NPTEL Week 9 Live Sessions

on Deep Learning (noc24_ee04)

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

- ✓Quiz 7, ✓Quiz 8 Solution
- VGG16, ResNet implementation

Powered by:

PMRF
Prime Minister's Research Fellows
Ministry of Education
Government of India

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

Content of the live session

1. Solution of Week-7, Week-8 quiz
2. Python coding on VGG-16, ResNet18
3. Solving numerical problems from week 9 content

Week-7

1 What is the main advantage of layer-by-layer pre-training for deep autoencoders?

$y = Wx + b = [W_1 W_2 \cdots W_n]\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix} + b$

$= [W_1 W_2 \cdots W_n\ b]\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\\1\end{bmatrix}$

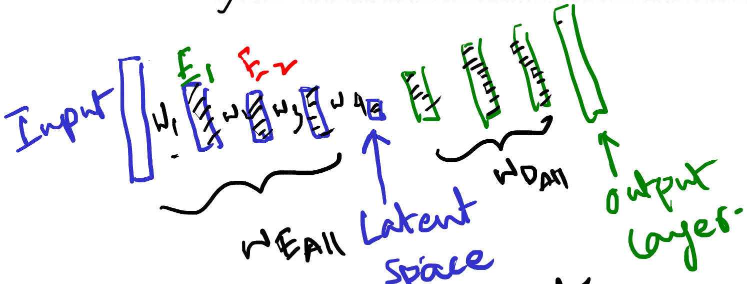
Stacked Autoencoders

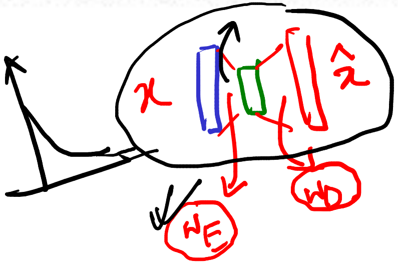
a. It reduces the total number of weights and simplifies the optimization process

b. It provides better initial weight values for the entire network ✓

c. It allows for parallel training of different hidden layers.

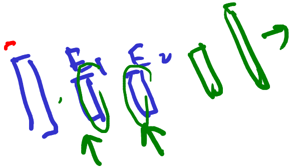
d. It guarantees perfect reconstruction of the input data with minimal error.





$\mathcal{L} = MSE(x, \hat{x})$

$\arg\min_{W_E, W_D} \mathcal{L}(x, \hat{x})$



$W(0) \rightarrow rand$

These pretrained weights you can use as the first initialized weight values rather than intializing them using a rand(·) function.

$W(n+1) \leftarrow W(n) - \eta \frac{\partial E}{\partial W}$.

② Select the correct option about Denoising autoencoders?

✓ Statement A: The loss is between the original input and the reconstruction from a noisy version of the input

✓ Statement B: Denoising autoencoders can be used as a tool for feature extraction.

a. Both the statements are false

b. Statement A is false but Statement B is true

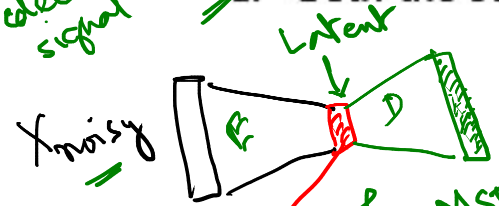
c. Statement A is true but Statement B is false

[✓ d. Both the statements are true] (Correct Option)

$x + \eta \approx x_{noisy}$

↑ clean signal.

↑ Noise



$\hat{x}$ (denoised output).

$\mathcal{L} = MSE(x, \hat{x})$

Hidden features: representation from the input noisy data.

5) Which of the following is a linear operator? → Superposition Theorem.

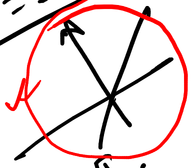
a. Sigmoid

b. Rectified Linear Unit

[✓ c. Convolution function]

d. None of the above

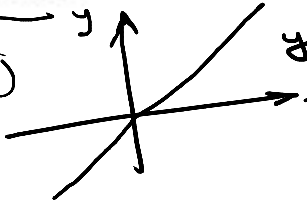
$V = IR.$



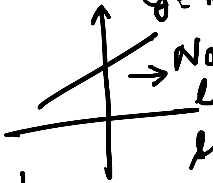
$\phi = LI.$

$Q = CV.$

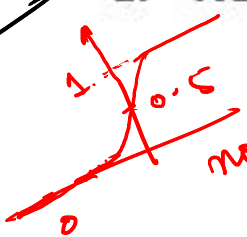
1)



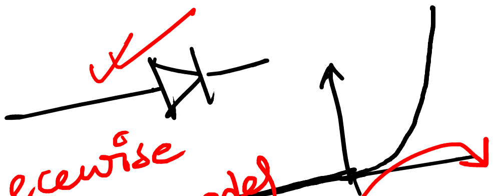
$y = mx$ → linear system.



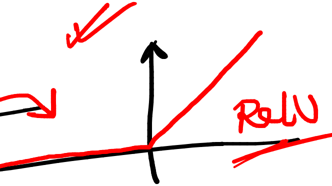
$y = mx + c.$ → Non linear system.



non linear operator.



Piecewise linear model of diode.



non linear in nature

$\mathrm{Relu}(x) = \max(0, x).$

Non linear.

$$x(n) \to \boxed{h(n)} \to y(n)$$

$$y(n) = h(n) * x(n) = x(n) * h(n).$$

$$x_1(n) \to \boxed{h(n)} \to y_1(n) = x_1(n) * h(n)$$

$$x_2(n) \to \boxed{h(n)} \to y_2(n) = x_2(n) * h(n)$$

$$y^s_{total} = y_1(n) + y_2(n) = x_1(n) * h(n) + x_2(n) * h(n).$$

$$x_1(n) + x_2(n) \to \boxed{h(n)} \to y_0(n)$$

Conv → linear op.

$$y_0(n) = (x_1(n) + x_2(n)) * h(n).$$

$$y_0(n) = x_1(n) * h(n) + x_2(n) * h(n)$$

Super position theorem obeyed $y_0(n) = y^s_{total}(n)$

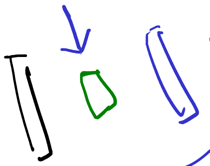
Which statement is TRUE about deep autoencoders?

a. They have a single hidden layer for representing the latent space.

b. They have few parameters compared to shallow autoencoders.

[✓ c.] They excel at capturing complex relationships and features ~~in~~ of high-dimensional data.

d. They require supervised learning with labeled data for training.



No weight ↓

Shallow AE.

$\mathcal{L} = MSE(X, \hat{X})$

encoder layer.

decoder

deep AE.

No of weight ↑

Unsupervised learning.

Dont need label.

Which of the following is false about autoencoder?

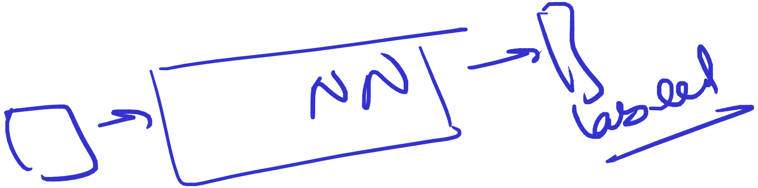
Anomaly detection
Image segmentation
Feature extraction.

a. Autoencoders possesses generalization capabilities True.

[✓ b.] Autoencoders are best suited for image captioning task X False

c. Its objective is to minimize the reconstruction loss so that output is similar to input $MSE(x, \hat{x})$. True

d. It compresses the input into a latent space representation and then reconstruct the output from it True



The Dirac delta function is $\underline{\infty}$ when t=0. Fill in the blanks.

a. ~~1~~

b. 0

[✓ c. Infinity]

d. None of the above

→ Continious time signal. t ∈ Real.

$\delta(t)$

→ Discrete time system. $\delta(n)$.

$t \to n$ ↑ sample.

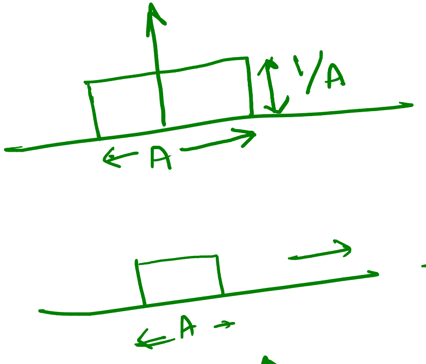
$n \in$ integer

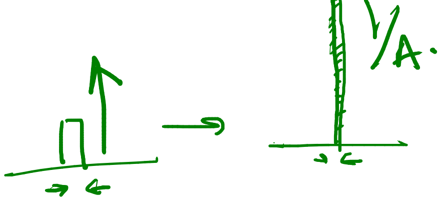
$\delta(n) = 1 ; n=0$

$= 0 ; n \neq 0$.

unit impulse function. elsewhere

Area = $A \times \frac{1}{A} = 1$



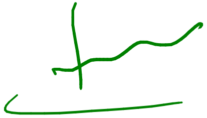


A → 0.

$A \to 0$.

$\frac{1}{A} \to \infty$.

$A\downarrow \quad \frac{1}{A}\uparrow$



① $\delta(t)$ exists at $t=0$ only

② $\delta(t) = 0 ; t \neq 0$.

→ Area = $\lim_{A\to 0} \left(A \times \frac{1}{A}\right) = 1$.

③ Amplitude of $\delta(t)$ at $t=0 ; \lim_{A\to 0} \frac{1}{A} \to \infty$

$\delta(t) \to \infty$.

④ $\int_{-\infty}^{\infty} \delta(t)\, dt = 1$.

Impulse response is the output of ________________ system due to impulse input applied at time=0. Fill in the blanks from the options below.

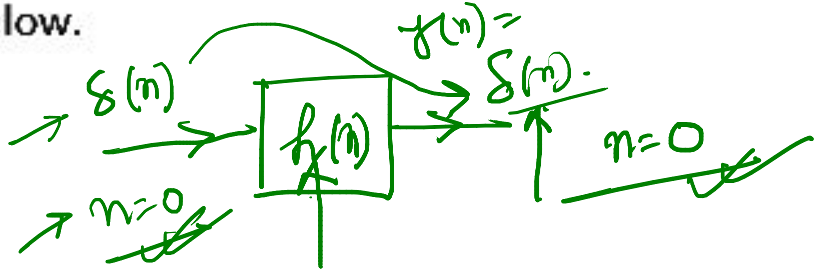
Sample (n)=0

a. Linear

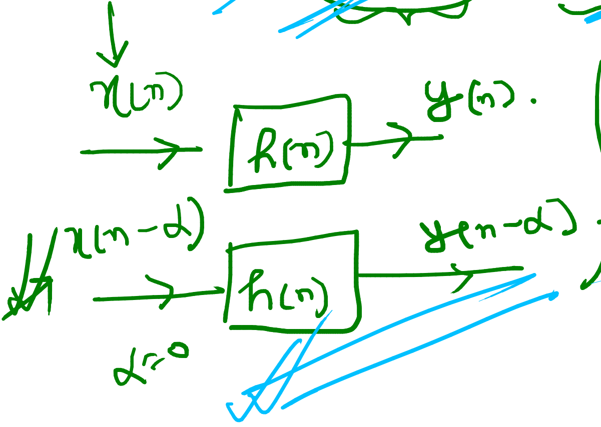
b. Time Varying

c. Time Invariant

[✓ d. Linear And Time Invariant]

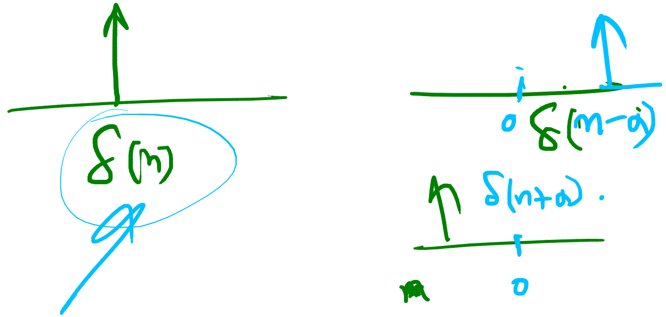


y(n) = δ(n) .

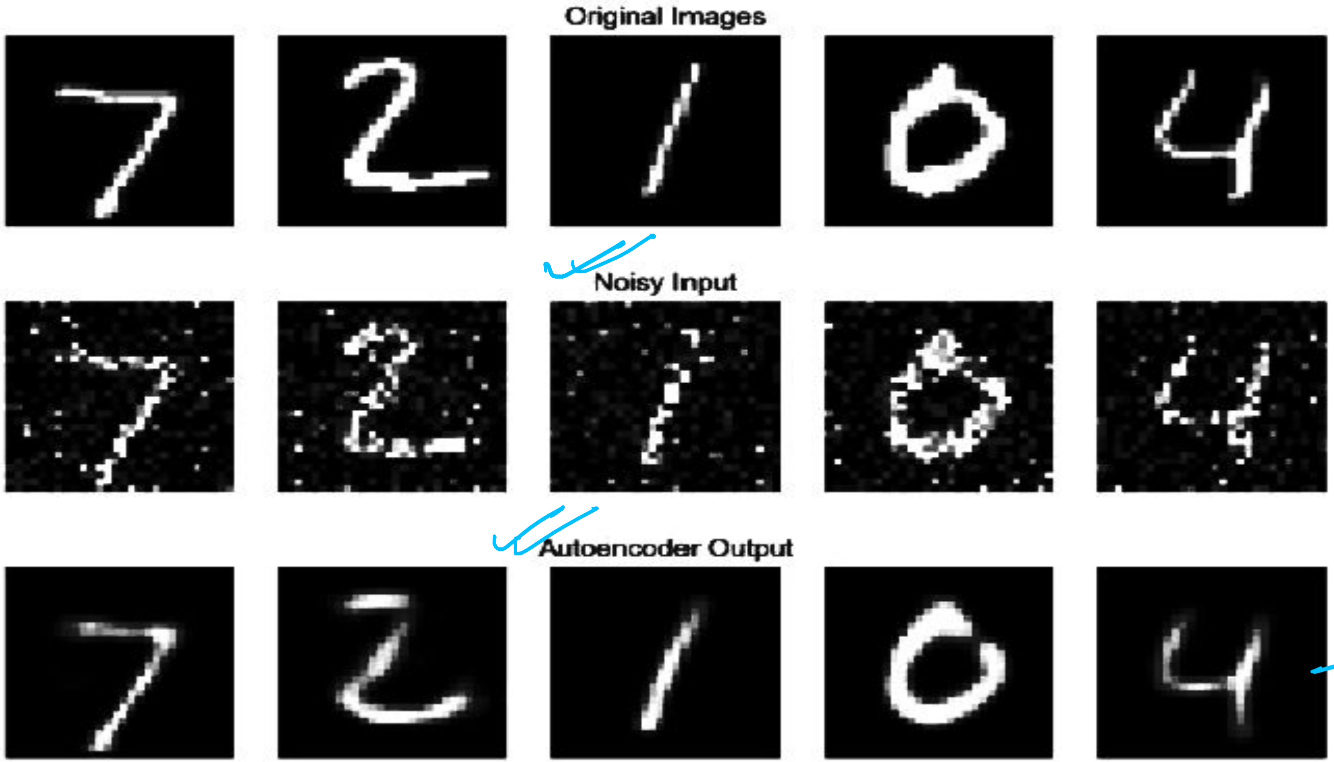


Time invariant system

δ(n) * δ(n) = δ(n).



Given the image below where, Row 1: Original Input, Row 2: Noisy input, Row 3: Reconstructed output. Choose one of the following variants of autoencoder that is most suited to get Row 3 from Row 2.



→ Denoising AE.

a. Stacked autoencoder

b. Sparse autoencoder

[✓ c. Denoising autoencoder]

d. None of the above

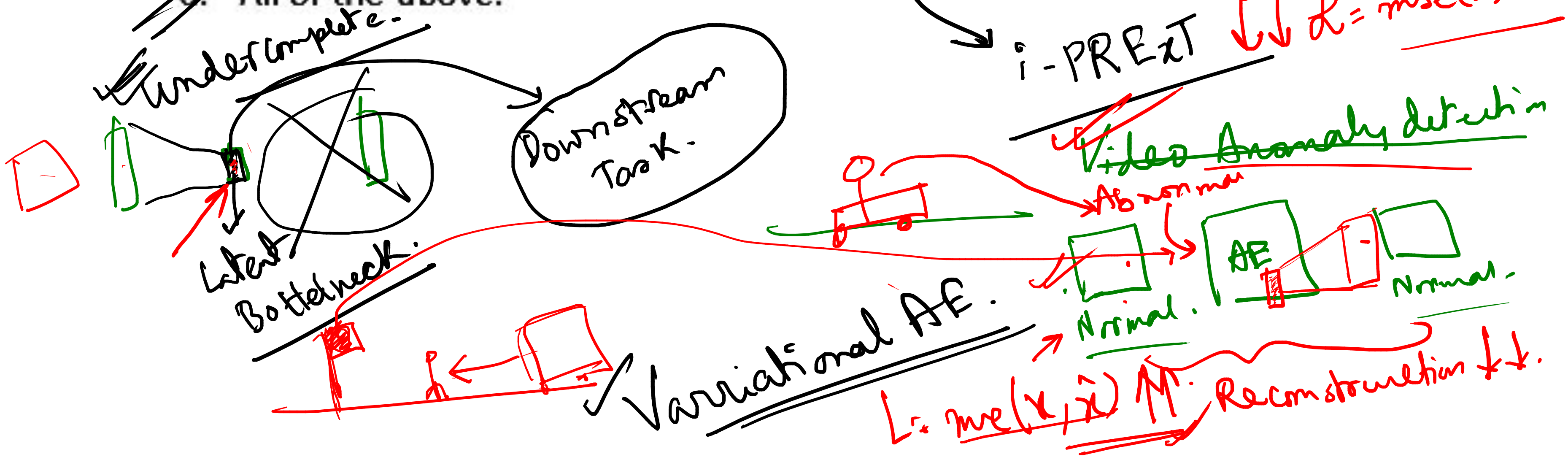
Which application of deep autoencoders utilizes the extracted latent space representation?

a. Anomaly detection by comparing data points to the known latent space distribution.

b. Image segmentation by classifying individual pixels based on their latent features.

c. Pre-training for other deep learning models like convolutional neural networks.

d. All of the above.



Week-8

① The input image has been converted into a matrix of size 256x256x3 and 4 kernel/filter of size 7x7 with a stride of 2 and no padding. What will be the size of the convoluted matrix?

a. 127x127x3

b. 128x128x4

c. 124x124x3

[✓ d. 125x125x4]

I/P = 256 x 256 x3.

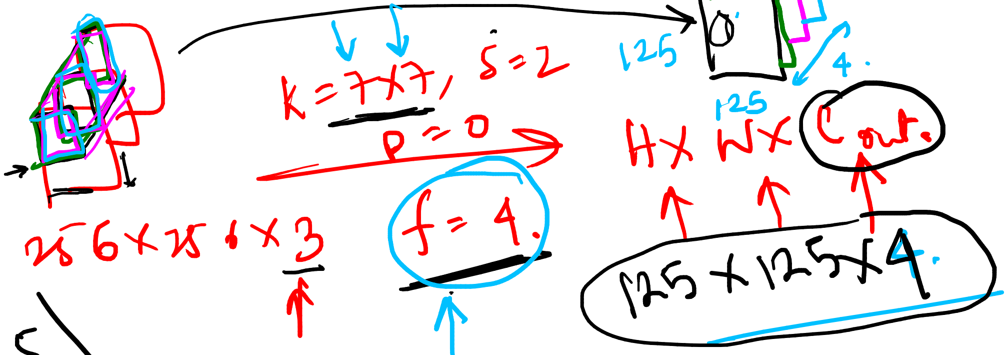
f = 4, K = 7x7, S = 2, P = 0.

$$O/P = \left\lfloor \frac{I/P - K + 2P}{S} + 1 \right\rfloor$$

$$= \left\lfloor \frac{256-7}{2} + 1 \right\rfloor$$

$$O/P = \left\lfloor \frac{249}{2} + 1 \right\rfloor = \lfloor 125.5 \rfloor$$

$$= 125.$$

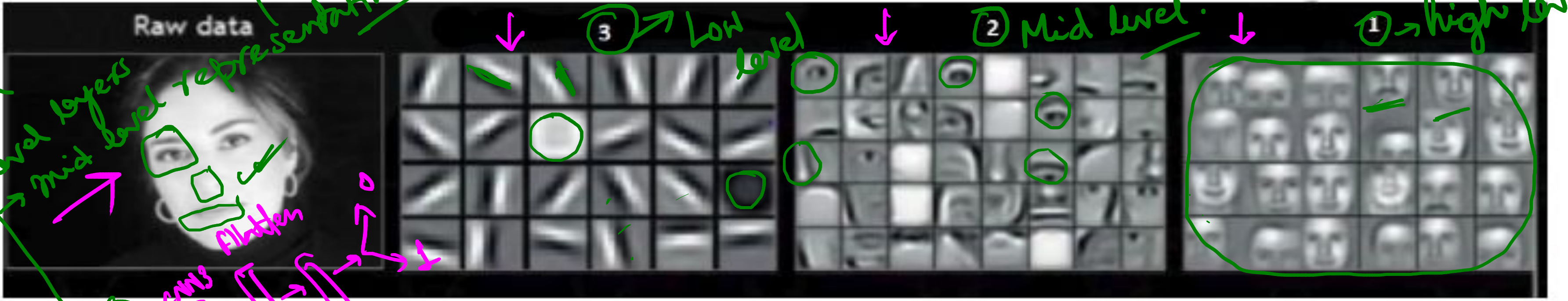


Cin

no of filter = Cout

eye, nose, lips.→ ear,

The figure below shows image of a face which is input to a convolutional neural net and the other three images shows different levels of features extracted from the network. Can you identify from the following options which one is correct?



→ Low level

Mid level.

→ high level.

mid level layers

mid level representation

flatten

conv1

conv2

conv3

FC

a. Label 3: Low-level features, Label 2: High-level features, Label 1: Mid-level features

higher features

high level feature → generating face (Task specific feature)

Lower

b. Label 1: Low-level features, Label 3: High-level features, Label 2: Mid-level features

c. Label 2: Low-level features, Label 1: High-level features, Label 3: Mid-level features

[✓ d. Label 3: Low-level features, Label 1: High-level features, Label 2: Mid-level features]

Low level representation

↳ Line, edge, boundary.

Which of the following statement is False about ReLU layer?

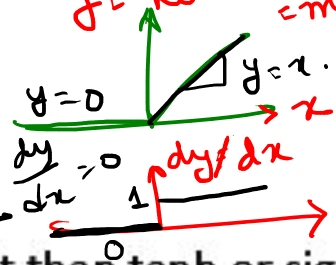
a. ReLU has expression f(x)=max(0,x) ✓ True .

b. The derivative of ReLU is 1 if x>0; o otherwise True .

False → c. Implementation of ReLU has more computational cost than tanh or sigmoid

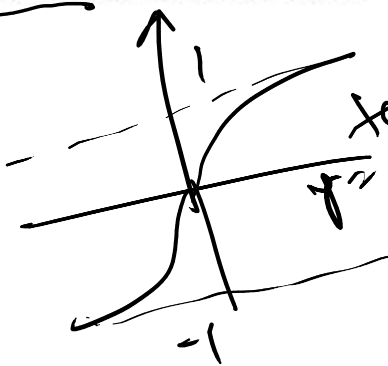
d. ReLU activation function was introduced in AlexNet architecture.

$y = \text{Relu}(x) = \max(0,x)$



sigmoid(x)

$$\sigma'(x) = \sigma(x)\,(1 - \sigma(x))$$



tanh(x)??

$\frac{dy}{dx} = ?$

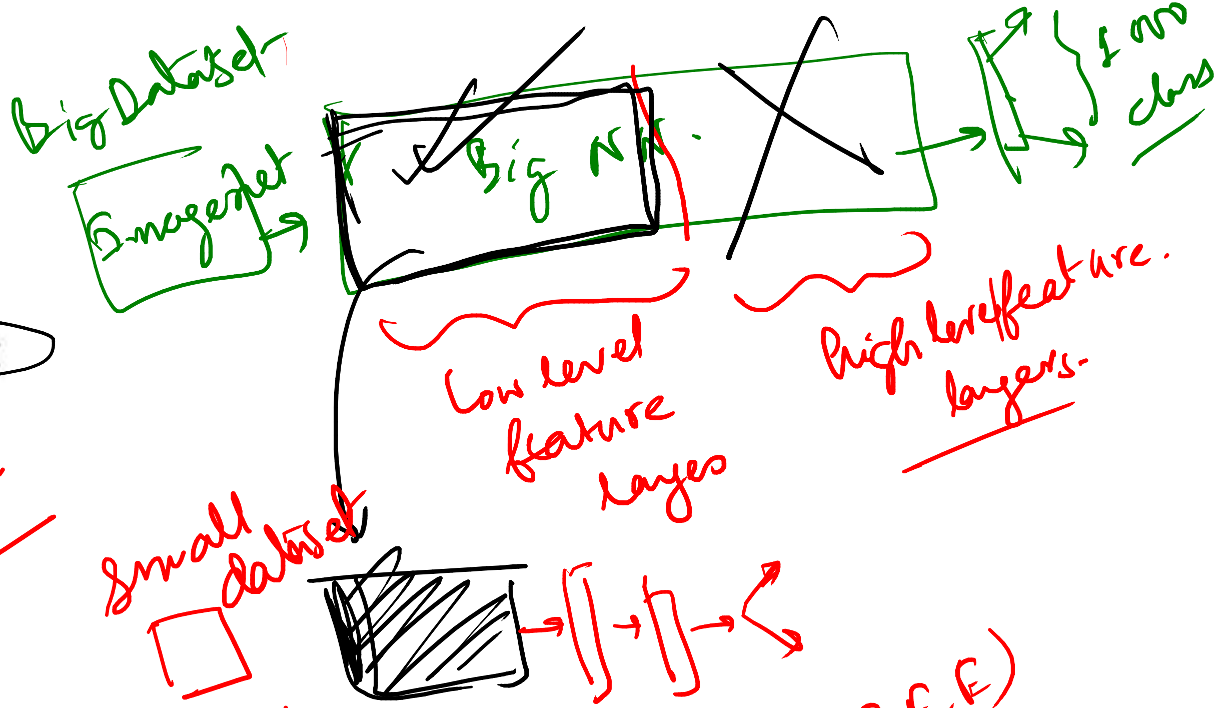
For a transfer learning task, which layers according to you can be more generally transferred to another task?

a. Higher layers

b. Lower layers

c. Task specific

d. ~~Cannot comment~~



Audo pretrained model → finetuning.

Imagenet

→ PASCAL

→ CIFAR 10

mini

→ chest XRay (EPFE)

Suppose your input is a 256 by 256 color (RGB) image, and you use a convolutional layer with 100 filters that are each 5x5. How many parameters does this hidden layer have (**with bias**)

$y = Wx + b.$

3 such weight

a. 2501

b. 2600

c. 7500

d. 7600

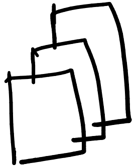
$y = W_1 x + b_1$

$= W_2 x + b_2$

$= W_3 x + b_3$

1st filter

5x5x3.



256x256x3

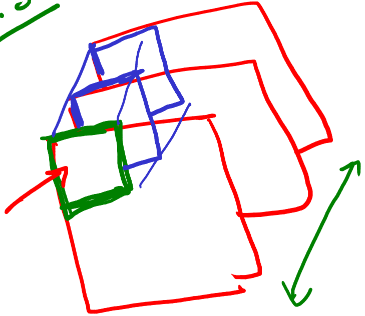
$\xrightarrow[\text{Bias.}]{f = 100,\ k = 5\times5.}$

Total no of parameters.

Parameter.

⇒ 5x5x3 → single Kernel.

* 100 such Kernel → (5x5x3)x100.



I/P.

256x256x3.

for 1 Kernel → 1 bias

for 100 Kernel → 100 bias

weight value.

Total parameter = total weight + Total Bias.

=

Suppose your input is a 256 by 256 color (RGB) image, and you use a convolutional layer with 100 filters that are each 5x5. How many parameters does this hidden layer have (**with bias**)

a. 2501

b. 2600

c. 7500

[✓ d. 7600]

Total weight + Total Bias

parameter

= Total ~~total~~

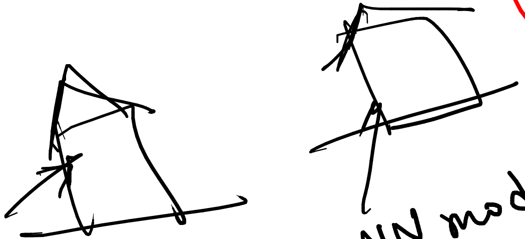
$(5 \times 5 \times 3) \times 100 + 100$

$(5 \times 5 \times 3)a + a.$

$a[(5 \times 5 \times 3) + 1].$

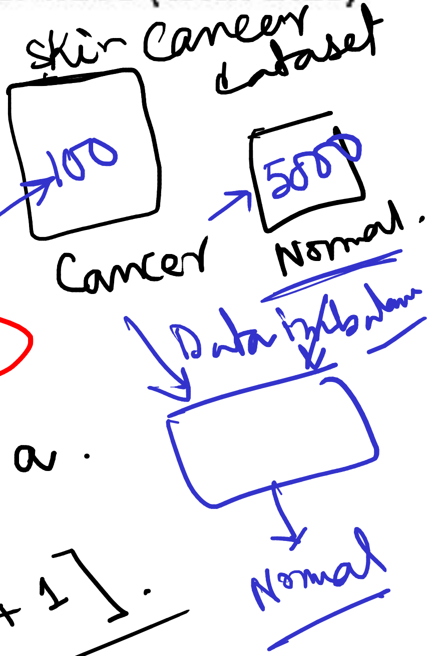
$a = 100.$

$= 7600$



CNN model is biased on Tajmahal =

→ It has more inclination towards Tajmahal



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.

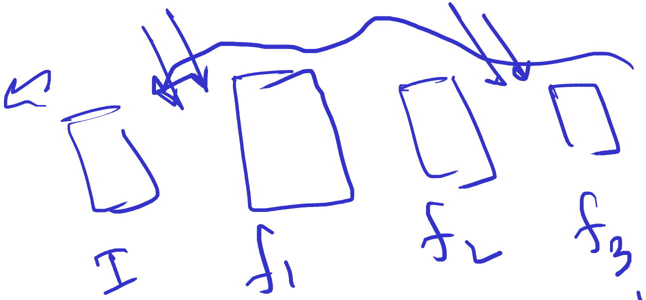
a. Statement 1 is correct

[✓ 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.

✓ $O = f_1 f_2 f_3 f_4 f_5 f_6 (I)$.



$f_4 \quad f_5 \quad f_6 . O$

$$\frac{\partial O}{\partial I} =$$

chain

chain

$\frac{1}{4}(1-\frac{1}{4})$ $\quad \sigma(n)(1-\sigma(n))$

o

$w(n+1) \Leftarrow w(n) - \eta \frac{\partial E}{\partial w}$.

Which of the following is false about CNN?

True .

a. Output should be flattened before feeding it to a fully connected lyer

False b. There can be only 1 fully connected layer in CNN wRong .

c. We can use as many convolutional layers in CNN

True .

d. None of the above

Vgg-16 . (13)

Resnet 18 (18)

50

101

Densenet 121 ,

Let us consider a Convolutional Neural Network having three different convolutional layers in its architecture as:

$$39\times39\times3 \xrightarrow[s=1,\ p=0]{3\times3,\ f=10} \underset{①}{37\times37\times10} \rightarrow \underset{②}{17\times17\times20} \rightarrow \underset{③}{\boxed{7\times7\times40}}$$

↳ Tensor (Multi dementional array)

↓

Flattening operation

↓

Fully connected layer — Dense

Vector: M x 1 = $\boxed{1960\times1}$

Layer-1: Filter Size – 3 X 3, Number of Filters – 10, Stride – 1, Padding – 0

Layer-2: Filter Size – 5 X 5, Number of Filters – 20, Stride – 2, Padding – 0

Layer-3: Filter Size – 5 X5 , Number of Filters – 40, Stride – 2, Padding – 0

Layer 3 of the above network is followed by a fully connected layer. If we give a 3-D image input of dimension 39 X 39 to the network, then which of the following is the input dimension of the fully connected layer.

[✓ a.] 1960

b. 2200

c. 4563

d. 13690

$$\frac{39-3+0}{1}+1 = 37 \qquad \Bigg| \qquad \frac{37-5+0}{2}+1 \quad 17 \qquad \Bigg| \qquad \frac{17-5+0}{2}+1$$

$$M = 7\times7\times40 = 49\times40 = 1960$$

Consider a CNN model which aims at classifying an image as either a rose, a marigold, a lily or orchid (consider the test image can have only 1 of the images at a time) . The last (fully-connected) layer of the CNN outputs a vector of logits, L, that is passed through a ✓ _____ activation that transforms the logits into probabilities, P. These probabilities are the model predictions for each of the 4 classes.

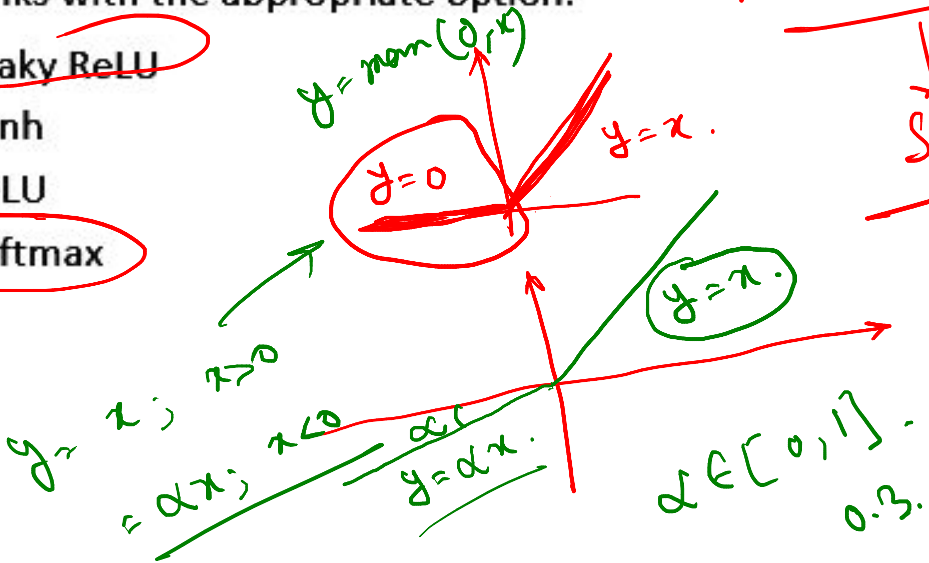
Fill in the blanks with the appropriate option.

a. Leaky ReLU

b. Tanh

c. ReLU

[✓ d. Softmax]



Multi class classification.

↓

Softmax → Probability of occurrence for each class.

10 Imagine you're training a CNN for Autonomous driving vehicle to distinguish between pedestrian, bicycle, bike and cars in images. You have two options:

Option 1: A shallow CNN with just one convolutional layer having 1024 kernels and a few fully connected layers.

Option 2: A deeper CNN with 4 convolutional layers having 128 kernels in first layer,128 kernels in second layer, 256 kernels in third layer and 512 kernels in fourth layer and then fully connected layers.

Both options use the same total number of kernels convolutional layer (=1024)

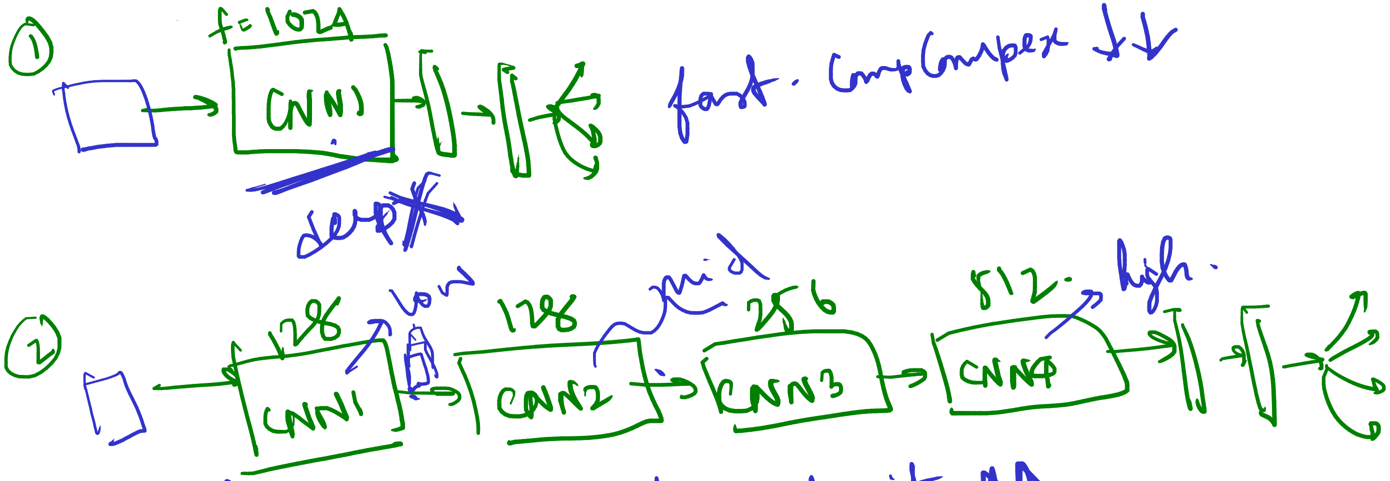
Choose the incorrect statement:

a. Option 1 will give higher inference speed since it can operate convolutions in parallel whereas option 2 can't be parallelized as results of next convolutional layers are dependent on past layers } True.

b. Option 2 , Deeper CNN, with multiple layers, can perform hierarchical feature extraction, thereby has higher representational power and accuracy } True.

[✓ c.] Option 1 , Shallow CNN, with one convolutional, can extract 1024 features from image, thereby has higher representational power and accuracy } False.

d. Option 2 , Deeper CNN, with multiple layers, can extract more abstract features that depend on features of shallower layers and therefore has higher representational power and accuracy } True



fast. CompComplex ↓↓

→ high computational complexity ↑↑

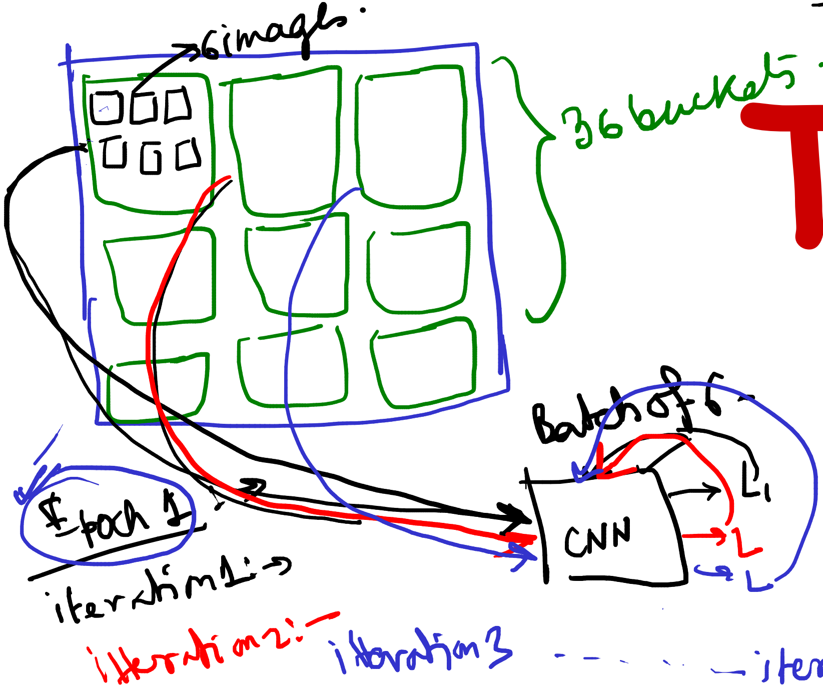
→ wide variety of feature

low
mid
high.

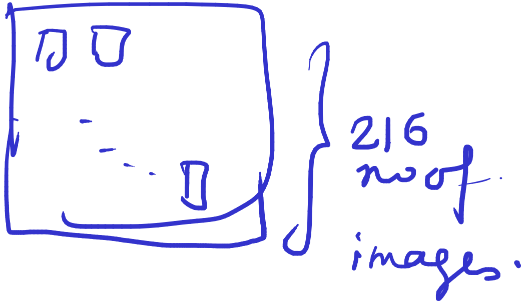
Dataset :→ 216. (sample/Images) .

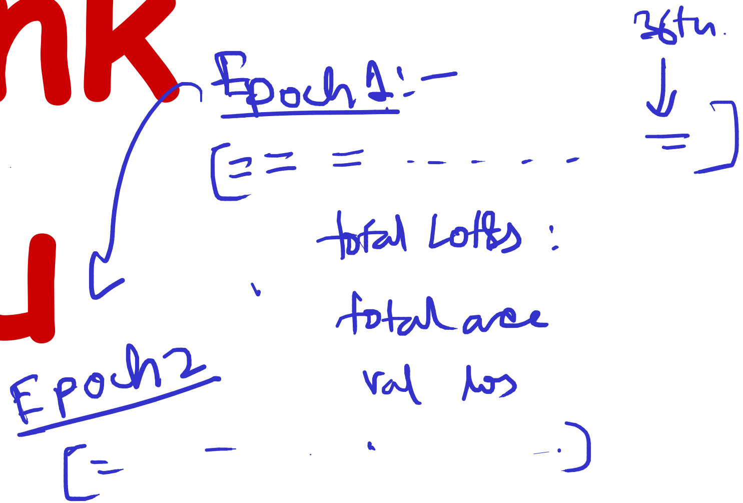
Batch Size = 6

Six epochs



Thank You




Epoch 1:—
36th
[= = = ... =]
total Loss :
total acc
val loss
Epoch 2
[= — . . .]