Ques No 1

What is the need of data standardization
What are the main steps you will
go through.

Data standardization is the process of standardizing the data. The data on which a model is to be applied upon many be very dimensional. There may be variability in the dataset. The data may have extreme values. These extreme values cause biasness for the model and results in bad accuracy. In order to avoid it, the data is standardized. This process involves subtracting the mean from the data points and subtracting it with standard deviation. This causes the mean between data points to be '0' and the standard deviation to be '1'

How to overcome overfitting in CNN:

Overfitting occurs when model is performing good on traing training set but generalizing badly on testing set. It can be avoided in the following ways.

L1 Regularization:-

L1 regulizer also known as lasso regulizer adds a regularization term to the new weight of the model. It generally makes most of the weights 0 hence making the model less complex

- $\theta_n - \frac{\lambda}{2m} ||w||$

L2 Regulizers:

Also known as Ridge regulizer, it adds a regulizer term which promotes weights close to zero but not actually zero

- $w_{old} - \frac{\lambda}{2m} |w|^2$

Loss Functions:
Loss Functions are used to find the accuracy of the model or to find out how good a model is performing. Examples of loss function are MSE, MAE, RMSE, Hinge Loss, etc.

10

Optimization Algorithm:
Optimization Algorithm optimize the trainable parameters of a model. Most common optimizer are Adam, gradient descent, Adagrad, Adadelta, Momentum, etc.

Hyper Parameters:
Hyperparameters are defined before the ~~traning~~ training starts. They remain constant throughout the training. Learning rate, decay rate, early stopping are hyperparameters.

QUES No 3

Scanned with CamScanner

Standardization : $\sum_{i=1}^{n} \frac{x_i - \text{mean}}{\text{standard Deviation}}$

Ques No 2

Parameters:-

Parameters are the values in a model which can be trained. These are the internal values of the models. They are updated throughout the training process for better results. Weights are an example of parameters.

Activation Function:-

Activation function is applied on the weighted sum of a neuron. It is applied to bring the weighted sum to a certain degree or domain.

Relu : $[0, \infty)$

softmax : $[0, 1]$

sigmoid : $(0, 1)$ or $(-1, 1)$

Data Augmentations:- Data Augmentation involves augmenting the data like shear, rotate, zoom in and out. It makes the data more versatile.

Dropout:- Adding dropout layer also reduces biasness. Dropout neuron will drop a random neuron and its connection based on some probability whereas dropout connection will drop a certain weight based on probability.

Early Stopping:- Early stopping involves stopping the process of training based on a predefined number of epochs when the testing accuracy started dropping.

Other Methods:-

i) Increasing data size
ii) Decreasing complexity of model
iii) Using batch normalization.

Ques No 4

1) Forward Pass:-

20

	Act. Shape	Act. size	Trainable Para.
Input	(6,6)	36	0
Conv1	(4,4)	16	~~17~~ 10
Layer Activation	–	–	–
Pool	(2,2)	4	0
Flatten	(4,1)	4	0
Output	(1)	1	5

Conv1 (Seperable):-

$$\begin{bmatrix} 1 & 2 & 1 & 0 & 3 & 1 \\ 3 & 1 & 2 & 1 & 1 & 0 \\ 2 & 0 & 1 & 1 & 2 & 1 \\ 1 & 1 & 2 & 3 & 1 & 1 \\ 1 & 1 & 1 & 2 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 \end{bmatrix} * \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & -1 \end{bmatrix}$$

$C_{11} = 1+1-2-1 = -1$

$C_{12} = 2+0+0-1 = 1$

$C_{13} = 1+3-1-2 = 1$

$C_{14} = 0+1-1-1 = -1$

$C_{21} = 3+2-1-2 = 2$

$C_{22} = 1+1-1-3 = -2$

$C_{23} = 2+2\,1-2-1 = 0$

$C_{24} = 1+0-3-1 = -3$

$C_{31} = 2+1-1-1 = 1$

$C_{31} = 0+1-1-2 = -2$

$C_{33} = 1+2-1-1 = 1$

$C_{34} = 1+1-2-0 = 0$

$C_{41} = 1+2-1 = 1$

$C_{42} = 1+3-1 = 3$

$C_{43} = 2+1-1-1 = 1$

$C_{44} = 3+1-1 = 3$

Without Bias $\qquad\qquad\qquad\qquad$ With bias.

$$C = \begin{bmatrix} -1 & 1 & 1 & 1 \\ 2 & -2 & 0 & 3 \\ 1 & -2 & 1 & 0 \\ 1 & 3 & 1 & 3 \end{bmatrix} \qquad \begin{bmatrix} 0 & 2 & 2 & 2 \\ 3 & -1 & 1 & 4 \\ 2 & -1 & 2 & 1 \\ 2 & 4 & 2 & 4 \end{bmatrix}$$

Activation Layer :-

$$A = \begin{bmatrix} 0 & 2 & 2 & 2 \\ 3 & 0 & 0 & \cancel{3}4 \\ 2 & 0 & 2 & \cancel{0}1 \\ 2 & \cancel{3}4 & 2 & \cancel{3}4 \end{bmatrix}$$

Pool 1 :-

$$P = \begin{bmatrix} 3 & \cancel{3}4 \\ \cancel{3}4 & \cancel{3}4 \end{bmatrix}$$

Flatten Layer :-

$$D =$$
$$F = \begin{bmatrix} 3 \\ \cancel{3}4 \\ \cancel{3}4 \\ \cancel{3}4 \end{bmatrix}$$

Output :-

$= \sum Fw + b$

$= (3+4+4+4)+1$

$= 16$

sigmoid (16)

$= \frac{1}{1+e^{-16}}$

$= 0.99$

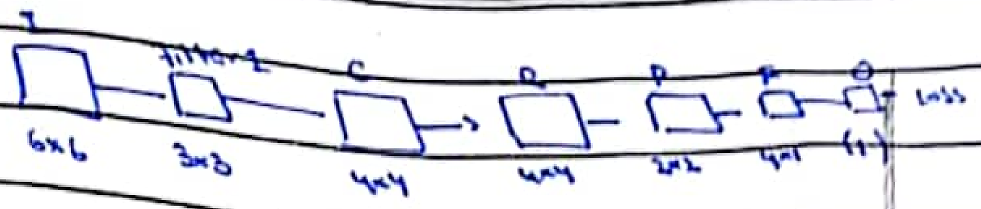
2) Calculate Cost/Error :-

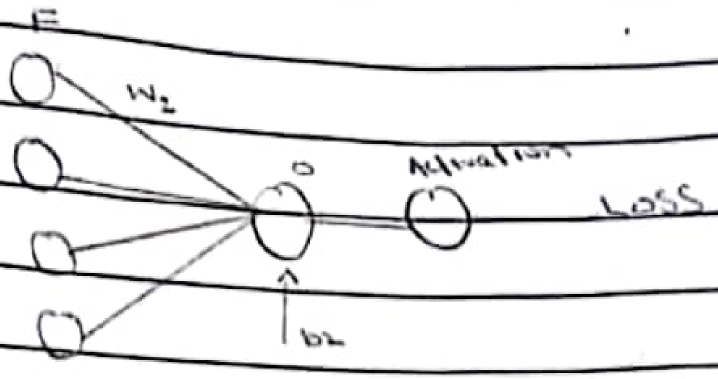
$\text{Error} = -y\log(\hat{y}) - (1-y)\log(1-\hat{y})$

$= -1\log(\overset{0.99}{1}) - (0)\log(1-0.99)$

$= -4.36 \times 10^{-3}$

3) Backpropage the error and update weights :-





$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial A} \times \frac{\partial A}{\partial O} \times \frac{\partial O}{\partial w_2}$$

$$\frac{\partial L}{\partial A} = \partial\left(-O\log(1-O) - (1-O)\log(1-\hat{O})\right)$$

$$= \frac{-O}{\hat{O}} + \frac{(1-O)}{(1-\hat{O})}$$

$$= \frac{-O(1-O) + (1-O)\hat{O}}{O(1-\hat{O})}$$

$$= -a - a^2 - a + a^2 \qquad \frac{-O - O^2 O - O}{O(1-O)}$$

$$= \frac{2}{O(2-O)}$$

$$\frac{\partial A}{\partial O} = \sigma(\partial O) \quad \because \text{Derivative of sigmoid}$$

$$= O(1-O)$$

$\frac{\partial x}{\partial w} = \sum c_i w_i + b$

where i ranges from 0,k

$\frac{\partial L}{\partial w_i} = c_i$

$\frac{\partial b}{\partial g_i} \quad \frac{\partial L}{\partial w_i} = 1$

Only convolution and output layers have trainable parameter.

$\frac{\partial L}{\partial w} = c_i \times \frac{0}{0(1-0)} \times (1-0)$

$= c_i \times 0$

convolution? OK

how?

$$W_1 = \begin{bmatrix} 0.9 & 0 & 0.9 \\ 0 & 0 & 0 \\ -0.9 & 0 & -0.9 \end{bmatrix}$$

4) Trainable parameters are same as we drawn before.

$c_1 = 0.9 + 0.9 - 0.18 - 0.9 = -0.9$

$c_2 = 0.18 + 0 + 0 - 0.9 = 0.9$

$c_3 = 0.9 + 0.27 - \cancel{0}.9 - 0.18 = 0$

$c_4 = 0 + 0.9 - 0.9 - 0.9 = -0.9$

Name ~~Sarmad Siddique~~ ~~FA21-BCS-000~~ Sheet No 19426

$= [1,1,1,1] - 0.1 [2.7, 3.6, 3.6, 3.6]$

$= [1,1,1,1] - [0.27, 0.36, 0.36, 0.36]$

$= [0.73, 0.64, 0.64, 0.64]$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial O} \times \frac{\partial O}{\partial F} \times \frac{\partial F}{\partial P} \times \frac{\partial P}{\partial R} \times \frac{\partial R}{\partial C} \times \frac{\partial C}{\partial w_1}$$

$$\frac{\partial O}{\partial F} = \text{Reshape}(2,2)$$

$$\frac{\partial F}{\partial P} = \text{upsample}(4,4)$$

$$= \begin{bmatrix} 0 & 0 & 0 & 0 \\ x_1 & 0 & 0 & x_2 \\ 0 & 0 & 0 & 0 \\ 0 & x_3 & 0 & x_4 \end{bmatrix}$$

$$\frac{\partial L}{\partial R} = \frac{\partial L}{\partial O} \times \frac{\partial O}{\partial F} \begin{bmatrix} 0 & 0 & 0 & 0 \\ x_1 & 0 & 0 & x_2 \\ 0 & 0 & 0 & 0 \\ 0 & x_2 & 0 & x_4 \end{bmatrix}$$

$$\begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x < 0 \end{cases}$$

$$\frac{\partial O}{\partial w_2} =$$

$$O = w_2 F + b$$

$$\frac{\partial O}{\partial w_2} = F$$

$$\frac{\partial O}{\partial b} = 1$$

$$\frac{\partial L}{\partial w_2} = \frac{2\hat{O}}{O(1-O)} \times O(1-O) \times F$$

$$= 2\hat{O}F$$

$$\frac{\partial L}{\partial w_2} = \frac{2\hat{O}}{O(1-O)} \times O(1-O) \times 1$$

$$= 2\hat{O}$$

Updating w_2..

$$= w_2 - \alpha \frac{\partial L}{\partial w_2}$$

$$\frac{\partial L}{\partial w_2} = [1,1,1,1] \times [3,4,4,4] \times [0.9]$$

$$= [2.7, 3.6, 3.6, 3.6]$$

$C_{21} = 0.27 + 0.18 - 0.9 - 0.18$

$= 0.18$

$C_{22} = 0.9 + 0.9 + -0.9 - 0.27 = 0.18$

$C_{23} = 0.18 + 0.9 - 0.9 - 0.9 = 0$

$C_{24} = 0.9 + 0 - 0.27 + 0 = -0.18$

$C_{31} = 0.9 + 0.18 - 0.9 - 0.9 = 0.18$

$C_{32} = 0 + 0.9 - 0.9 - 0.18 = -0.18$

$C_{33} = 0.9 + 0.18 + -0.9 - 0.9 = 0.9$

$C_{34} = 0.9 + 0.9 - 0.18 + 0 = 0$

$C_{41} = 0.9 + 0.18 + 0 - 0.9 = 0.9$

$C_{42} = 0.9 + 0.27 - 0.9 = 0.27$

$C_{43} = 0.18 + 0.9 - 0.9 - 0.9 = 0.9$

$C_{44} = 0.27 + 0.9 - 0.9 = 0.27$

$$= \begin{bmatrix} -0.9 & 0.9 & 0 & -0.9 \\ -0.18 & 0.18 & 0 & -0.18 \\ 0.18 & -0.18 & 0.9 & 0 \\ 0.9 & 0.27 & 0.9 & 0.27 \end{bmatrix}$$

$$\text{Relu} = \begin{bmatrix} 0 & 0.9 & 0 & 0 \\ 0 & 0.18 & 0 & 0 \\ 0.18 & 0 & 0.9 & 0 \\ 0.9 & 0.27 & 0.9 & 0.27 \end{bmatrix}$$

$$\text{Max} : \begin{bmatrix} 0.18 & 0 \\ 0.27 & 0.27 \end{bmatrix}$$

$= [0.17 \times 0.73 + 0 + 0.27 \times 0.64 \times 0.27] \times 0.67$

$= 0.134 + 0.477$

$= \dfrac{1}{1+e^{-0.477}}$

$= 0.61$

6) <u>Calculate Loss:-</u>

$= -(1)\log(1-0.61) - (1-0.61)\log(1-0.61)$

$= 0.21 - (-0.41)(0.39)$

$= 0.21 + 0.1599$

$= 0.3699$

7) <u>Compare Error:-</u>

The error before (-4.36×10^{-3}) was less than the current loss 0.3699.

8) <u>Predicted Outputs:-</u>

Predicted output on first iteration was 0.99 whereas on second iteration was 0.3699

for complete rep