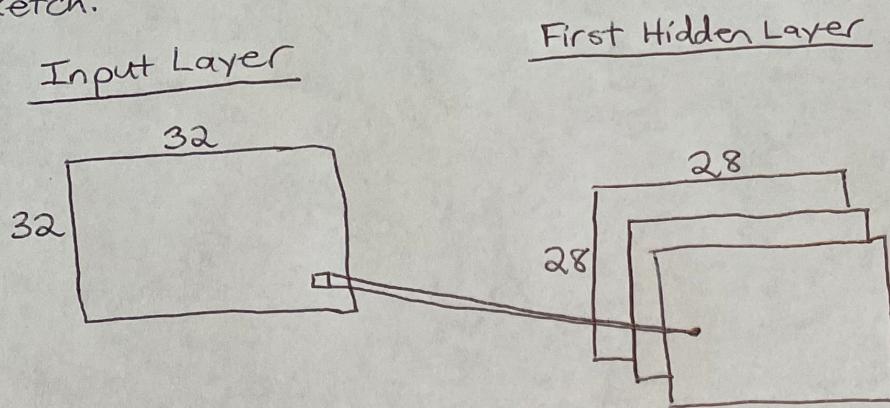


4.)

(a) Sketch.



→ Why are each of the 3 two-dimensional feature maps in the first hidden layer 28 by 28?

- Problem says no boundary padding.
- Since each convolutional filter is 5 by 5, in each of the 2 dimensions, there are 28 positions for a line of length 5 to be positioned along a line of length 32. General formula for each dimension:

$$\begin{aligned}\text{Dim-Hidden-Layer} &= \text{Dim-Prior-Layer} - (\text{Dim-Filter}-1) \\ &= 32 - (5-1) \\ &= 28\end{aligned}$$

(b.) ~~QUESTION~~ How many parms?

$$\text{Number Parameters} = 3 \times [28 \times 28] = \boxed{2,352 \text{ parameters}}$$

(c.) Not sure if this is the solution the authors were looking for:

→ To frame my response, I first show via the example on page 413 how to recast a convolutional layer as a feed forward one...

$$\begin{bmatrix}
 a & b & c & d & e & f & g & h & i & j & k & l \\
 d & e & f & g & h & i & j & k & l & a & b & c \\
 g & h & i & j & k & l & a & b & c & d & e & f
 \end{bmatrix} \cdot \begin{bmatrix} \alpha & 0 \\ \beta & \alpha \\ 0 & \beta \\ \gamma & 0 \\ \delta & \gamma \\ 0 & \delta \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix}
 ad + b\beta + dy + e\delta & ba + cp + ey + f\delta \\
 da + e\beta + gy + h\delta & ea + fb + hy + i\delta \\
 ga + th\beta + jy + k\delta & ha + i\beta + ky + l\delta
 \end{bmatrix}$$

$X$

$\Theta$

→ Here,  $\Theta = \begin{bmatrix} X_{11} & X_{12} \\ \vdots & \vdots \\ X_{12,1} & X_{12,2} \end{bmatrix}$ ; And all  $X_{ij} = 0$  except  $X_{11}, X_{21}, X_{41}, X_{22}, X_{32}, X_{52}, X_{62}$ , and  $X_{51}$

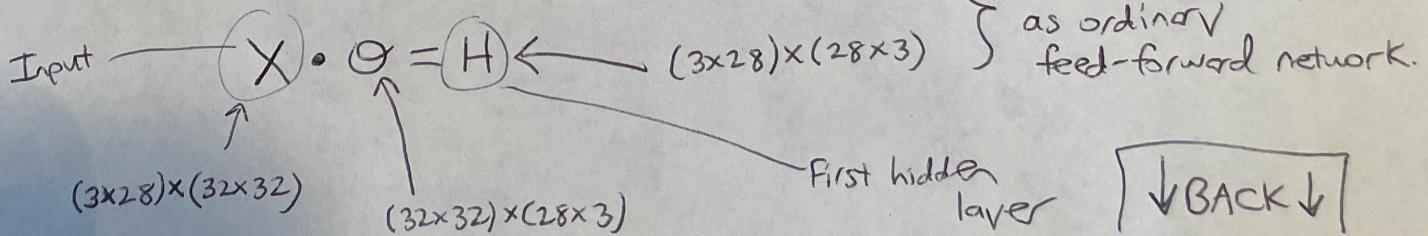
$$\text{Also, } X_{11} = X_{22}$$

$$X_{21} = X_{32}$$

$$X_{41} = X_{52}$$

$$X_{51} = X_{62}.$$

→ Analogously, back in the problem in #4(c), we desire 3 two-dimensional feature maps, which I will place side-by-side, I reshape the Input Layer as a  $(3 \times 28) \times (32 \times 32)$  tensor and the weights map as a  ~~$(32 \times 32) \times (28 \times 3)$~~  tensor, thus:



→ Analogously as above, since there are  $5 \times 5 = 25$  parameters in the convolutional filter, thus in each column of  $\Theta$ , all except 25 parameters are 0.

Further, the 25 parms in any col i must equal their counterparts in any other column j = the constraints.

(d.) If there were no constraints, there would be

$(32 \times 32) \times (28 \times 3) = \underline{86,016 \text{ parameters}}$  in the neural network.  
in (c)