

BME/ECE 695 Deep Learning  
Midterm I Solution  
March 11, Spring 2021

**Q1.**

2 Points

**Rules:** I understand that this is an open book exam that shall be done within the allotted time of 120 minutes. I can use my notes, and web resources. However, I will not communicate with any other person other than the official exam proctors during the exam, and I will not seek or accept help from any other persons other than the official proctors.

*Upload a scan of your signature here:*

**Name:** (4 pt) \_\_\_\_\_

## Q2 Minimum of Continuous Function

(10 Points)

Let  $f : [0, 1]^N \Rightarrow \mathfrak{R}$  be a continuous function.

Does  $f$  take on a global minimum? Either prove the result or give a counter example.

---

---

**Solution:**

**Q2**

Yes,  $f$  takes on a global minimum.

Proof:

Let  $\mathcal{A}$  denote the set  $[0, 1]^N \subset \mathfrak{R}^N$ . Since  $\mathcal{A} = [0, 1]^N$  is the N-dimensional Cartesian product of the closed and bounded interval  $[0, 1]$ ,  $\mathcal{A}$  is both closed and bounded. Hence,  $\mathcal{A}$  is a compact set.

Given that  $f$  is a continuous function, it must attain a maximum and a minimum on a compact set. Therefore,  $f$  takes on a global minimum.

### Q3 Minimum of Convex Function

(10 Points)

Let  $f : \mathfrak{R}^N \Rightarrow \mathfrak{R}$  be a convex function.

Does  $f$  take on a global minimum? Either prove the result or give a counter example.

---

**Solution:**

**Q3**

No,  $f$  does not necessarily take on a global minimum.

Counter example:

A convex function may not achieve a minimum value over its domain.

For example,

$$f(x) = Ax + b$$

where  $x \in \mathfrak{R}^N$ ,  $A \in \mathfrak{R}^{1 \times N}$ ,  $b \in \mathfrak{R}$ .

**Q4**

(10 Points)

Let  $f : \mathfrak{R}^N \Rightarrow \mathfrak{R}$  be a convex function with local minimum  $x^* \in \mathfrak{R}^N$ .

**Q4.1 Miminum**

(5 Points)

Is  $x^*$  a global minimum? Either prove the result or give a counter example.

**Q4.2**

(5 Points)

Is  $x^*$  a unique local minimum? Either prove the result or give a counter example.

**Solution:****Q4.1**

Yes,  $x^*$  is a global minimum.

Proof:

Since  $x^* \in \mathfrak{R}^N$  is a local minimum, then for  $\alpha \in [0, 1]$  and  $x \in \mathfrak{R}^N$  we have,

$$f(x^*) \leq f(x^* + \alpha\Delta x) = f(x^* + \alpha(x - x^*)) = f(\alpha x + (1 - \alpha)x^*)$$

By the definition of convex function,

$$f(\alpha x + (1 - \alpha)x^*) \leq \alpha f(x) + (1 - \alpha)f(x^*)$$

Hence,

$$f(x^*) \leq \alpha f(x) + (1 - \alpha)f(x^*)$$

$\Rightarrow f(x^*) \leq f(x)$ , which implies that  $x^*$  is a global minimum of convex function  $f$ .

**Q4.2**

No,  $x^*$  is not necessarily a unique local minimum.

Counter example:

The local minimum of a convex function is not necessarily unique.

For example, considering the constant function

$$f(x) = c$$

where  $x \in \mathfrak{R}^N$  and  $c \in \mathfrak{R}$  is a constant, the local minimum exists but not unique because every point is a local minimum.

### Q5 Gradient Descent

(35 Points)

Consider the loss function  $L : \Re^2 \Rightarrow \Re$  where

$$L(x) = a(x_1 + x_2)^2 + b(x_1 - x_2)^2$$

where  $a > 0$  and  $b > 0$  are scalar parameters and  $x = [x_1, x_2]$ .

#### Q5.1 Mimumum

(5 Points)

Prove that  $L$  convex.

Hint: You can use the “linear transform” property of convexity from the notes.

#### Q5.2

(5 Points)

Does  $L(x)$  take on a global minimum?

Justify your answer, and if the answer is ”yes”, then give its value?

#### Q5.3

(5 Points)

Calculate the gradient  $\nabla L(x)$ .

#### Q5.4

(5 Points)

Write out the gradient descent algorithm with step size  $\alpha$ .

#### Q5.5

(5 Points)

If  $b = 100$  and  $a = 1$ , then sketch a contour plot of the function  $L(x)$ . Label the key dimensions and axes.

#### Q5.6

(5 Points)

If  $b = 100$  and  $a = 1$ , then sketch a typical path of the gradient descent algorithm on a contour plot of the loss function.

#### Q5.7

(5 Points)

Explain in words why gradient descent optimization will be slow when  $b \gg a$ .

---

---

**Solution:**

**Q5.1**

Proof:

Let  $s : \Re \Rightarrow \Re$  denote the square function where  $s(y) = y^2$  for  $y \in \Re$ , and  $f : \Re^2 \Rightarrow \Re$  be

$$f(x) = as(x_1) + bs(x_2) = ax_1^2 + bx_2^2$$

Given  $a > 0, b > 0$ ,  $f(x)$  is a convex function since  $s$  is convex and  $f$  is a linear combination of the convex function.

Let  $A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ , the linear transform of the convex function  $f(x)$ :

$$L(x) = f(Ax) = f\left(\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = f\left(\begin{bmatrix} x_1 + x_2 \\ x_1 - x_2 \end{bmatrix}\right) = a(x_1 + x_2)^2 + b(x_1 - x_2)^2$$

is convex by “linear transformation” property of convexity.

**Q5.2**

$L(x)$  takes on a global minimum because it’s continuously differentiable and convex on  $\Re^2$ . By solving  $\Delta_x L(x) = 0$ , we know the function achieves global minimum value 0 at  $[x_1, x_2] = [0, 0]$ .

**Q5.3**

The gradient is

$$\begin{aligned} \Delta L(x) &= \left[ \frac{\partial L(x)}{\partial x_1}, \frac{\partial L(x)}{\partial x_2} \right] \\ &= [2ax_1 + 2ax_2 + 2bx_1 - 2bx_2, 2ax_1 + 2ax_2 - 2bx_1 + 2bx_2] \\ &= [2(a+b)x_1 + 2(a-b)x_2, 2(a-b)x_1 + 2(a+b)x_2] \end{aligned} \tag{1}$$

**Q5.4**

**Repeat until converged**{

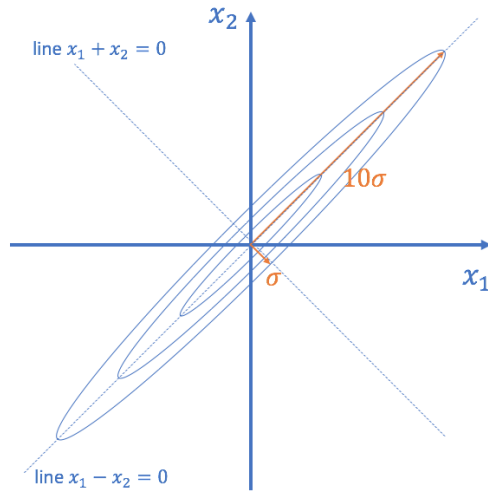
$$d \leftarrow -\Delta_x L(x)$$

$$x \leftarrow x + \alpha d^t$$

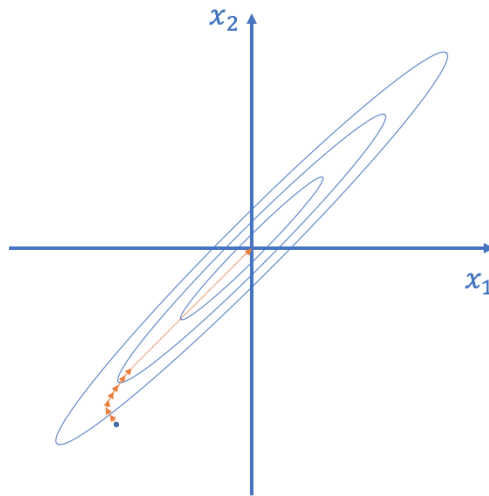
}

**Q5.5**

Given  $b = 100$  and  $a = 1$ , the contour plot is shown below

**Q5.6**

An example path of gradient descent algorithm is shown below

**Q5.7**

Define the following two unit vectors.

$$e_1 = \frac{[1, 1]}{\sqrt{2}}$$

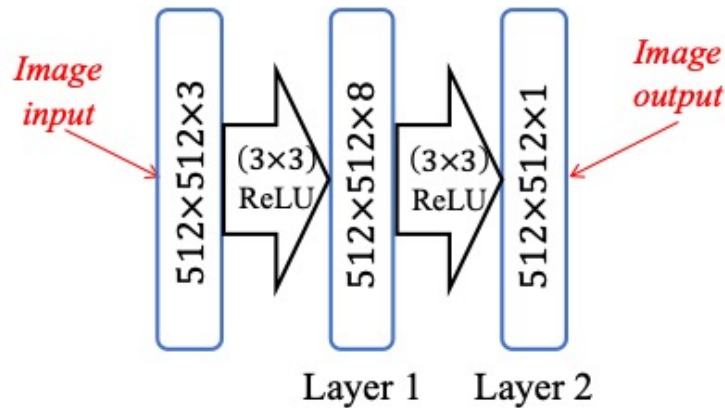
$$e_2 = \frac{[1, -1]}{\sqrt{2}}$$

When  $b \gg a$ , the gradient changes rapidly along the  $e_2$  direction, but the gradient changes slowly along the  $e_1$  direction. Consequently, a step size that avoids oscillations along the  $e_2$  direction, will result in slow convergence along  $e_1$  direction. So the convergence of the overall algorithm will be slow.

## Q6 Convolutional Neural Networks

(30 Points)

Consider the following convolutional neural network pictured below 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  $b$ .



### Q6.1

(5 Points)

For layer 1, what are the shapes of the tensors  $w$  and  $b$ ?

### Q6.2

(5 Points)

For layer 1, what are the number of parameters?

### Q6.3

(5 Points)

For layer 2, what are the shapes of the tensors  $w$  and  $b$ ?

### Q6.4

(5 Points)

For layer 2, what are the number of parameters?

### Q6.5

(10 Points)

What are the advantages of a convolutional layer over a fully connected layer?



---

---

**Solution:**

**Q6.1**

For layer 1, the shape of tensor  $w$  is  $3 \times 3 \times 3 \times 8$  and the shape of  $b$  is  $8 \times 1$ .

**Q6.2**

The total number of parameters for layer 1 is

$$3 \times 3 \times 3 \times 8 + 8 \times 1 = 224$$

**Q6.3**

For layer 2, the shape of tensor  $w$  is  $3 \times 3 \times 8 \times 1$  and the shape of  $b$  is  $1 \times 1$ .

**Q6.4**

The total number of parameters for layer 2 is

$$3 \times 3 \times 8 \times 1 + 1 \times 1 = 73$$

**Q6.5**

The advantages of a convolutional layer over a fully connected layer include:

- The convolutional layer requires fewer model parameters.
- The convolutional layer requires less training data.
- The convolutional layer requires less computation and memory.
- The convolutional layer is space invariant.
- The convolutional layer typically achieves better accuracy in computer vision and image processing tasks.