Ans. to the Ques. No: 4

(a)

Here,

500 training cat images,

where

correctly detected as cat = 450

~~re~~ wrongly " as dog = 50

Again 500 training dog images,

where,

correctly detected as dog = 430

wrongly " as cat = ~~8~~70

(a) ∴ Training Accuracy $= \frac{450+430}{500+500}$

$= 0.88$

$= 88\%$

(b) for validation,

Validation Accuracy $= \frac{60+65}{100+100}$

$= 0.625$

$= 62.5\%$

(c)

The most likely problem with the trained classifier based on the figure is overfitting.

(d)

Two possible ways to fix this problem are given below:

1. <u>Regularization:</u>
In this solution, the cost function is ~~being~~ modified such way that weight or parameter's values do not be large and the values remain distributed.

To make this solution work, it will add a regularization term of the parameter which can be square, cubic or any term of the parameter.

regularization ~~would~~ term will have a constant part ~~of~~ ~~log~~ and using this constant ~~p~~ variable α, we can modify the regulation as well as cost value. One example of regularization term:

$$\text{regularization term} = \alpha \sum_{i=1}^{n} \theta_i^2$$

② Simplify the model:

~~Be~~ If the model is trained with less complex parameters, the prediction would be more simple and generalized. For example, if the ~~to~~ extra feature x^n, x^m of the model, are added in the model, it becomes ~~very~~ tough and less generalized. So, ~~by~~ not this additional parameter might cause overfitting.

Ans. to the Ques. NO: 3

(a) 3

(b) This is a regression problem. Because, the ~~target~~ / output (y) is a continuous variable. As continuous variable, this is a regression problem.

(c) We should perform normalization a a preprocessing step in this case because the ranges of the giving parameters are different to each other and because of this type of range differences the cost function would create an ellipse type ~~data~~ graph. For this reason, at one direction the graph might come highly skewed.

So, to make ~~the~~ all the variable values in one range without changing the relation betn the features, we ~~we~~ should perform normalization as a preprocessing step in this case.

(d) MSE should be used here as this ~~t~~ is a regression problem.

(e) hypothesis function would be:

$$x = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix}$$ here, x_0 = bias term = 1

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix}$$

So, ~~h~~ $h_\theta(x) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$

$\therefore h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$

This is the hypothesis function.

Cost function :

$$J_\theta(x) = \frac{1}{n} \sum_{i=1}^{n} J_i(\theta)$$

where,

$$J_i(\theta) = [h_\theta(x^i) - y^i]^2$$

$$= [\theta^T x^i - y^i]^2$$

So, $J_\theta(x) = \frac{1}{n} \sum_{i=1} [\theta^T x^{(i)} - y^{(i)}]$

(f) here given, two sets, let,

$$\theta_A = [\theta_0, \theta_1, \theta_2, \theta_3] = [0, 0.1, 0.008, 0.001]$$

$$\theta_B = [\theta_0, \theta_1, \theta_2, \theta_3] = [0, 0.5, 0.008, 0.002]$$

After normalizing given data:

Minmax normalization:

x_1	x_2	x_3	y
0.25	0	0.55	0.59
1	1	1	0.37
0	0.27	0	-0.27

So, for A case,

$$h_{\theta_1}(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$$

$$= 0 + 0.1 \times 0.25 + 0.008 \times 0 + 0.001 \times 0.55$$

$$= 0.02555$$

$$h_{\theta_2}(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$$

$$= 0 + 0.1 \times 1 + 0.008 \times 1 + 0.001 \times 1$$

$$= 0.109$$

$$h_{\theta_3}(x) = 0.27 \times 0.008$$

$$= 0.00216$$

$$J_1(\theta) = \left[h_{\theta_1}(x) - y^1\right]^2$$

$$= (0.02555 - 0.59)^2$$

$$= 0.318$$

$$J_2(\theta) = \left[0.109 - 0.37\right]^2$$

$$= 0.068$$

$$J_3(\theta) = \left[0.00216 + 0.27\right]^2$$

$$= 0.074$$

$$\therefore J\theta_A = \frac{1}{3} \sum_{i=1}^{3} J_i(\theta)$$

$$= \frac{1}{3}(0.318 + 0.068 + 0.074)$$

$$= 0.153$$

for case B,

$h_{\theta_1}(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 \cancel{+\theta_4}$

$= 0 + 0.5 \times 0.25 + 0 + 0.55 \times 0.002$

$= 0.1261$

$h_{\theta_2}(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$

$= 0.5 + 0.008 + 0.002$

$= 0.51$

$h_{\theta_3}(x) = 0.27 \times 0.008$

$= 0.00216$

$J\theta_B = \frac{1}{3}\{(h_{\theta_1}(x) - 0.59)^2 + (h_{\theta_2}(x) - 0.37)^2 + (h_{\theta_3}(x) + 0.27)^2\}$

$= \frac{1}{3}(0.215 + 0.0196 + 0.074)$

$= 0.103$

As, $J\theta_B < J\theta_A$,

the second set of parameters is the better for this model. Because

~~the~~ we know the less cost function value, the more the model is better.

Ans. to the Ques. NO: 1

ⓐ

ILSVRC is is an ~~am~~ annual competition which helped between 2010 and 2017 ~~ha~~ in which challenge tasks use subset of the ImageNet dataset. The full form of ILSVRC is ImageNet Large Scale Visual Recognition Challenge.

In this completition, the competitors are asked to work on the subset of ~~their~~ dataset of image. There were many challenges like localization, classification etc.

~~TB~~ In this competition, an algorithm is ~~ak~~ asked to represent where 5 class labels are needed to ~~be~~ predict for one ~~pic~~ image in decreasing order of confidence. ~~and~~

In this competition, the competitors over the 7 years came up with brilliant solutions which keeps very important impact on image classification and localization.

for instance, the accury of an error rate of an model (2010 winner's model) is 28.2 ~~which~~ ~~is~~ ~~for~~ ~~another~~ and in the ~~last~~ ~~ye~~ 2017, the predicted error rate was only 2.3 for the winner's model of 2017. By seeing this error rate, ~~it~~ this we might say that due to this competition

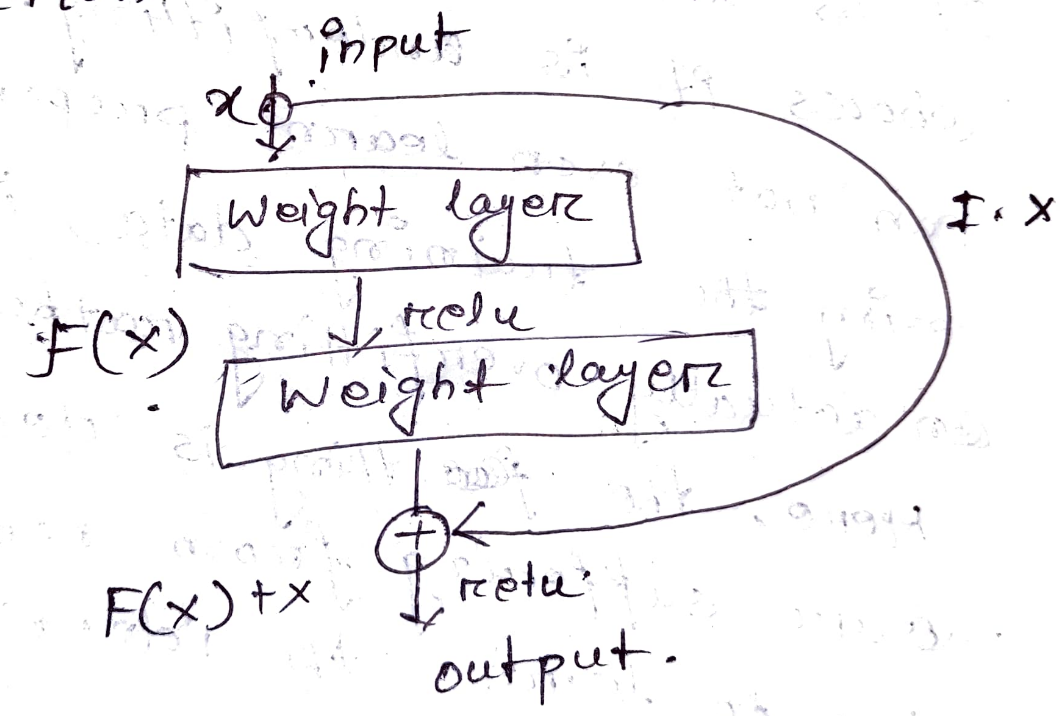
~~the error rate on~~ the evaluation ~~is~~ of image classification ~~is~~ was very ~~vital~~ important.

(b)

Here, Mr X's hypothesis is deeper networks always perform better than shallow networks. But experimenting with a deep and a shallow networks he got the ~~for the~~ given results. It is because the training data is underfitting. The model shows it is can not even learn properly. By seeing the training data, it is not actually overfitting rather underfitting. Here, the ~~fa~~ thing is deeper models are suffering from not being able to learn the identity function. ~~The~~ Actually, the problem is in optimization and the model ~~che~~ is not being able to optimize properly.

(c)

The next year's winner's model was ResNet. They overcome the problem by adding a skip connection.
Suppose two layers are there in stacking way. The input is ~~connect~~ concatinated by a skip connection with the output.



Here input = output when

means, $X = F(x) + X$

$\Rightarrow X = 0 + X$

So, the weights are learnable parameter and can be edited by the preferences of the model.

Ans. to the Ques. NO: 2

~~mAP@0.5~~
By running object detector on all test images (with NMS); the mAP@0.5 is calculated:
• mAP = avg of AP for each catagory
where AP = area under precision vs Recall curve.

(a)

(b)

~~So, ... the~~
The values at mAP@0.75 will be lower than the values of mAP@0.5 as the IOU is less ~~Area of A Ans~~ than 0.75.

~~B~~ Model A is predicted 60% and Model B 70%. The ~~#~~ given threshold is 75%. As the past predictions are ~~a~~ less than 75%, it can be said mAP@0.75 ~~is not~~ ~~f~~ do not fall in True positive ~~a~~ term for both Model and so, mAP@0.75 is lower.

(c)

By comparing performance between the given categories, multistage methods have more mAP than single stage methods. Here, the single stage ~~f~~ belongs to single model A belongs to single stage method and Model B belongs to under multistage.

So, ~~It~~ Model B has ~~less~~ greater computational expenses that the Model A. Because ~~single~~ multistage method has ~~less~~ greater computational expenses.