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

Name: KEY.

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) (2.5 pt each)

For each of the following, check the **one** box that **best** corresponds to the function's properties.

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

- [x] Convex
- [ ] Concave
- [ ] neither
- [ ] Both

b) $f(x) = x$ for $x \in \mathfrak{R}$

- [ ] Convex
- [ ] Concave
- [ ] neither
- [x] Both

c) $f(x) = x^2$ for $x \in \mathfrak{R}$

- [x] Convex
- [ ] Concave
- [ ] neither
- [ ] Both

d) $f(x) = x^3$ for $x \in \mathfrak{R}$

- [ ] Convex
- [ ] Concave
- [x] neither
- [ ] Both

e) $f(x) = |x|$ for $x \in \mathfrak{R}$

- [x] Convex
- [ ] Concave
- [ ] neither
- [ ] Both

f) $f(x) = |x|^3$ for $x \in \mathfrak{R}$ for $x \in \mathfrak{R}$

- [x] Convex
- [ ] Concave
- [ ] neither
- [ ] Both

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

- [x] 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 $x \in \mathfrak{R}$, $a_k \geq 0,\ b_k \geq 0, c_k \geq 0,\ \mu_k \in \mathfrak{R}$

- [x] Convex
- [ ] Concave
- [ ] neither
- [ ] Both

Problem 2) (20 pt) (4 pt each)
Mark each of the following statements as **only one** of the three following labels:
T-"true"; F-"false"; or U-"Undecidable given the information that is provided".

a) Gobbly gook is always blue.

U

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 $f(x)$.

T

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 $f(x)$.

F

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.

F

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.

T

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:

2' 1. The shape of the tensor w;

$3 \times 3 \times 3 \times 8$

2' 2. The number of parameters in w;

216

2' 3. The shape of the vector b;

8×1

2' 4. The number of parameters in b.

8

1.5' 5. The total number of parameters in the layer.

224

b) For layer 2, give:

2' - The shape of the tensor w;

$3 \times 3 \times 8 \times 1$

2' - The number of parameters in w;

72

2' - The shape of the vector b;

1×1

2' - The number of parameters in b.

1

1.5' - The total number of parameters in the layer.

73

1' **c) The total number of parameters in the model.**

$224 + 73 = 297$

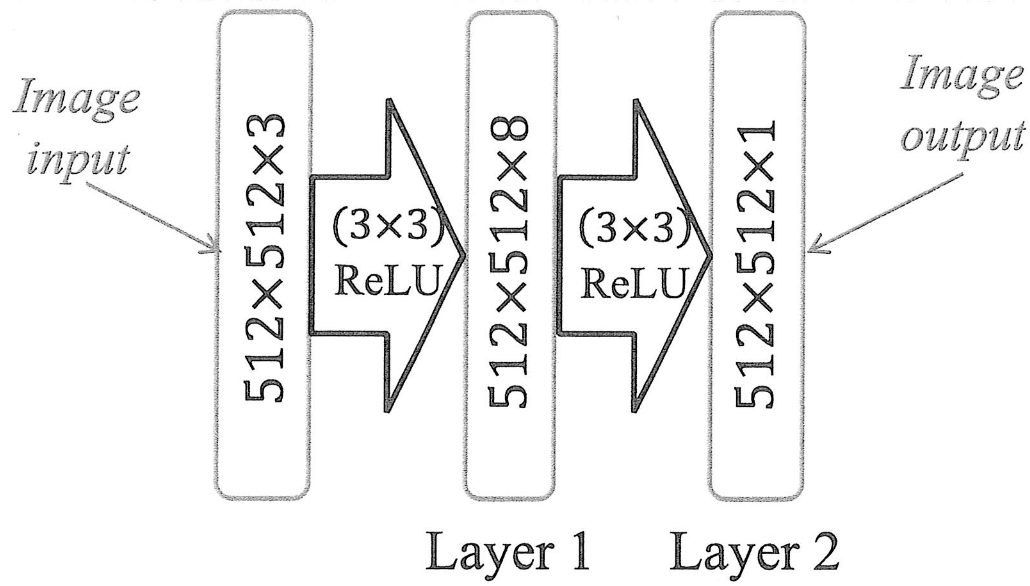


Figure 1: Convolutional Neural Network for Problem 3

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.

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

4' 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.)

2

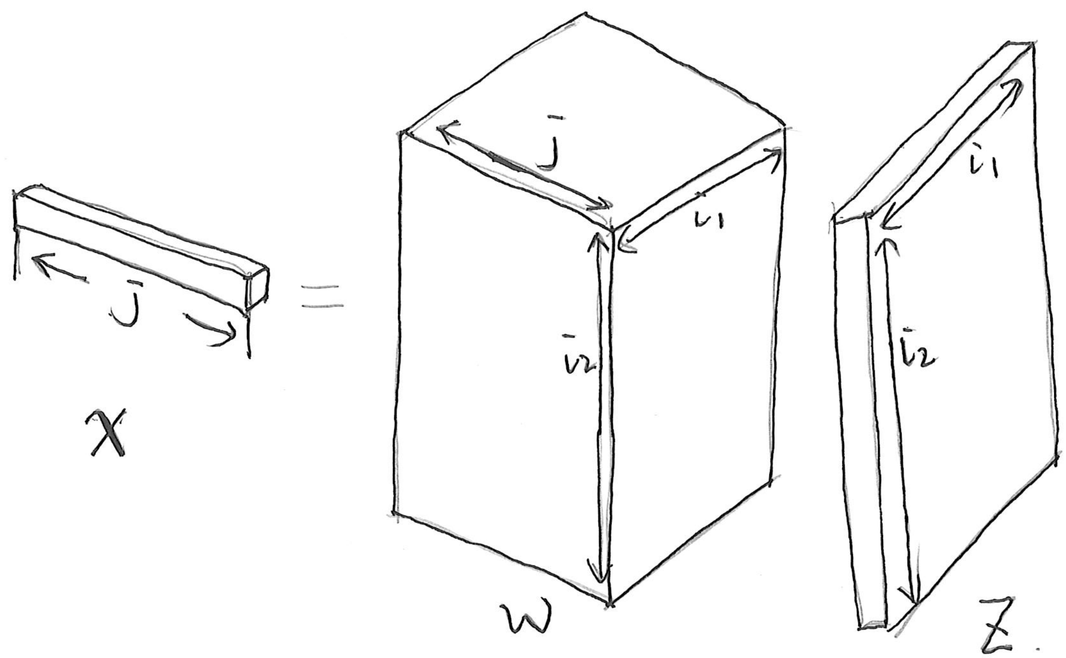
4' b) What is the rank of the tensor x^j?

1

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

3

8' 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 $\theta = [A, b]$.

(a) Mean Squared Error Loss.

(b) $N_x \times N_y + N_x$

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

where $\nabla_b f_\theta(y_k) = I$.

Therefore $g_{1,i} = [\nabla_b l(\theta)]_i = -\frac{2}{K}\sum_{k=0}^{K-1}[x_k]_i - [f_\theta(y_k)]_i$

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

where $\left[\nabla_A f_\theta(y_k)\right]^i_{j_1,j_2} = \delta^i_{j_1}[y_k]^{j_2}$

Therefore $g_{2,i,j} = \left[\nabla_A l(\theta)\right]_{i,j}$

$$= -\frac{2}{K}\sum_{k=0}^{K-1}\left[(x_k - f_\theta(y_k))^t\right]_m \left[\nabla_A f_\theta(y_k)\right]^m_{i,j}$$

$$= -\frac{2}{K}\sum_{k=0}^{K-1}\left[x_k - f_\theta(y_k)\right]_i [y_k]^j$$

(e) Initialize A, b, α

while not converge{

$d = -[\nabla_A l(\theta), \nabla_b l(\theta)]$

$[A,b] \leftarrow [A,b] + \alpha d$

}