BME/ECE 695 Deep Learning
Midterm I
February 27, Spring 2020

Name:____.

Instructions:
This is a 75 minute exam containing five (5) problems.

- You may only use your brain and a pencil (or pen) to complete this exam.
- You may not use or have access to your book, notes, any supplementary reference, a calculator, or any communication device including a cell-phone, computer, smart watch, etc.
- You may not communicate with any person other than the official proctors during the exam.

Good Luck.

Problem 1) (20 pt)
For each of the following, check the **one** box that **best** corresponds to the function's properties.

a) $f(x) = e^{-x}$ for $x \in \Re$

☐ **Convex** ☐ **Concave**
☐ **neither** ☐ **Both**

b) $f(x) = x$ for $x \in \Re$

☐ **Convex** ☐ **Concave**
☐ **neither** ☐ **Both**

c) $f(x) = x^2$ for $x \in \Re$

☐ **Convex** ☐ **Concave**
☐ **neither** ☐ **Both**

d) $f(x) = x^3$ for $x \in \Re$

☐ **Convex** ☐ **Concave**
☐ **neither** ☐ **Both**

e) $f(x) = |x|$ for $x \in \Re$

☐ **Convex** ☐ **Concave**
☐ **neither** ☐ **Both**

f) $f(x) = |x|^3$ for $x \in \Re$ for $x \in \Re$

☐ **Convex** ☐ **Concave**
☐ **neither** ☐ **Both**

g) $f(x) = \sum_{k=0}^{K}(x - \mu_k)^2$ for $x \in \Re$, $\mu_k \in \Re$

☐ **Convex** ☐ **Concave**
☐ **neither** ☐ **Both**

h) $f(x) = \sum_{k=0}^{K}\{a_k e^{-x} + b_k x + c_k(x - \mu_k)^2\}$ for $x \in \Re$, $a_k \geq 0,\ b_k \geq 0, c_k \geq 0,\ \mu_k \in \Re$

☐ **Convex** ☐ **Concave**
☐ **neither** ☐ **Both**

Problem 2) (20 pt)
Mark each of the following statements as **<u>only one</u>** of the three following labels:

T-"true"; **F**-"false"; or **U**-"Undecidable given the information that is provided".

a) Gobbly gook is always blue.

b) Let $f(x)$ be a function of $x \in \Re$. For all x^*, if x^* is a global minimum of $f(x)$, then x^* must also be a local minimum of $f(x)$.

c) Let $f(x)$ be a function of $x \in \Re$. For all x^*, if x^* is a local minimum of $f(x)$, then x^* must also be a global minimum of $f(x)$.

c) Let $f(x)$ be a continuously differentiable function for $x \in \Re$. If $\frac{d}{dx}f(x^*) = 0$, then x^* is a local minimum.

d) Let $f(x)$ be a continuously differentiable and convex function for $x \in \Re$. If $\frac{d}{dx}f(x^*) = 0$, then x^* is a global minimum.

Problem 3) (20 pt)
Consider the following convolutional neural network (see diagram on next page) with a color image as input, and a gray-scale image as output. Each layer uses a ReLu activation function, and denote the convolution kernel by w and the offsets by b.

a) For layer 1, give:

1. The shape of the tensor w;
2. The number of parameters in w;
3. The shape of the vector b;
4. The number of parameters in b.
5. The total number of parameters in the layer.

b) For layer 2, give:

- The shape of the tensor w;
- The number of parameters in w;
- The shape of the vector b;
- The number of parameters in b.
- The total number of parameters in the layer.

c) The total number of parameters in the model.

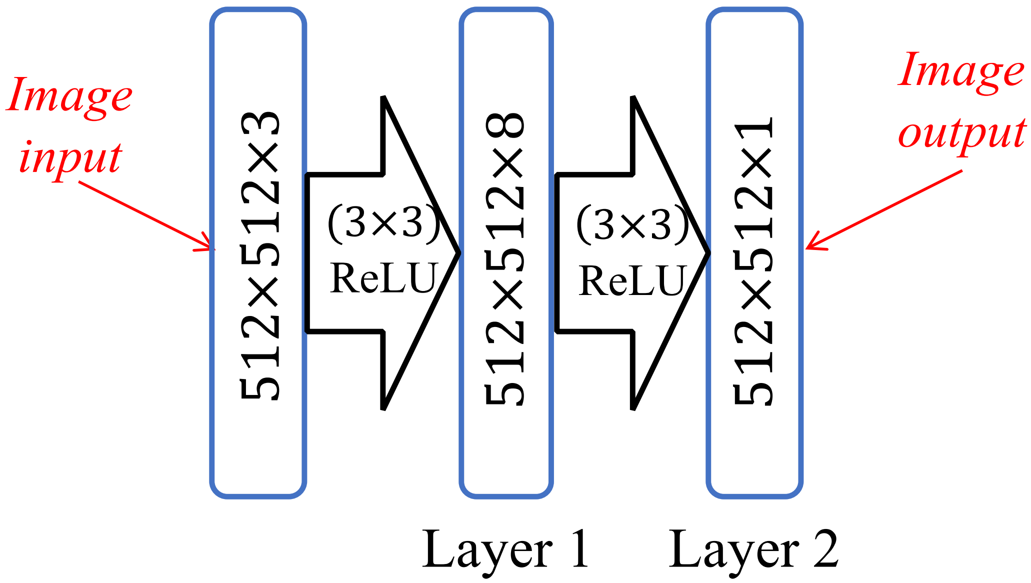

Figure 1: Convolutional Neural Network for Problem 3

Problem 4) (20 pt)
Consider the following tensor operation where z^{i_1,i_2} is the input, x^j is the output, and $w^j{}_{i_1,i_2}$ is the kernel.

$$x^j = w^j{}_{i_1,i_2} z^{i_1,i_2}$$

a) What is the rank of the tensor z^{i_1,i_2}?
(Hint: The rank of a tensor is the number of axes in the tensor.)

b) What is the rank of the tensor x^j?

c) What is the rank of the tensor $w^j{}_{i_1,i_2}$?

d) Draw a 3D picture illustrating this operation.

Problem 5) (20 pt)

Consider a machine learning (ML) system,

$$\hat{x} = f_\theta(y) = Ay + b \,,$$

where $y \in \Re^{N_y}$, $\hat{x} \in \Re^{N_x}$, and $\theta = [A, b]$ where $A \in \Re^{N_x \times N_y}$ and $b \in \Re^{N_x}$. Assume we have training data pairs given by $(x_k, y_k)|_{k=0}^{K-1}$, and a loss function given by

$$l(\theta) = \frac{1}{K} \sum_{k=0}^{K-1} \| x_k - f_\theta(y_k) \|^2 \,.$$

a) What is the commonly used name for this loss function?

b) How many scalar parameters are in this model, i.e., what is the dimension of θ?

c) Calculate a theoretical expression for $\nabla_b l(\theta)$, the gradient of the loss function w.r.t. b. You can express this in the form $g_{1,i} = [\nabla_b l(\theta)]_i$ for $0 \le i < N_x$.

d) Calculate a theoretical expression for $\nabla_A l(\theta)$, the gradient of the loss function w.r.t. A. You can express this in the form $g_{2,i,j} = [\nabla_A l(\theta)]_{i,j}$ for $0 \le i < N_x$ and $0 \le j < N_y$.

e) Write a pseudo-code algorithm for gradient descent of the parameter $\theta = [A, b]$.