

Homework 7

Instructions

This homework contains **4** concepts and **6** programming questions. In MS word or a similar text editor, write down the problem number and your answer for each problem. Combine all answers for concept questions in a single PDF file. Export/print the Jupyter notebook as a PDF file including the code you implemented and the outputs of the program. Make sure all plots and outputs are visible in the PDF.

Combine all answers into a single PDF named `andrewID_hw7.pdf` and submit it to Gradescope before the due date. Refer to the syllabus for late homework policy. Please assign each question a page by using the “Assign Questions and Pages” feature in Gradescope.

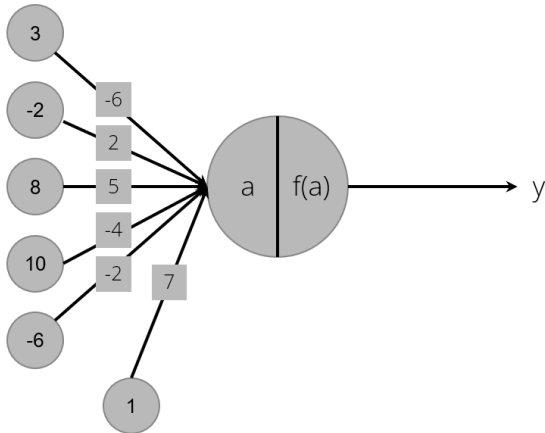
Here is a breakdown of the points for programming questions:

Name	Points
M7-L1-P1	10
M7-L1-P2	5
M7-L2-P1	5
M7-L2-P2	10
M7-HW1	30
M7-HW2	30

Problem 1 (2.5 points)

Consider the following perceptron. Compute the output y , using $\sigma(a)$, the sigmoid activation function

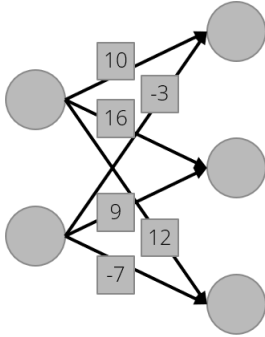
$$\sigma(a) = \frac{1}{1 + \exp(-a)}$$



Problem 2 (2.5 Points)

(Multiple choice)

Which of the following weight matrices is correct for the provided fully connected layers?



$$W_1 = \begin{bmatrix} 10 & 16 & 12 \\ -3 & 9 & -7 \end{bmatrix} \quad W_2 = \begin{bmatrix} 10 & -3 & 16 \\ -3 & 9 & -7 \end{bmatrix}$$

$$W_3 = \begin{bmatrix} 10 & -3 \\ 16 & 9 \\ 12 & -7 \end{bmatrix} \quad W_4 = \begin{bmatrix} 10 & 9 \\ -3 & 12 \\ 16 & -7 \end{bmatrix}$$

Problem 3 (2.5 Points)

The following question concerns the sigmoid, tanh, and softmax activation functions.

(Multiple choice choose one)

Select the true statement:

1. The tanh activation function is suitable in the output layer for binary classification problems because its output has a probabilistic interpretation
2. The softmax activation function is used in the output layer for multi-class classification problems to produce a probability distribution over multiple classes
3. Unlike other activation functions, the sigmoid activation function does not suffer from the problem of vanishing gradients in networks with many hidden layers
4. All of the above

Problem 4 (2.5 Points)

The following question considers the ReLU, Leaky ReLU and GELU activation functions.

(Multiple choice choose one)

Select the true statement:

1. The GELU activation function is a smooth approximation of the ReLU function, which means its derivative is continuous
2. The derivatives of ReLU and Leaky ReLU have a discontinuity at $x = 0$
3. Leaky ReLU is a variant of ReLU that allows a small non-zero gradient for negative input values,
4. All of the above