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$

(ii) CNN : 4 layers

3x3 filter 1 stride.

no padding no pooling.

output size 1x1 input size ?

(iii) NN 3 hidden layer

hidden nodes ↓

effect : bias & variance?

(b) (i) margin of SVM-2 change

(ii) SVM-3 $\frac{w}{2} \frac{b}{2}$

(iii) SVM-3 margin?

(c) (i) 7th epochs : parameter estimates

Solution (i)

$$\textcircled{1} \quad X = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 2 \\ 1 & 3 & 1 \end{bmatrix} \quad H = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

$$\textcircled{2} \quad \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

$$= 1 \cdot 1 + 2 \cdot 1 + 2 \cdot 1 + 3 \cdot 2$$

$$= 1 + 2 + 2 + 6$$

$$= 11$$

$$\textcircled{3} \quad \begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

$$= 2 + 1 + 3 + 4$$

$$= 10$$

$$\textcircled{4} \quad \begin{bmatrix} 2 & 3 \\ 1 & 3 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

$$= 2 + 3 + 1 + 6$$

$$= 12$$

$$\textcircled{5} \quad \begin{bmatrix} 3 & 2 \\ 3 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

$$= 3 + 2 + 3 + 2$$

$$= 10$$

\textcircled{6} Hence the 2×2 output is

$$\begin{bmatrix} 11 & 10 \\ 12 & 10 \end{bmatrix}$$

(ii) ① assume the input image size is $N \times N$

$$\frac{N - \text{filter size} + 2 \times \text{padding}}{\text{stride}} + 1 = \text{output size}$$

$$\frac{N - 3 + 2 \times 0}{1} + 1 = N - 2$$

② 4 layers

$$N \rightarrow N - 2 \rightarrow N - 4 \rightarrow N - 8 = 1$$

$$\text{So } N = 9$$

Thus the input image must be a 9×9

(iii) ① Bias will tend to increase, ^{image.}
due to decreasing the number of hidden
units in each layer reduces the model
capacity

② Variance will tend to decrease, due to
the model become less sensitive
to small fluctuation in the training
data

(b) Solution (i)

① SVM - 1 pin the support vectors at ± 1

So, its margin $d = \frac{2}{\|w\|}$

② SVM - 2 : constraints are just $w \cdot x + b \geq 0$

or < 0 , one can freely rescale (w, b) and still satisfy the same inequalities.

That means SVM-2's margin isn't pinned down at a finite value.

The solution isn't unique and the margin is not well-defined.

cii) ① for a class - 1 support vector x
with $w \cdot x + b = 1$

$$\frac{w}{2} \cdot x + \frac{b}{2} = \frac{1}{2} > 0$$

② for a class - 2 support vector x
with $w \cdot x + b = -1$

$$\frac{w}{2} \cdot x + \frac{b}{2} = -\frac{1}{2} < 0$$

③ So no training point switches side,
and SVM - 3 make no classification
errors on the training set.

(iii) ① SVM - 1 : $d = \frac{2}{\|w\|}$

② SVM - 3 :
support vector is

$$\frac{w}{2}x_1 + \frac{b}{2} = \frac{1}{2} \quad (1)$$

$$\frac{w}{2}x_2 + \frac{b}{2} = -\frac{1}{2} \quad (2)$$

$$(1) - (2) : \frac{w}{2}(x_1 - x_2) = 1$$

$$w \cdot (x_1 - x_2) = 2$$

$$\|w\| d = 2$$

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

Hence SVM - 3 has exactly
the same margin as SVM - 1

Solution (c)(i)

① By inspection, the training loss goes monotonically down while validation loss eventually turns upward, means overfitting.

② The validation-loss curve is lowest at epoch 4. That point achieves the best generalization on the validation set.

(ii)

① The plots shows that by epoch 7, the training loss is very low but the validation loss has risen sharply.

② Overfitting has occurred

③ Hence, the parameter at epoch 7 will likely generalize poorly

(iii)

① Because the training loss can continue going down to near-zero while the validation loss increases, the model clearly has sufficient capacity to overfit the training data.

(iv)

① We would use regularization, such as L1 regularization or L2 ~

② Or we could use techniques, such as early stopping, dropout or data augmentation

(v) ① Training loss typically goes up slightly, because with much more data, the network cannot "memorize" them all as easily

② Validation loss often goes down ,
because having many more training
examples improves generalization