Ain Shams University, Faculty of Engineering,
Computer & Systems Engineering Department
CSE463: Neural Networks
Sheet #03: Activation and Loss Functions

1- A multi-layer perceptron network is used to solve the XOR problem. It has two nodes in the hidden layer. All nodes has a sign activation function. Tune mathematically the weighting parameters of such a network. Visualize the classification boundaries.
2- Find the derivative of the sigmoid function. List the advantages and disadvantages of this function.
3- Find the derivative of the tanh function. List the advantages and disadvantages of this function.
4- The following data set of 2D points, {(-1, -1), (+1, -1), (-1, +1), (+1, +1)} and their corresponding labels {+1, +1, +1, -1} is trained with a single neuron neural network.
    a. Find the total loss if, $\mathbf{W} = [0.5\ -0.3\ 0.8]^T$ if the activation function is the identity.
    b. Find the total loss if, $\mathbf{W} = [0.5\ -0.3\ 0.8]^T$ if the activation function is sigmoid.
    c. Find the total loss if, $\mathbf{W} = [0.5\ -0.3\ 0.8]^T$ if the activation function is the sign function. Use the bipolar perceptron criteria.
    d. Find the total loss if, $\mathbf{W} = [0.5\ -0.3\ 0.8]^T$ if the activation function is the sign function. Use the SVM criterion.
5- Given a data set of RGB colors, {(0, 0, 0), (255, 0, 0), (0, 255, 0), (0, 0, 255), (255, 255, 0), (0, 255, 255), (255, 0, 255), (255, 255, 255)} and their corresponding labels {+1, +1, +1, -1, +1, -1, -1, +1}:-
    a. Using the bipolar perceptron criterion with gradient descent, find and visualize a classification boundary adopting the online training.
    b. Using the bipolar SVM criterion with gradient descent, find and visualize a classification boundary adopting the online training.
    c. Using the bipolar SVM criterion with gradient descent, find and visualize a classification boundary adopting the batch training.
6- Given a data set, $D = \{(\mathbf{X}_1, y_1), (\mathbf{X}_2, y_2)... (\mathbf{X}_N, y_N)\}$ of labelled feature vectors where $\mathbf{X}$ is the feature vector and y is the class label. Note that $y \in \{1, 2\dots m\}$ and m represents the number of categories. A multilayer neural network is used for classification with a soft-max layer.
    a. What is the number of nodes in the input layer?
    b. What is the number of nodes in the output layer?
    c. Write an expression for the loss function using sum of squared differences.
    d. Write an expression for the loss function using the log-likelihood criterion.
7- A neural network is designed to classify the handwritten numerical digits. It has 10 ten output nodes to represent the image categories 0, 1, 2… 9. Assume that the output vector for a certain image input is produced as (3.1, -9.3, 7, 8.7, 3.6, 5.2, 4.7, -2.2, 3.1, -6.6)
    a. Find the outputs of the soft-max layer.
    b. Find the log-likelihood loss if the correct class label is the fourth.
    c. Find the log-likelihood loss if the correct class label is the second.