24-51-03 (Q(a)(i)) $X = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 2 \\ 1 & 3 & 1 \end{bmatrix}$ $H = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$ (ii) CNN: 4 layers 3x3 filter 1 stride. no padding no pooling. output size 1x1 input size? (iii) NIY 3 hidden layer hidden notes L effect: bias & variance? (b) (i) margin of SVM-2 change (ii) SVM-3 <u>w</u> <u>b</u> 2 (iii) SVM-3 margine? (C) (i) 7th epochs: pavameter estimaly Solution (i)

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

$$\mathcal{D}\begin{bmatrix} 3 & 2 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

(ii)
$$\odot$$
 assume the input image size is N×N

N-filter size + 2×paddiby

stride.

$$N-3+2x0$$
 +1 = $N-2$

② 4 layers

$$N \rightarrow N-2 \rightarrow N-4 \rightarrow N-8 = 1$$
 $So N = 9$

Thus the input image must be a 9x9 (iii) O 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) Solwtion (i)

① SVM - 1 pin the support vectors at ± 1 So, its margin $d = \frac{2}{11w11}$

②SVM-2: constraints are just wix+b≥o 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 0.

- (ii) O for a class 1 support vector xwith $\frac{w}{z} \cdot x + \frac{b}{z} = 1$
 - ② for a class-2 support vector x with $\frac{w}{z} \cdot x + \frac{b}{z} = -1$
 - 3 SUM discriminate function is = (w.x+b) = 0 and it is total same with SUM-1
- (4) So no training point switches side, and SVM-3 make no classification errors on the training see.

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

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

$$\frac{w}{2}X_{L}f + \frac{b}{2} = -| \qquad (2)$$

$$C(1)-C(2): \frac{\omega}{2}(\chi_1-\chi_2) = 2$$

$$W \cdot (X_1 - X_2) = 4$$

Hence, the margin of SVM-3 is twice as large as that of SVM-1

Solution (c)(i)

- By inspection, the training loss goes mono tonically down while validation loss even tually turns upward, means over fitting,
 - 2) The validation-loss carse is lowest at epoch 4. That point a chieves the best generalization on the validation set.
 - Othe 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

(:::)

OBecause 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.

(11)

- DWe would use regularization, such as L1 regularization or L2 ~
- Or we could use techniques, such as early stopping, dropout or data augmentation
- (V) O Training loss typically goes up slighty, because with much more data, the network cannot "memorize" them all as easily

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