CSE428: Image Processing

Final Exam
Date: 22 September, 2021
Window: 7.00 PM - 9.45 PM
Time: **2 Hours 30 Minutes (Exam) + 15 Minutes (Submission)**
Full Marks: **40** (10 points x 4 questions)
Handwritten/Typed both PDF submissions are acceptable.
Answer all 4 questions.

Question 1

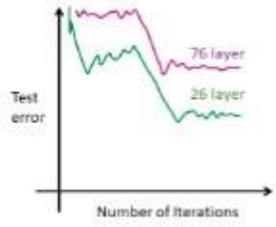
a. What is the ILSVRC challenge? Briefly explain how the performance of an algorithm was evaluated in the ILSVRC image classification criteria. **[1+2]**

b. In 2014, when GoogLeNet won the ILSVRC classification challenge (which had 3x more layers than the previous ILSVRC winner) Mr. X hypothesized that *deeper networks always perform better than shallow networks*. But experimenting with a deep and a shallow network, he got the following results, which was the exact opposite to his hypothesis.

Question 1

a. What is the ILSVRC challenge? Briefly explain how the performance of an algorithm was evaluated in the ILSVRC image classification criteria. **[1+2]**

b. In 2014, when GoogLeNet won the ILSVRC classification challenge (which had 3x more layers than the previous ILSVRC winner) Mr. X hypothesized that *deeper networks always perform better than shallow networks*. But experimenting with a deep and a shallow network, he got the following results, which was the exact opposite to his hypothesis.



If Mr. X's hypothesis is true, judging from the data provided, what seems to be wrong with the deeper models? **[4]**

c. How did the winner of the next year's ILSVRC classification challenge overcome the problem with deeper networks? What was their main idea? **[3]**

Question 2

On a particular object detection task the performance of two object detection models are listed below:

Model	Number of object classes	Performance	Speed
A	5	mAP@0.5 = 0.6	40 FPS
A	5	mAP@0.75 = ?	40 FPS
B	5	mAP@0.5 = 0.7	10 FPS
B	5	mAP@0.75 = ?	10 FPS

a. Explain how the value of mAP@0.5 was calculated for model A. **[2]**
b. For models A & B, will the values of mAP@0.75 be higher or lower than the values of mAP@0.5? Why? **[2]**
c. CNN based state of the art object detector models can be broadly categorized into
   i. Single-stage methods &
   ii. Multi-stage methods.

   What are the main differences between these two broad categories in terms of working principle (architecture, training) & performance (accuracy, speed)? Give an actual example of each kind. Which categories do A & B fall into? **[4+1+1]**

Question 3

Suppose you have a database, stored as a csv file, on which you want to perform a supervised learning algorithm. The first 3 rows of the dataset are shown below. The last column (y) is the target.

x1	x2	x3	y
3	216	-139	0.59
9	452	-112	0.37
1	282	-172	-0.27

a. How many features are there in this case? **[0.5]**
b. Is this a classification problem or a regression problem? Why? **[0.5 + 0.5]**
c. Explain why we should perform normalization as a preprocessing step in this case. **[1]**
d. To check whether a given predictive model is good or bad, which performance metrics should be used here? Why? **[0.5 + 1]**
e. Suppose you are using linear regression to predict **y** given **x1**, **x2**, and **x3**. Write the structure of the hypothesis function and the cost function. **[2]**
f. Among the following two sets of parameters, which one is the "better" model? Why? (**Hint**: What is the role of the cost function?) **[3]**

$$i)\ \theta = [\theta_0, \theta_1, \theta_2, \theta_3] = [0, 0.1, 0.008, 0.001]$$
$$ii)\ \theta = [\theta_0, \theta_1, \theta_2, \theta_3] = [0, 0.5, 0.008, 0.002]$$

Question 4

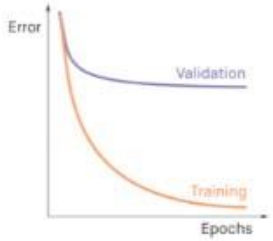
A certain neural network based image classifier is trained on 1000 images (500 cats, 500 dogs) to predict whether a given image is of a cat or a dog. After training 100 epochs with this dataset, the model's performance is as follows:

- On the 500 training cat images, the model identifies 450 images correctly (and hence misidentifies 50 cat images as dog images). On the other hand, out of 500 training dog images, the model identifies 430 images correctly and hence misidentifies 70 dog images as cat images.
- On the validation set, out of 100 cat images, the model identifies 60 images correctly, while out of 100 dog images on the validation set, 65 images are identified correctly by the model.

Clearly, the training performance of the model is superior compared to the validation performance.

a. Calculate the training accuracy **[2]**
b. Calculate the validation accuracy **[2]**

To further investigate the problem, the training loss (binary cross-entropy loss) for both validation set and training during training (per epoch) was calculated. The plot is given below:



[**Source**: Google image search]

c. Based on the figure above, what is the most likely problem with the trained classifier? **[2]**
d. Suggest at least **two** possible ways to fix this problem. **[4]**