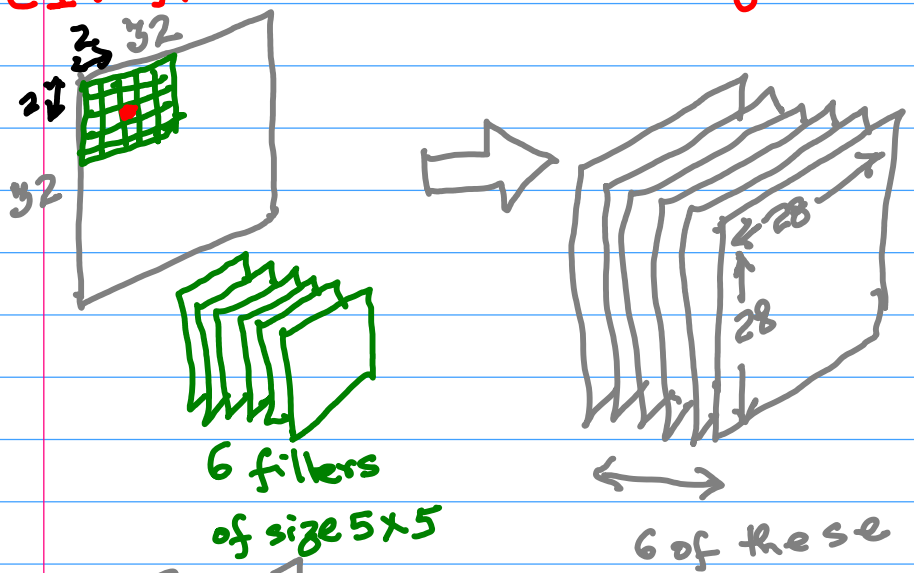
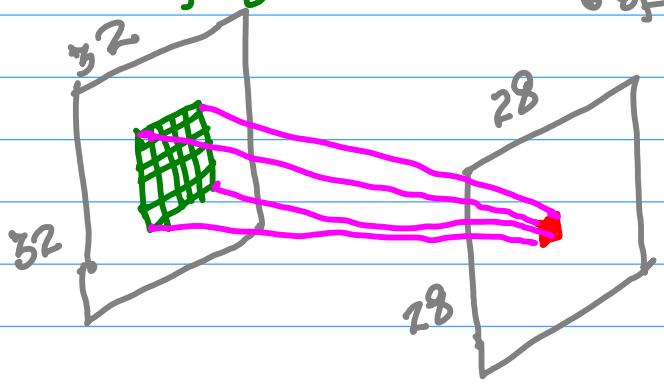
LeNet-5 DETAILS

C1: first convolutional layer





$$a_j = \sum_{i=1}^{25} w_{ji}^{(1)} x_i + w_{j0} \quad \text{(bias)}$$

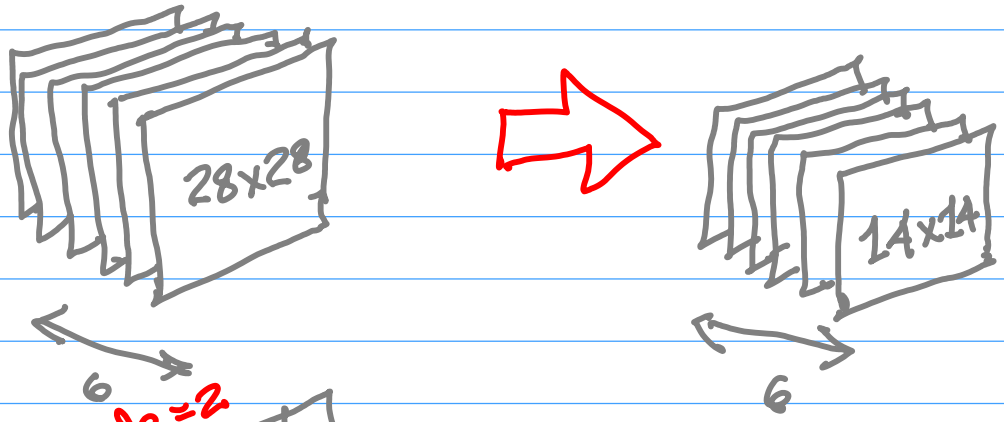
parameters

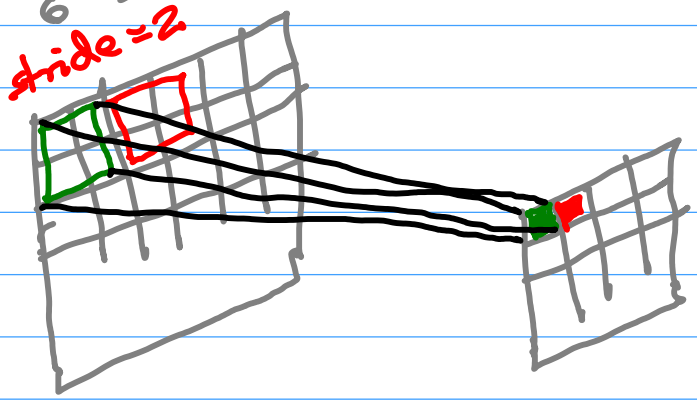
$$6 \times [5 \times 5 + 1] = 6 \times 26 = 156$$

connections: for each of the 6 filters, for each pixel in the resultant 28×28 image

$$28 \times 28 \times 6 \times \underbrace{[5 \times 5 + 1]}_{26} = 1,22,304$$

Second layer (S2) "Sub-sampling" or Average Pooling layer





Stride = 2

Average pooling
same weight on each 4

$$a_j = \sum \left(w_{ji} x_i + w_{j0} \text{ or } b_j \right)$$

resultant: $\frac{w}{4}$ → 1 parameter

bias → 1 parameter

this is throughout the image, independent of the image size (14x14)

2 x 6 = 12 parameters

How many connections:

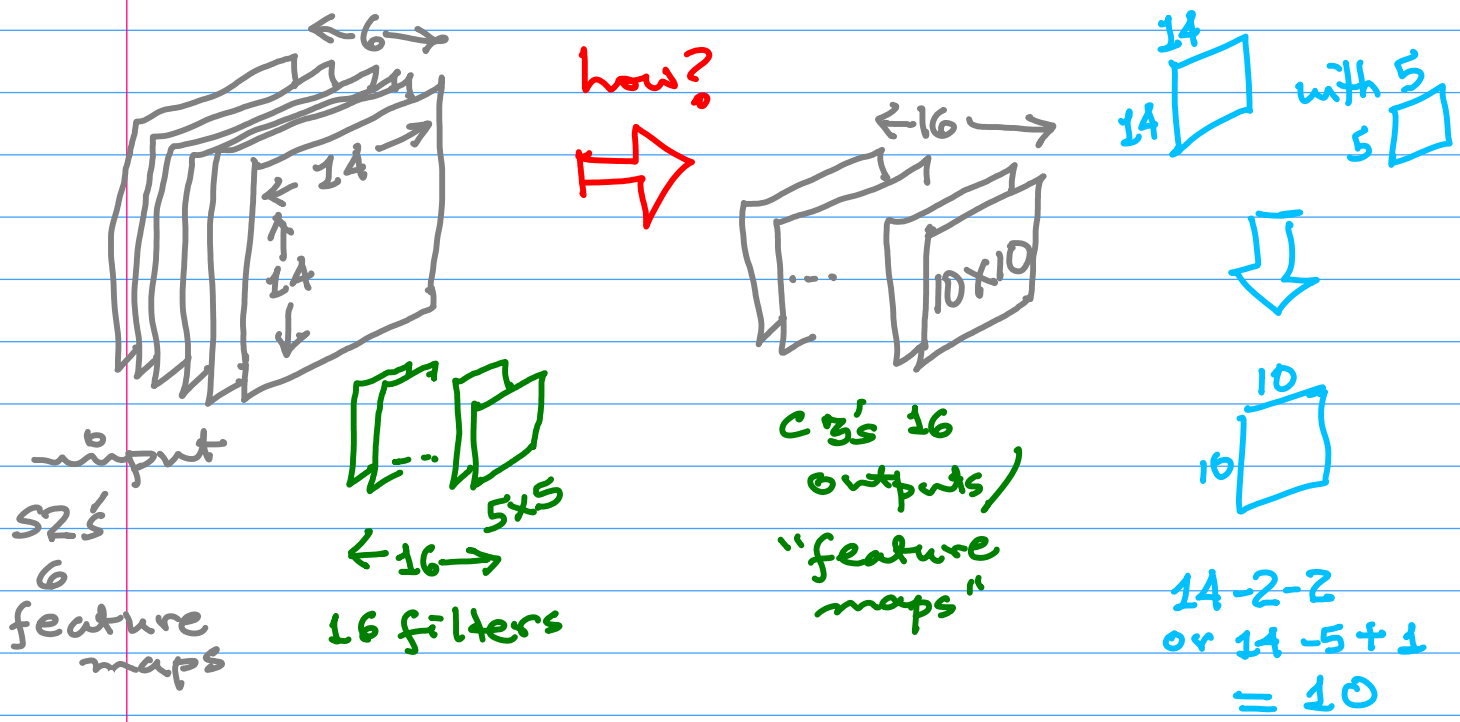
$$6 \times (4+1) \times 14 \times 14 = 5{,}880 \text{ connections}$$

6: for each filter/channel

$(4+1)$: connections

14×14: image

C3: 3rd Convolutional layer



How do 6 images of size 14 x 14 map onto 16 images of size 10x10

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	x				x	x	x			x	x	x	x		x	x
1	x	x				x	x	x			x	x	x	x		x
2	x	x	x				x	x	x			x		x	x	x
3		x	x	x			x	x	x	x			x		x	x
4			x	x	x			x	x	x	x		x	x		x
5				x	x	x			x	x	x	x		x	x	x

"3 at a time" — "4 at a time" → "all 6"

Each column indicates which feature maps in S2 are combined by the units in a particular feature map of C3

→ To break the symmetry in the network

→ to keep the number of connections within reasonable bounds.

16 filters of size 5×5 give us 16 output feature maps of size 10×10

However, which ones of the 6 input feature maps of the previous layer do we consider, to take inputs from? Any one / some / all ?

Answer: 3 at a time + 4 at a time + 6 at a time

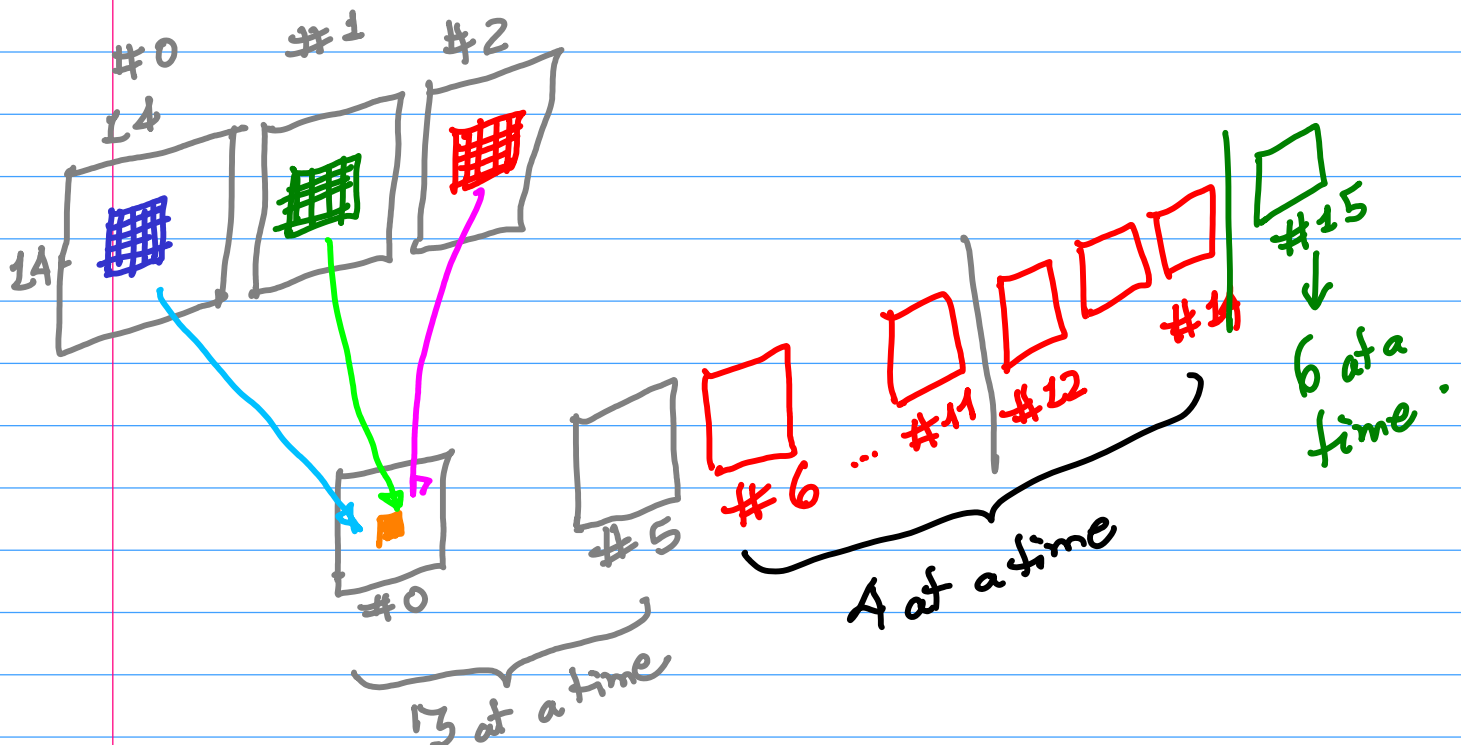
How many parameters ?

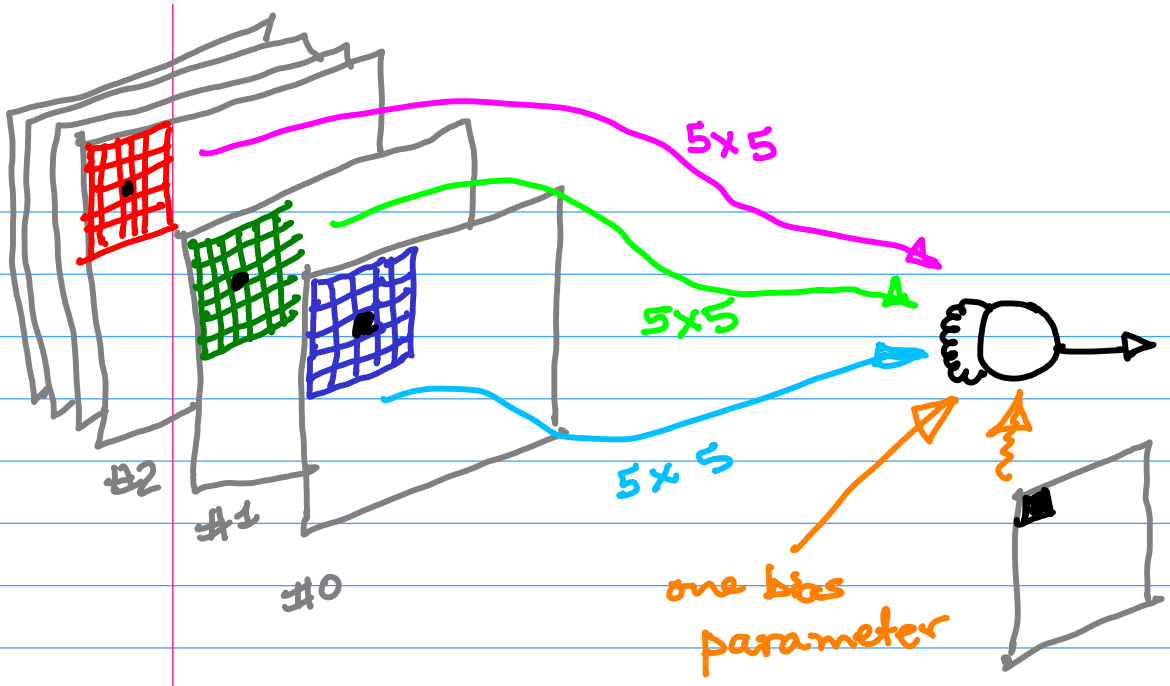
"3 at a time"	"4 at a time"	"all 6"
(#0 to #5: 6 of these)	(#6 to #14: 9 of these)	(#15: 1 of this)
take input from 3	take input from 4	takes input from 6

$6 \times (3 \times 5 \times 5 + 1) + 9 \times (4 \times 5 \times 5 + 1) + 1 \times (6 \times 5 \times 5 + 1)$

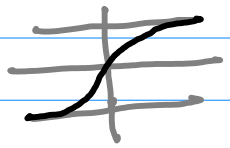
$= 6 \times (75+1) + 9 \times (100+1) + 1 (150+1) = 1516.$

How many connections? for each pixel in the 10×10 image $= 1516 \times 100 = 1,51,600$ connections.





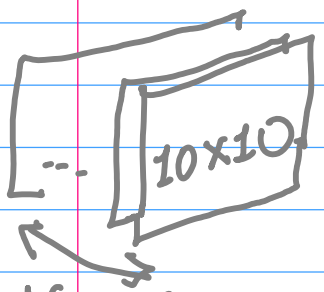
$h(\cdot) = \tanh(\cdot)$



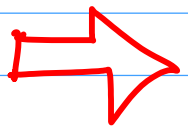
Example: for the first out of the "5 at a time" cases.

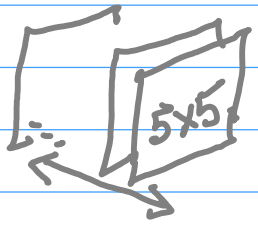
(4) "S4" The fourth (subsampling) layer

how? half the dimensions
⇒ stride = 2

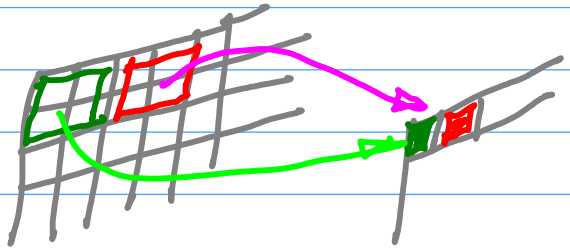


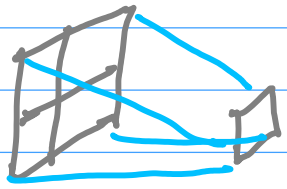
16 of these
Input





16 of these





Same weight for all 4 connections
+ 1 bias = 2 parameters,
across the entire image, for each of the
16 output 5x5 images
= (1+1) x 16 = 32.

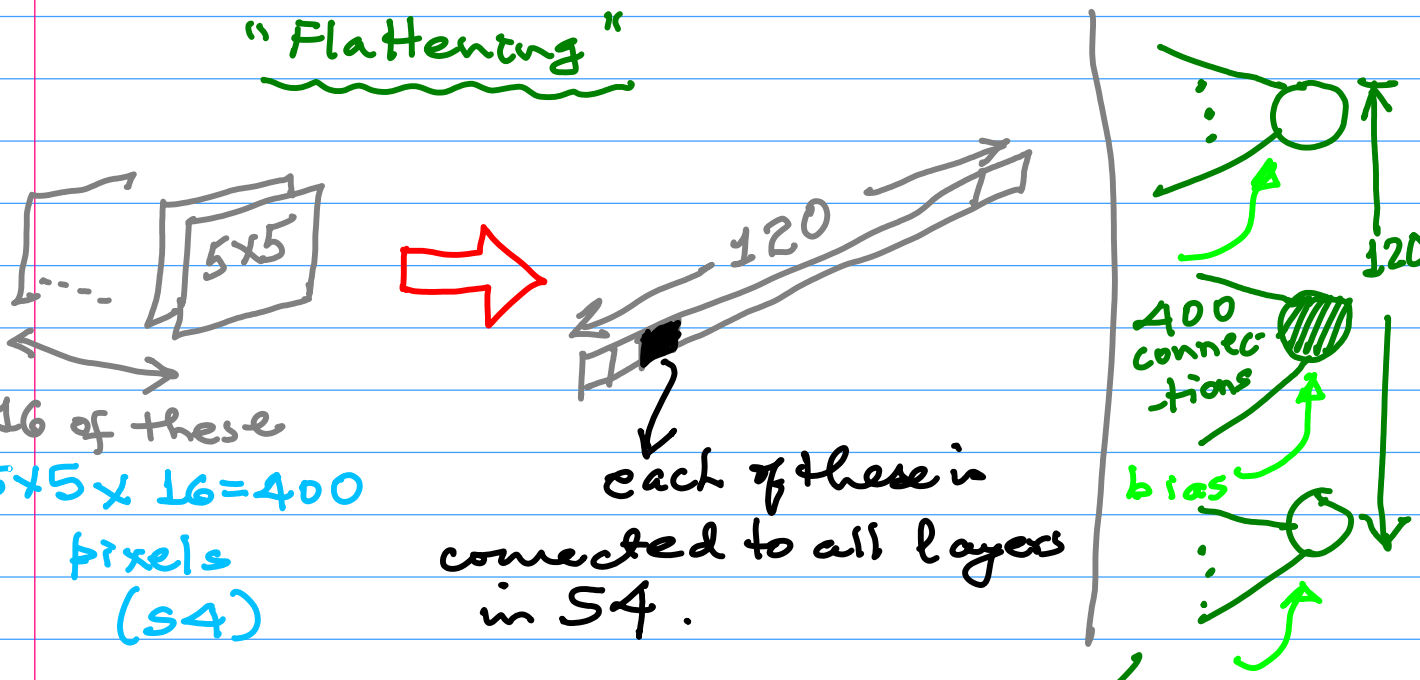
Number of connections: The above structure repeats for each pixel in the 5×5 output image: 16 of them:

$$\Rightarrow (5\times5)\times 16 \times (4+1) = 5\times5\times80 = 2000$$

(5×5: size; 16: 16 images; 4+1: square 2×2 + bias)

(5) "C5" 5th layer, C ⇒ convolution

"Flattening"



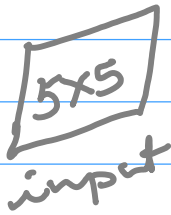
Parameters: For each of these 120 neurons e.g., the jth one: there are 400 weights + 1 bias term

$$\Rightarrow 120\times(400+1) = 120\times401 = 48{,}120 \text{ parameters}$$

Connections: The number of connections is the same as the number of parameters. This is

a fully connected layer.

Why is this called a convolutional layer

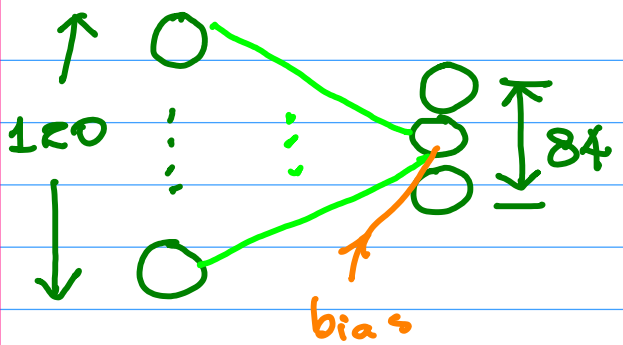


with a 5x5 mask → results in a 1x1 image.

there are 120 of these.

Why 120 & not any other number?
(arbitrary!)

(6) "F6" Fully connected layer



Parameters:

$84 \times [120 + 1]$

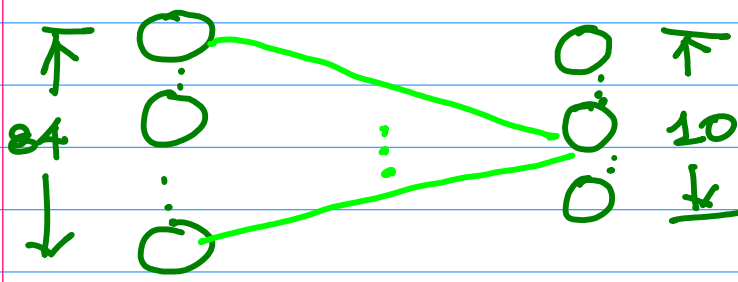
- 84 → # of outputs
- 120 → weights
- +1 → bias

$= 84 \times 121$

$= 10{,}164$

Connections: (same)

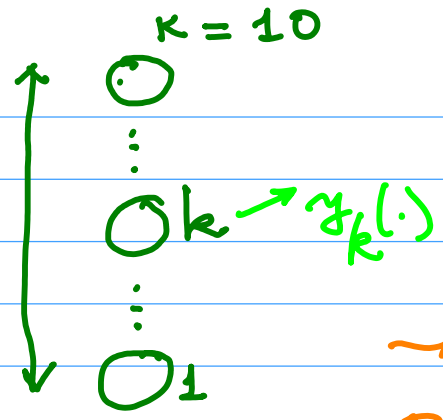
(7) Output layer



The activation fcts at all other layers were the tanh(.) function

(smooth version of the sign fct of the Perceptron)

For this output layer, it is the SOFTMAX



$$\sigma_k(y) \triangleq \frac{e^{y_k}}{\sum_l e^{y_l}}$$

why?

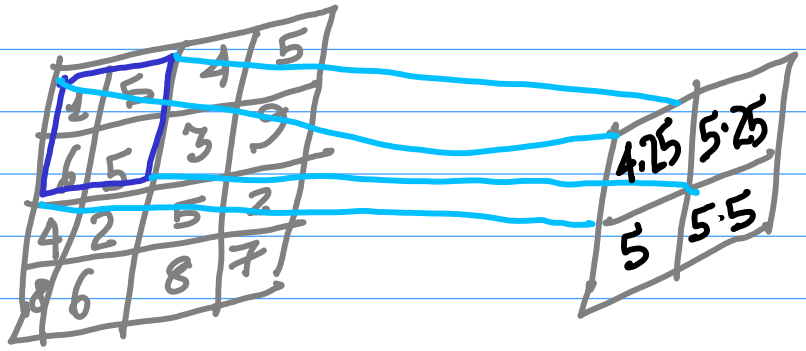
normalisation: getting a probability. → like a predictor.

Why did we raise a number to an exponential? → large #, & not a small ratio.

SOME IMPORTANT CONCEPTS

POOLING:

1) Average Pooling (LeNet)



2) Max Pooling (AlexNet)