Problem Set 2

CS 4277: Deep Learning

Name (print clearly): ______________________________ Section: (e.g., 01) ________

Signature: ____

Student account username (e.g., msmith3): ______________________________

Signing signifies that you agree to comply with the **Academic Honor Code** and course policies stated in the syllabus.

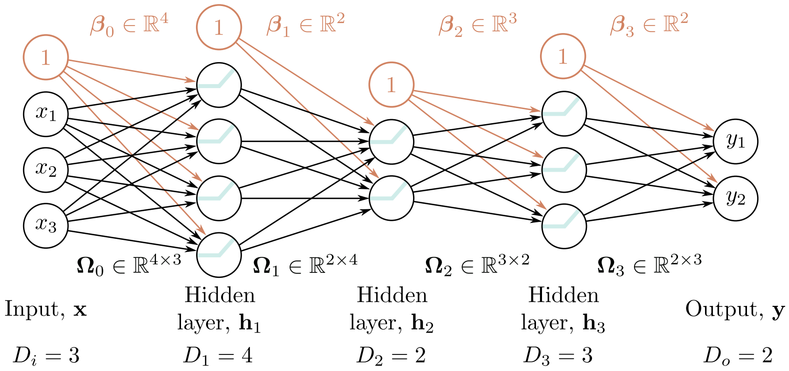
Choose one of these two options for turn-in:

1. Print this document, write or answers, scan your finished homework to a PDF, name the PDF `cs4277-ps2-<your-student-account-username>.pdf`, e.g., `cs4277-ps2-msmith3.pdf` and submit the PDF to the assignment on D2L.

2. While viewing this document in your web browser, in the address bar change `.pdf` to `.tex`, save the LaTeX source as a text file, add your answers in appropriate LaTeX markup in the appropriate spaces, compile to a PDF named as in the instructions above, and submit the PDF file to the assignment on D2L.

Question:	1	2	3	4	5	6	7	8	9	Total
Points:	10	10	15	10	15	14	13	10	15	112
Score:										

1. (10 points) Problem 4.2 Identify the hyperparameters in Figure 4.6, reproduced here for convenience (Hint: there are four!).


$\boldsymbol{\beta}_0 \in \mathbb{R}^4$
$\boldsymbol{\beta}_1 \in \mathbb{R}^2$
$\boldsymbol{\beta}_2 \in \mathbb{R}^3$
$\boldsymbol{\beta}_3 \in \mathbb{R}^2$
1
1
1
1
x_1
x_2
x_3
y_1
y_2
$\boldsymbol{\Omega}_0 \in \mathbb{R}^{4\times 3}$
$\boldsymbol{\Omega}_1 \in \mathbb{R}^{2\times 4}$
$\boldsymbol{\Omega}_2 \in \mathbb{R}^{3\times 2}$
$\boldsymbol{\Omega}_3 \in \mathbb{R}^{2\times 3}$
Input, $\mathbf{x}$
Hidden layer, $\mathbf{h}_1$
Hidden layer, $\mathbf{h}_2$
Hidden layer, $\mathbf{h}_3$
Output, $\mathbf{y}$
$D_i = 3$
$D_1 = 4$
$D_2 = 2$
$D_3 = 3$
$D_o = 2$

2. (10 points) Problem 4.4 Write out the equations for a deep neural network that takes $D_i = 5$ inputs, $D_o = 4$ outputs and has three hidden layers of sizes $D_1 = 20$, $D_2 = 10$, and $D_3 = 7$, respectively in both the forms of equations 4.15 and 4.16 (reproduced below for convenience). What are the sizes of each weight matrix $\Omega_\bullet$ and bias vector $\beta_\bullet$?

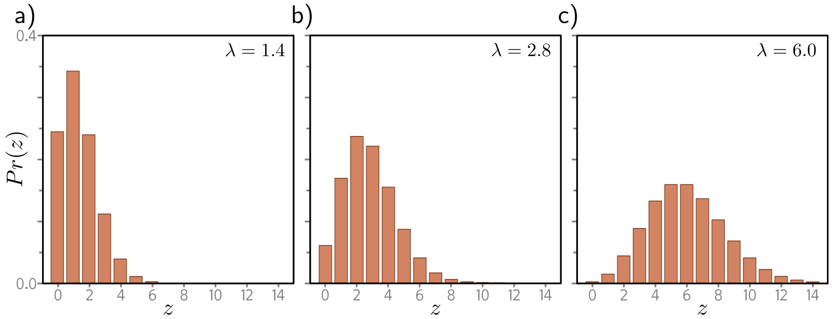
$$\boldsymbol{h}_K = a\,(\boldsymbol{\beta}_{K-1} + \boldsymbol{\Omega}_{K-1}\boldsymbol{h})_{K-1})$$

$$\boldsymbol{y} = \boldsymbol{\beta}_K + \boldsymbol{\Omega}_K\boldsymbol{h}_K \tag{4.15}$$

We can write the whole network as:

$$\boldsymbol{y} = \boldsymbol{\beta}_K + \boldsymbol{\Omega}_K a\,(\boldsymbol{\beta}_{K-1} + \boldsymbol{\Omega}_{K-1} a(...\boldsymbol{\beta}_1 + \boldsymbol{\Omega}_1 a(\boldsymbol{\beta}_0 + \boldsymbol{\Omega}_0 a))) \tag{4.16}$$

3. (15 points) Problem 5.6 Consider building a model to predict the number of pedestrians $y \in \{0, 1, 2, \dots\}$ that will pass a given point in the city in the next minute, based on data x that contains information about the time of day, the longitude and latitude, and the type of neighborhood. A suitable distribution for modeling counts is the Poisson distribution (figure 5.15 from book, reproduced below for convenience).



This has a single parameter $\lambda > 0$ called the rate that represents the mean of the distribution. The distribution has probability density function:

$$Pr(y = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Use the recipe in Section 5.2 to design a loss function for this model assuming that we have access to I training pairs $\{\boldsymbol{x_i}, y_i\}$.

4. (10 points) Problem 6.1 Show that the derivatives of the least squares loss function in equation 6.5:

$$L(\boldsymbol{\phi}) = \sum_{i=1}^{I} (\phi_0 + \phi_1 x_i - y_i)^2 \tag{6.5}$$

are given by the expressions in equation 6.7:

$$\frac{\partial \ell_i}{\partial \boldsymbol{\phi}} = \begin{bmatrix} \frac{\partial \ell_i}{\partial \phi_0} \\ \frac{\partial \ell_i}{\partial \phi_1} \end{bmatrix} = \begin{bmatrix} 2(\phi_0 + \phi_1 x_i - y_i) \\ 2x_i(\phi_o + \phi_i x_i - y_i) \end{bmatrix} \tag{6.7}$$

5. (15 points) Problem 6.2 A surface is convex if the eigenvalues of the Hessian $H(\boldsymbol{\phi})$ are positive everywhere. In this case, the surface has a unique minimum, and optimization is easy. Find an algebraic expression for the Hessian matrix,

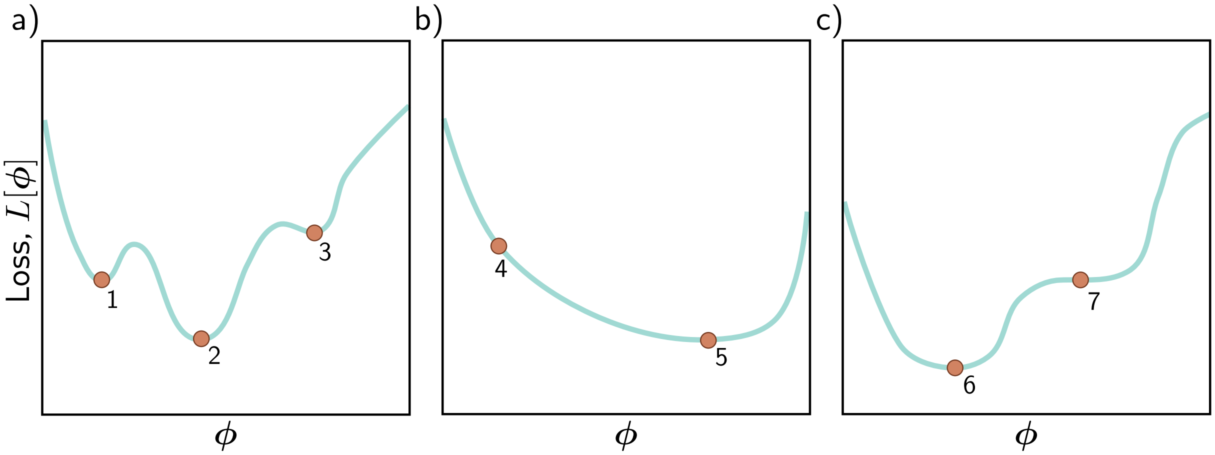
$$H(\boldsymbol{\phi}) = \begin{bmatrix} \frac{\partial^2 L}{\partial \phi_0^2} & \frac{\partial^2 L}{\partial \phi_0 \partial \phi_1} \\ \frac{\partial^2 L}{\partial \phi_1 \partial \phi_0} & \frac{\partial^2 L}{\partial \phi_1^2} \end{bmatrix} \tag{6.20}$$

for the linear regression model (equation 6.5):

$$L(\boldsymbol{\phi}) = \sum_{i=1}^{I} (\phi_0 + \phi_1 x_i - y_i)^2 \tag{6.5}$$

Prove that this function is convex by showing that the eigenvalues are always positive. This can be done by showing that both the trace and the determinant of the matrix are positive.

6. (14 points) Problem 6.6 Which of the functions in Figure 6.11 from the book (preproduced here for convencience) is convex?

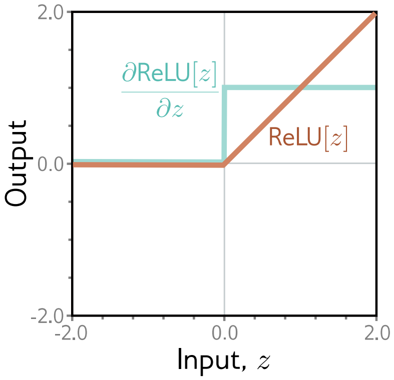


Justify your answer. Characterize each of the points 1-7 as (i) a local minimum, (ii) the global minimum, or (iii) neither.

7. (13 points) Problem 7.1 A two-layer network with two hidden units in each layer can be defined as:

$$
\begin{aligned}
y = \phi_0 &+ \phi_1 a(\psi_{01} + \psi_{11} a(\theta_{01} + \theta_{11}x) + \psi_{21} a(\theta_{02} + \theta_{12}x)) \\
&+ \phi_2 a(\psi_{02} + \psi_{12} a(\theta_{01} + \theta_{11}x) + \psi_{22} a(\theta_{02} + \theta_{12}x))
\end{aligned}
$$

where the functions $a(\bullet)$ are ReLU functions. Compute the derivatives of the output y with respect to each of the 13 parameters $\phi_\bullet$, $\theta_{\bullet\bullet}$, and $\psi_{\bullet\bullet}$ directly (i.e., not using the backpropagation algorithm). The derivative of the ReLU function with respect to its input $\frac{\partial a(z)}{\partial z}$ is the indicator function $\mathbb{I}(z > 0)$, which returns one if the argument is greater than zero and zero otherwise (Figure 7.6, reproduced here for convenience).



8. (10 points) Problem 7.2 Find an expression for the final term in each of the five chains of derivatives in equation 7.12 (reproduced here for convenience).

$$\frac{\partial \ell}{\partial h_3} = \frac{\partial f_3}{\partial h_3} \frac{\partial \ell_i}{\partial f_3} \tag{7.12}$$

9. (15 points) Problem 7.5 Calculate the derivative $\frac{\partial \ell_i}{\partial f(\boldsymbol{x_i}, \boldsymbol{\phi})}$ for the binary classification loss function:

$$\ell_i = -(1 - y_i)\log(1 - sig(f(\boldsymbol{x_i}, \boldsymbol{\phi}))) - y_i \log(sig(f(\boldsymbol{x_i}, \boldsymbol{\phi})))$$

where the function $sig(\bullet)$ is the logistic sigmoid and is defined as:

$$sig(z) = \frac{1}{1 + exp(-z)}$$