Name: Shadab Iqbal
ID: 19101072

AnsToTheQuestionNo. 1 (a)

ILSVRC (ImageNet Large Scale Visual Recognition Challenge) is an annual competition which is held between 2010 and 2017 and uses subsets of the ImageNet dataset.
The evaluation was done based on both classification and localization. The metric used for the evaluation would allow an algorithm to identify multiple objects within an image but would not penalize if the algorithm could identify an object which was actually present, but missing in the ground truth. Moreover, the algorithm had to produce 5 class labels in decreasing order of confidence and 5 bounding boxes for each of the class labels.

AnsToTheQuestionNo. 1 (b)

The problems deeper models are suffering from is, it is not being able to emulate the shallower models, which it should be able to do. As a result, the model keeps underfitting. The deeper models could not learn the identity function. If it could learn identity function, it could easily emulate shallower models just by adding some extra layers. So, what's happening in deeper models is that it is messing up the weights during backpropagation and thus making the model underfit.

AnsToTheQuestionNo. 1 (c)

The winner of the next year's ILSVRC classification challenge was ResNet. The structure is like VGG, but with residual blocks. ResNet uses skip connections. After every 2 convolutional layers, the input keeps being concatenated to the output. As the weights are learnable parameters, ResNet learns the weights such as it becomes 0. As a result, the input and output remain the same. That is why ResNet can work with a very large network and can still provide the optimal model like a shallower network.

AnsToTheQuestionNo . 2 (a)

For model A, mAP@0.5 is calculated by averaging the AP for each category. And the threshold IoU value, in this case, is 0.5. As the result is > 0.5, it will be marked as True positive.

AnsToTheQuestionNo . 2 (b)

Values of mAP@0.75 will be lower because we are increasing the threshold IoU value. That means, we will consider True positive when the predicted bounding box and GT bounding box overlaps more than 75%, which is a more difficult case. As a result, the performance will be degraded.

AnsToTheQuestionNo . 2 (c)

The main difference between single-stage and multi-stage methods is that -

In the multi-stage method, the model first divides the whole image into several regions of interest and then runs the classifier on each of those regions of interest to predict the probability of it being an object and the bounding boxes. Here, the training time is longer but provides great accuracy. But during testing, it works comparatively slower as it has greater computational complexity.

In the single-stage method, there is a backbone CNN which extracts meaningful features from the image and then is passed to the final layer and the final layer outputs a w * h * (B(C+5)) tensor, where B is the number of bounding boxes per cell and C is the number of classes. This single-stage method is fast but has less accuracy.

Example of single-stage: **YOLO**
Example of multi-stage: **Fast-RCNN**

Model **A** is in **single-stage method**
Model **B** is in **multi-stage method**

AnsToTheQuestionNo. 3 (a)

There are 3 features in this case.

AnsToTheQuestionNo. 3 (b)

This is a regression problem. The reason is that the target values are continuous.

AnsToTheQuestionNo. 3 (c)

When the feature values are of different ranges, the shape of the cost function becomes somewhat like an ellipse. This can lead to problems finding the global minima. In the given database, we can see different ranges of feature values. That is why we need to perform normalization as a preprocessing step in this case.

AnsToTheQuestionNo. 3 (d)

We should use the Mean Squared Error (MSE) performance metrics in this case. Because MSE is more sensitive to outlier target values. And in this question, we can see that the last value i.e -0.27 seems to be an outlier compared to the other target values.

AnsToTheQuestionNo. 3 (e)

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

Cost function:
$J(\theta) = (1/n)\ \Sigma$ i=0 to n $(J_i(\theta))$ where $J_i(\theta) = [y^{(i)} - h_\theta(x^i)]^2$

AnsToTheQuestionNo. 3 (f)

The first set of parameters will fit a better model. Because the values of the parameters in this set are lower compared to the values of the parameters of the second set. Again, if we plug in these values of the first set in the hypothesis function of linear regression, we will see that we will get a lower error than putting the values of the second set in the hypothesis function.

AnsToTheQuestionNo. 4 (a)

Correctly predicted cat image = 450 = TP
Correctly predicted dog image = 430 = TN

So, FP = 50 and FN = 70

Therefore,
Accuracy = (TP + TN) / (TP + TN + FP + FN)
= 880 / 1000
= 88%

AnsToTheQuestionNo. 4 (b)

Correctly predicted cat image = 60 = TP
Correctly predicted dog image = 65 = TN

So, FP = 40 and FN = 35

Therefore,
Accuracy = (TP + TN) / (TP + TN + FP + FN)
= 125 / 200
= 62.5%

AnsToTheQuestionNo. 4 (c)

The most likely problem with the trained classifier is that it is overfitting the model. Because we know that when the error becomes less during training but drastically rises during testing, the model is most likely overfitting.

AnsToTheQuestionNo. 4 (d)

One way to fix this is **Regularization**. What this does is, distributes the weights equally among the important features and then uses an alpha value to lessen the weight of the less important features.
Another way is **dropping out layers**. Here the network will remove certain features by setting them to 0 and thus reducing the fitting of the model overall.