21 – 51 – Q 3
(a) (i) SVM
(ii) CNN
(b) graph

Solution (a)(i) SVM

① remain the same

② In a linear SVM only the support vectors influence the optimisation problem.

③ A point that is well outside the margin has a Lagrange multiplier $\alpha = 0$

④ Deleting it leaves the quadratic program unchange

⑤ So, weigh vector w and bias b are unchanged.

(ii) ① w_1 : B w_2 A

     w_3 : C w_4 : C

② A : +1 /0/-1 vertical edge detector

defect left edge

   B: -1 /0/ +1 ~ defecte right edge

   C: indentify kernel leavel the image unchange

```
0 0 0 0 0 0 0 0
0 0 1 1 1 0 0
0 2 0 0 1 0 0
0 0 1 1 1 0 0
0 0 0 0 1 0 0
0 0 1 1 1 0 0
0 0 0 0 0 0 0 0
```
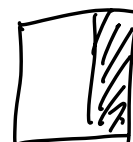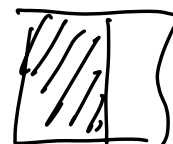
White : 0

black : 1, 2, 3 ...

padding : 1

A.
```
1 0 -1
1 0 -1
1 0 -1
```
→
```
-1 -1 -1  1  2
-2 -2 -1  2  3
-1 -1 -2  1  3
-2 -2 -1  2  3
-1 -1 -1  1  2
```
ReLu →
```
0 0 0 1 2
0 0 0 2 3
0 0 0 1 3
0 0 0 2 3
0 0 0 1 2
```


B.
```
-1 0 1
-1 0 1
-1 0 1
```
→
```
1 1  1 -1 -2
2 2  1 -2 -3
1 1  2 -1 -3
2 2  1 -2 -3
1 1  1 -1 -2
```
ReLu →
```
1 1 1 0 0
2 2 1 0 0
1 1 2 0 0
2 2 1 0 0
1 1 1 0 0
```

(iii)

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

$$= \frac{100 - 7 + 2 \times 0}{1} + 1$$

$$= 94$$

output : $94 \times 94 \times 1$

(b) (i) ① bias lower

② variance higher

(ii) ① bias stays roughly the same or grow slightly

② Because model capacity cannot fit more data

③ variance lower

④ Because model can improve generalization performance in the test set