# NPTEL Week 8 Live Sessions

on Deep Learning (noc24_ee04)

A course offered by: Prof. Prabir Kumar Biswas, IIT Kharagpur

- Python Coding: Autoencoder, Conv-autoencoder, Denoising autoencoder, PCA
- Week 8 practice question


By

**Arka Roy**

NPTEL PMRF TA

Prime Minister's Research Fellow

Department of Electrical Engineering, IIT Patna

Web: https://sites.google.com/view/arka-roy/home


Powered by:

PMRF

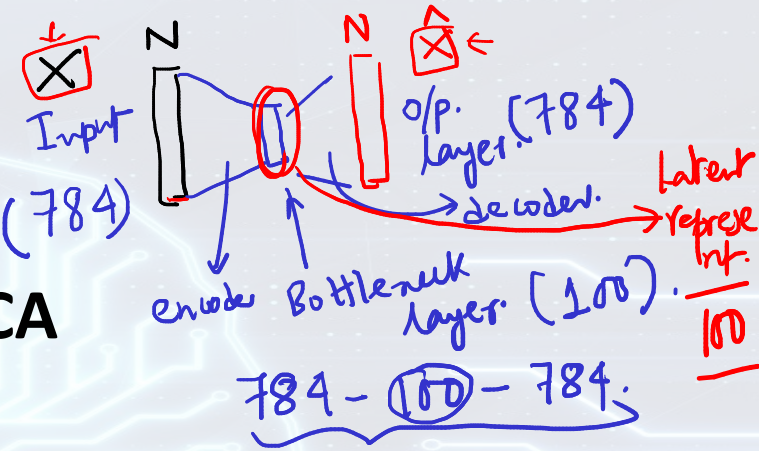
Prime Minister's Research Fellows

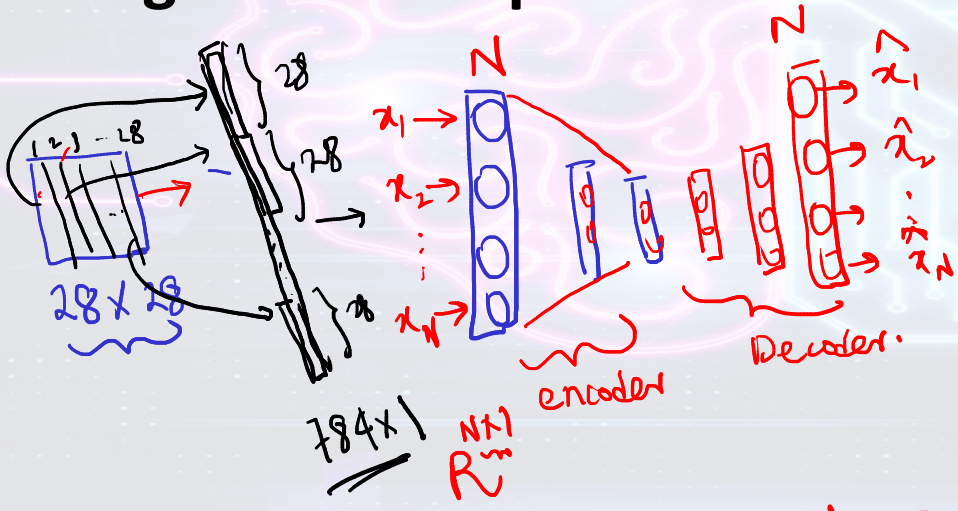
Ministry of Education

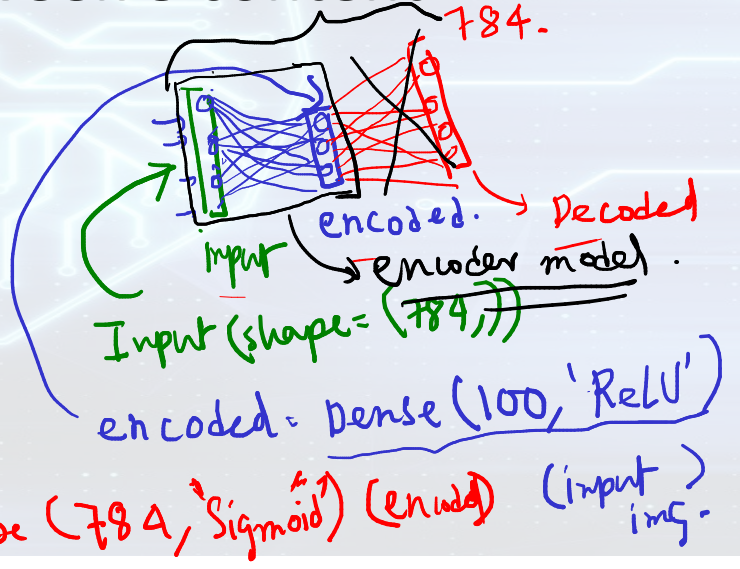
Government of India

# Content of the live session

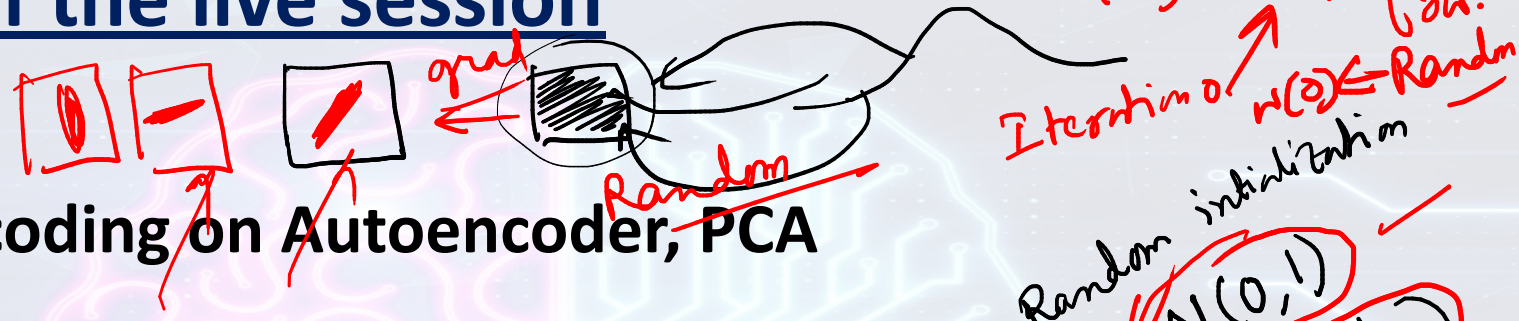
1. Python coding on Autoencoder, PCA
2. Solving numerical problems from week 8 content



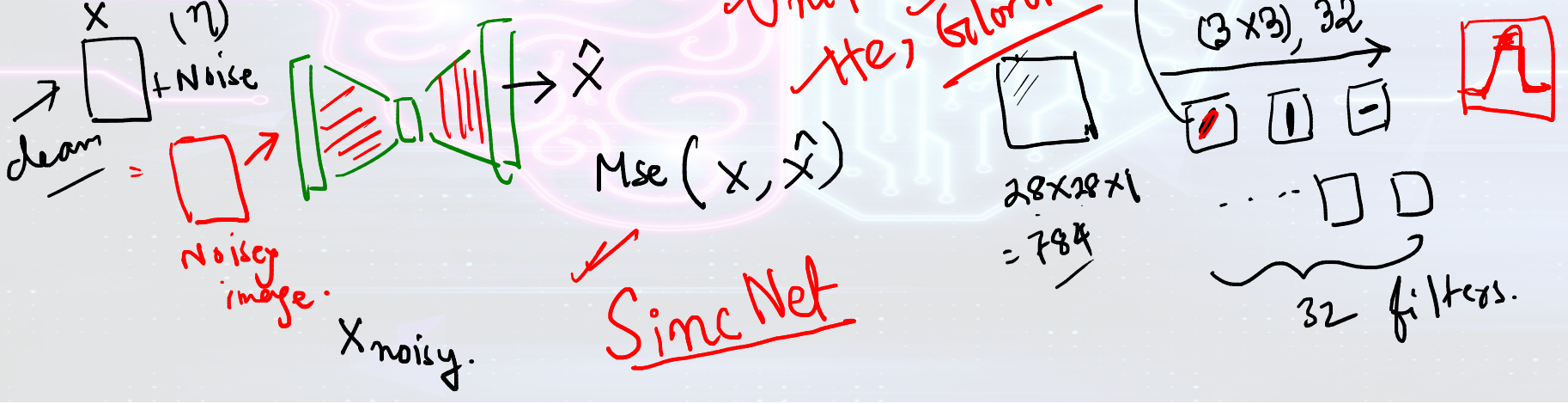



# Content of the live session

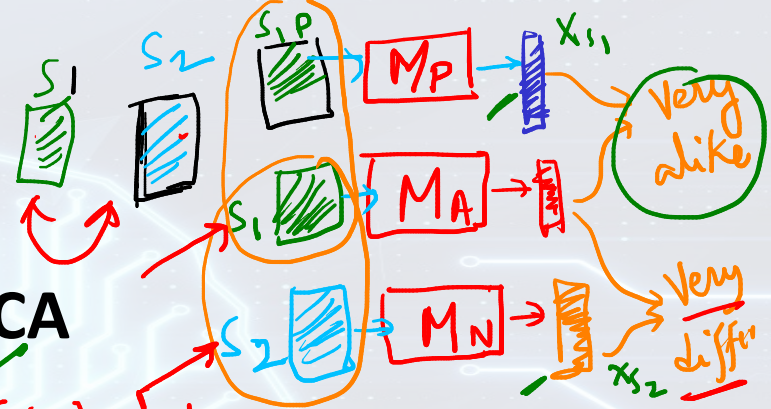



1. Python coding on Autoencoder, PCA

2. Solving numerical problems from week 8 content



# Content of the live session

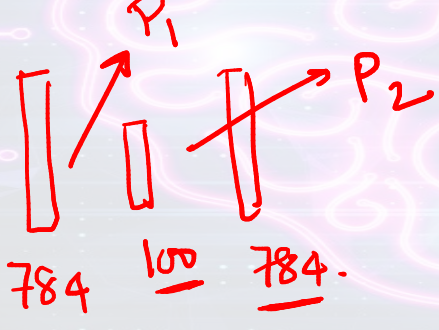



1. Python coding on Autoencoder, PCA

AsthmaSCELNet

Triplet Loss.

2. Solving numerical problems from week 8 content




$P_1 = (784 \times 100) + 100$

$= 78400 + 100$

$= 78500$


$P_2 = (784 \times 100) + 784$

$= 78400 + 784 = 79184$

An RGB input image has been converted into a matrix of size 257 X 257 X 3 and a kernel/filter of size 7 X 7 X 3 with a stride of 2 and padding = 3 is used for 2D convolution. What will be the size of the output of convolution?

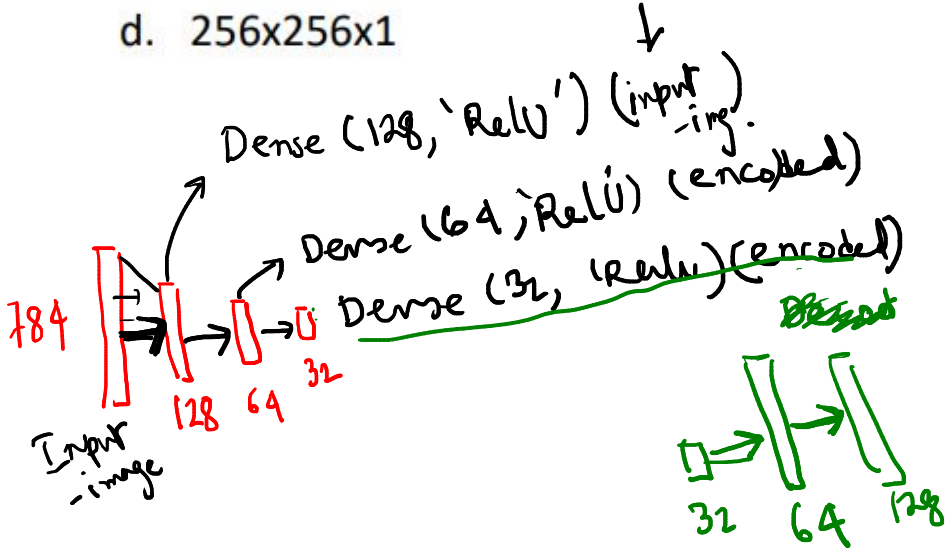
a. 129x129x1
b. 128x128x1
c. 254x254x3
d. 256x256x1

$$O/P = \frac{I/P - Kernel + 2padding}{Stride} + 1$$

$$= \frac{257 - 7 + 6}{2} + 1$$

$$= \frac{256}{2} + 1 = 129$$

129 x 129 x 1



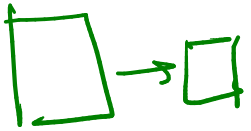
Primary reason for adding pooling layers is?

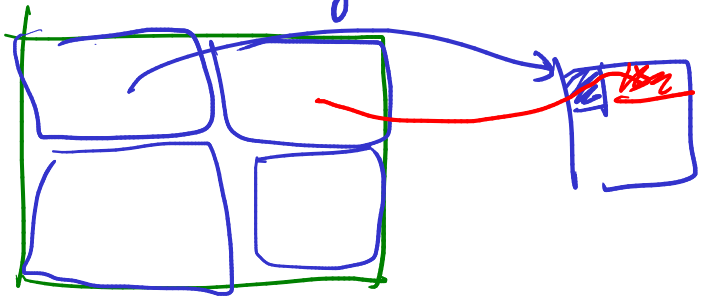
a. Promote small shift invariance

b. Reduce computations for subsequent layers

c. To produce activations that summarize filter response in local windows.

d. Both b and c

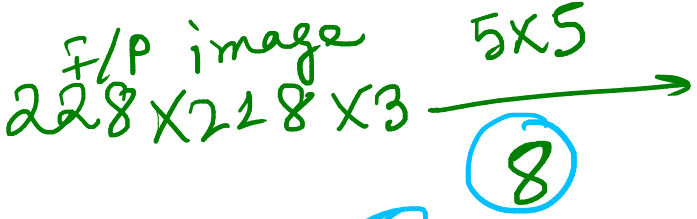





Blue Blue

Suppose you have 8 convolutional kernel of size 5 x 5 with no padding and stride 1 in the first layer of a convolutional neural network. You pass an input of dimension 228 x 228 x 3 through this layer. What are the dimensions of the data which the next layer will receive?

a. 224 x 224 x 3
b. 224 x 224 x 8
c. 226 x 226 x 8
d. 225 x 225 x 3




O/P channel is = no of filters applied

S = 1
P = 0

$$\frac{228 - 5 + (2 \times 0)}{1} + 1$$

$$= 223 + 1$$

$$= 224$$

224 X 224 X 8

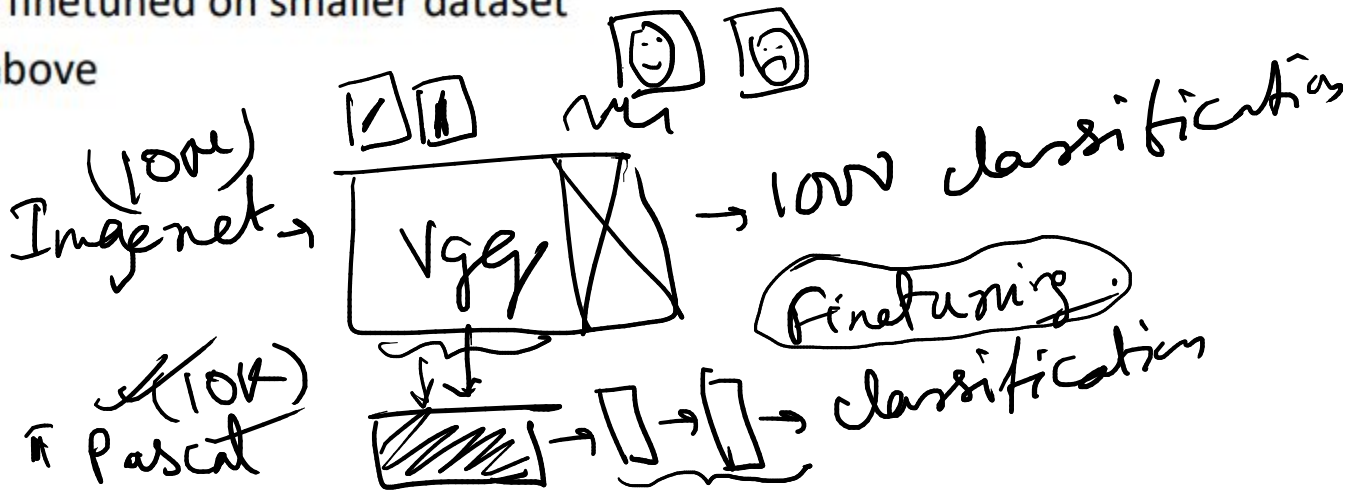
Choose the correct statement in context of transfer learning

a. Higher layers learn task specific features, where as lower layers learn general features

b. Transfer learning is generally used in situations when task specific dataset is very less

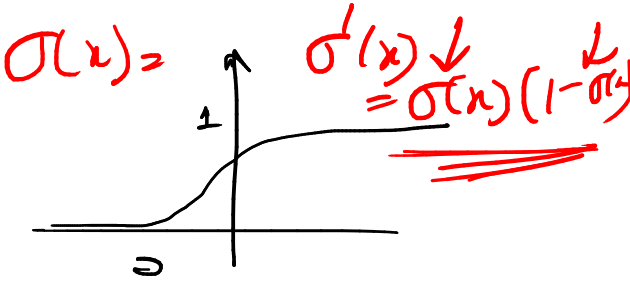
c. The weights of lower layers of pretrained CNN (trained on dataset like ImageNet etc) are copied and higher layers are random/gaussian initialized and entire network is finetuned on smaller dataset

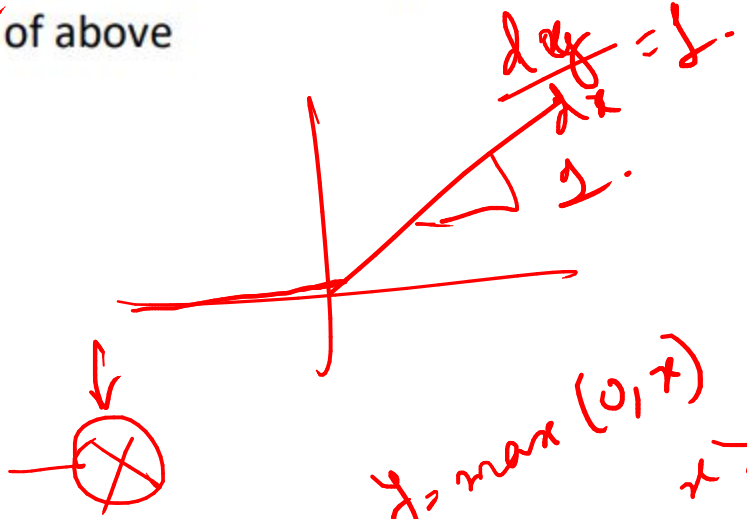
d. All of the above

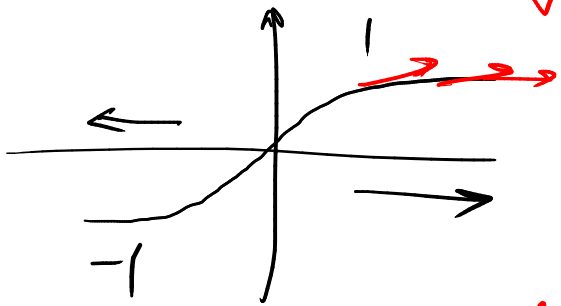


Advantage of ReLU over Sigmoid and TanH is

a. Low computational requirements
b. Alleviates vanishing gradient to some extent
c. Backpropagation is simpler
d. All of above




∇ → smooth
very
less
variation.
↓
Gradient
vanishing

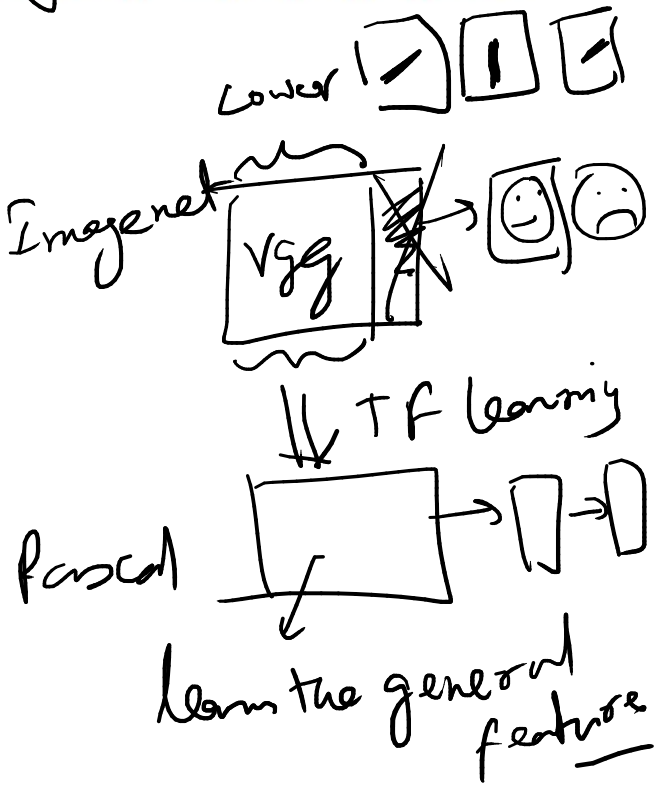
y = max(0, x)
x > 0
y = x
dy/dx = 1

Statement 1: For a transfer learning task, lower layers are more generally transferred to another task

Statement 2: For a transfer learning task, last few layers are more generally transferred to another task

Which of the following option is correct?

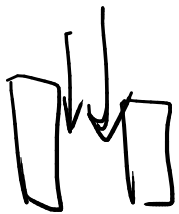
a. Statement 1 is correct and Statement 2 is incorrect
b. Statement 1 is incorrect and Statement 2 is correct
c. Both Statement 1 and Statement 2 are correct
d. Both Statement 1 and Statement 2 are incorrect



Statement 1: Adding more hidden layers will solve the vanishing gradient problem for a 2-layer neural network

Statement 2: Making the network deeper will increase the chance of vanishing gradients.

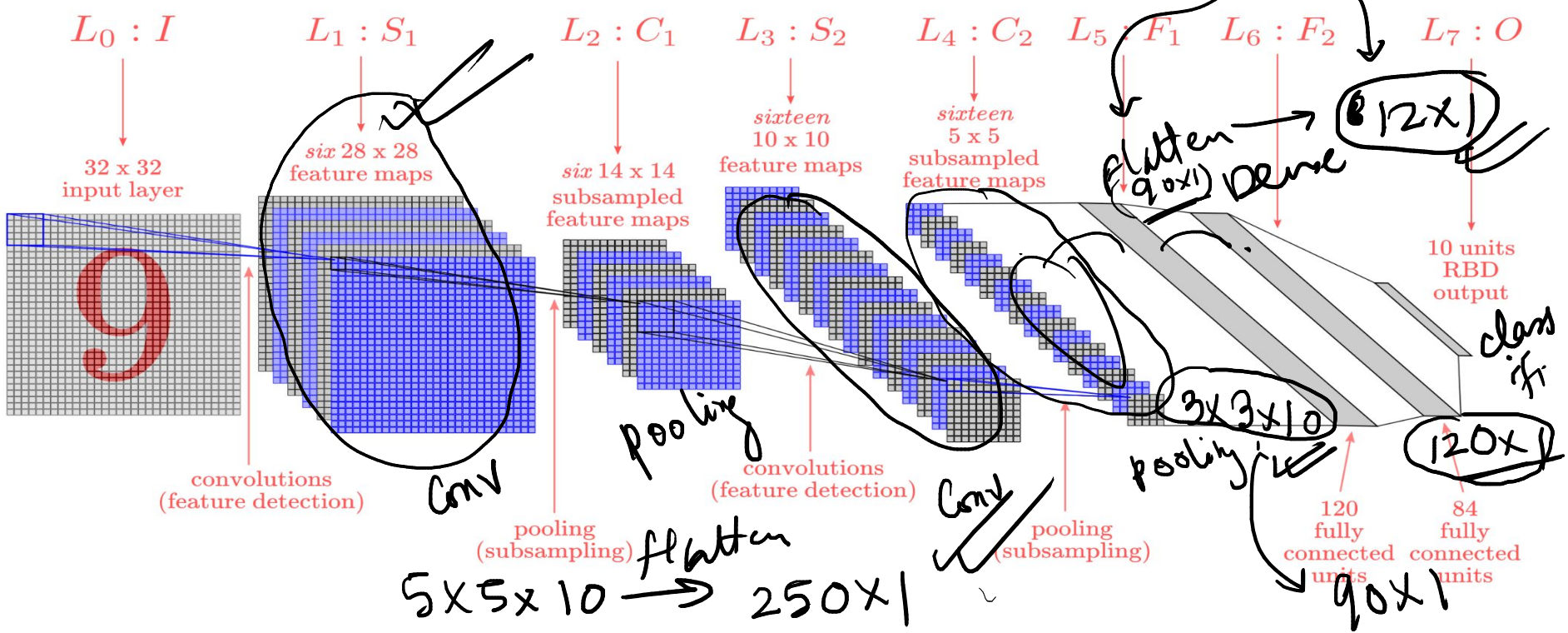
a. Statement 1 is correct
b. Statement 2 is correct
c. Neither Statement 1 nor Statement 2 is correct
d. Vanishing gradient problem is independent of number of hidden layers of the neural network.



$\sigma(n)(1-\sigma(n))$

How many convolution layers are there in a LeNet-5 architecture?

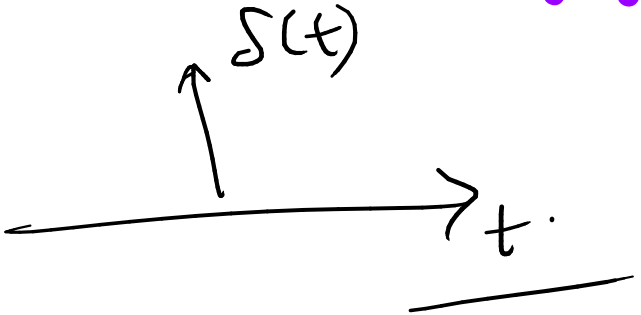
a. 2
b. 3
c. 4



delta dirac

$\delta(n) = \begin{cases} 1, & n = 0 \\ 0, & \text{elsewhere.} \end{cases}$


# Thank You




h(n)