Midterm for CAP5415

Computer Vision

Fall 2018

Thursday, October 04, 3-4:15pm

Full Name: _____

Student number: _____

Remarks:

This is a closed-book test. It is marked out of 100 points. Please answer ALL of the questions. Here is some advice:

- The questions are NOT arranged in order of difficulty, so you should attempt every question.
- If you think there may be many possible solutions for a question, answer them based on the lecture notes.
- Questions that ask you to "briefly explain" something only require short (1-3 sentence) explanations. Do not write a full page of text. I am looking for the main idea.
- None of the questions require long derivations. If you find yourself plugging through lots of equations, consider giving less detail or move on to the next question.
- Some questions have more than one right answer.
- Please write your answers carefully (it is only natural that if we cannot read your hand-writing, we cannot give you a point!).

Q1: _____ / 20 points
Q2: _____ / 20 points
Q3: _____ / 27 points
Q4: _____ / 16 points
Q5: _____ / 17 points **Total: _____ / 100**

Question 1: Each true or false question is 2 points _____ / (2 X 10 = 20 points in total)

True or False questions (circle one). Answer based on the lecture notes (justification is not needed):

a) Neural networks are inspired from the biological neurons: **True** or **False**

b) Loss function is a measure to compute the error over the entire dataset: **True** or **False**

c) Gradient Descent is an optimization technique: **True** or **False**

d) Stochastic gradient descent (SGD) needs to go through the individual data samples to update the weights: **True** or **False**

e) A single step of updating the parameters requires going through the entire dataset in (batch) gradient descent: **True** or **False**

f) Backpropagation algorithm uses chain rule to propagate the error starting from the output layer backwards to each layer within a deep architecture: **True** or **False**

g) If you have no GPU, you cannot use deep learning: **True** or **False**

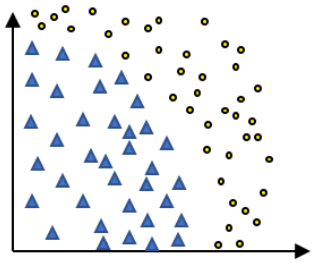
h) Imagine an image classification task where you have 7 image classes (categories) and you need to assign each given image into one of those 7 classes. You design a deep architecture. In the last layer of your architecture, you should not use logistic regression with single output: **True** or **False**

i) You should use convolutional layers in your earlier layers especially if you have spatial data (such as image) that comes with large height and weight sizes since convolutional layers require less number of parameters than FC layers to train typically: **True** or **False**

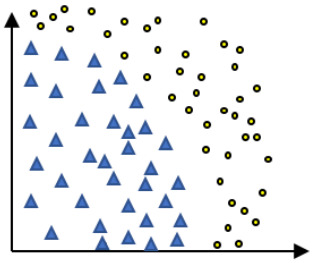
j) If you have a regression task, you can use the linear model at the last layer: **True** or **False**

Question 2: ____ /5 points for Figure1.
____ /5 points for Figure2.
____ /5 points for Figure3.
____ /5 points for both definitions.

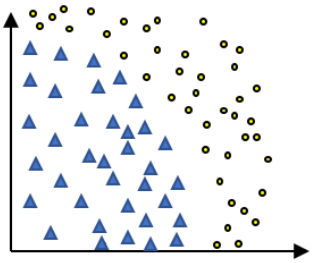
Please briefly **describe what is over-fitting and what is under-fitting** at the provided space below, **briefly**. Use the given three figures below and **illustrate a classifier** for each case. (Hint: remember the in-class discussion). circles represent class_1 and the triangles represents class_0 in the figures.



Draw: Under-fitting classifier example



Draw: Normal (expected) classifier example



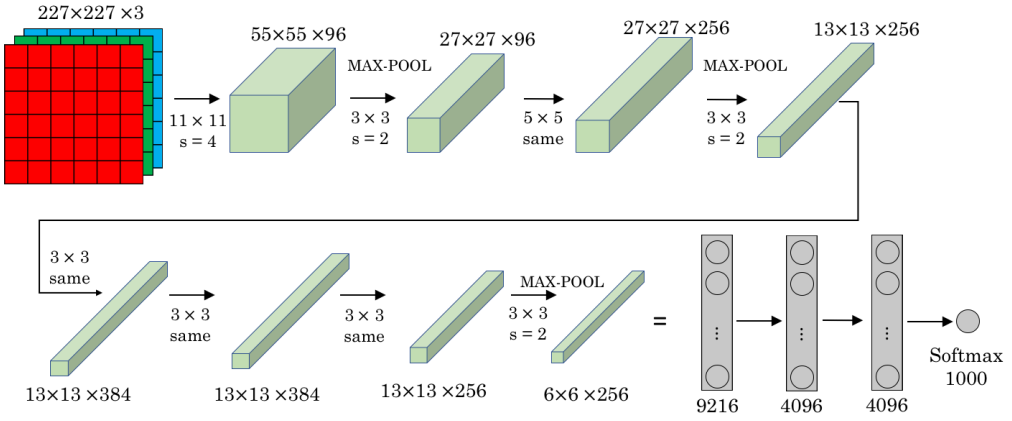
Draw: Over-fitting classifier example

Over-fitting definition:

Under-fitting definition:

Question 3: _____ / 27 points. Each item is 3 points (3X9=27 points for this question)

Consider the convolutional neural network architecture given in the figure below. This is a slightly modified version of the well-known AlexNet for image classification. Based on the given information in the figure, answer the questions below:



a) (Remember that pooling and convolutional layers are considered separate in this figure): Write down only the asked number next to each question below:
   - How many **convolutional layers** exist in the entire architecture:
   - How many **pooling layers** exist in the entire architecture:
   - How many **FC (fully-connected) layers** exist in the entire architecture (**including** the softmax layer):

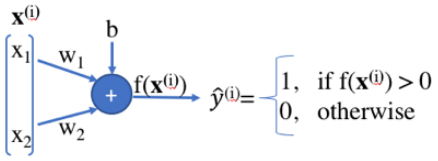
b) Consider that the indexing starts from the input side in the architecture (first, second, third, etc.). Write down the **total number of filters and filter dimensions as used**:
   - in the **first** convolutional layer:
   - in the **second** convolutional layer:
   - in the **softmax** layer:

c) Provide the **total number of trainable weights** for "**each** filter" (for this part, you can leave your expression as a product of integers; you do not need to actually compute the final product value):
   - in the **first** convolutional layer:
   - in the **second** convolutional layer:
   - in the **first** pooling layer:
   - in the **first FC** layer:

Question 4: _____ / 9 points for (a) where each data sample is 3 points. _____ / 7 points for (b).

Suppose that you are given the following two-dimensional dataset for a binary classification below. (Here you have only 3 training data where each data is two dimensional and you have their corresponding output label). Use the *linear model* given in the figure below (on the right) where you have the weights: w_1 and w_2 and the bias value: b. The notation: $f(\mathbf{x})$ defines the linear model. The prediction (estimated output label) is $\hat{y}$ and it is computed by thresholding the $f(\mathbf{x})$ value (the threshold is set to 0 in the given model below).

x_1	x_2	y
2	2	1
1	1	1
1	0	0



a) Write down (evaluate) the output inequalities for each of those three given input pairs using the parameters of the given model (at the end, you should obtain a set of 3 inequalities by using each input data):

The first inequality (for the first input pair: $\mathbf{x}^{(1)} = [2,2]^T$ and $y^{(1)} = 1$):

The second inequality (for the second input pair: $\mathbf{x}^{(2)} = [1,1]^T$ and $y^{(2)} = 1$):

The third inequality (for the third input pair: $\mathbf{x}^{(3)} = [1,0]^T$ and $y^{(3)} = 0$):

b) Find a set of parameters (including both weights and the bias value) satisfying all the inequalities you derived from the data above.
Note: that is similar to solving inequalities/equations in algebra and this problem has multiple solutions, therefore start with assigning an integer value to the parameter: b.

Question 5: _____ / 9 points for (a), _____ / 8 points for (b)

Consider a convolutional neural network that takes color images (of size: **260x260x3**) as input **for pedestrian localization** (this question is an open question since it has multiple answers).

a) Please **design a deep architecture** (and draw that below) that can be used for that application. In your figure, **be clear** in terms of the used layers, filter numbers, their size, padding and pooling values. You need to use multiple filters in each convolutional layer in your architecture. **Make sure to describe the format of your output layer in details** (the output format, number of neurons, what activation function(s) to be used and the meaning of the output values) for localization clearly.

b) Compute the dimensions of the feature maps after the first convolutional layer (i.e., the dims of the output volume of the first convolutional layer) in your designed architecture.