Deep Learning CS583 Fall 2022
Quiz 1

October 13th

Instructor: Jia Xu

Student name: Dhruv Vaghela

Student ID: 20015603

Student email address: dvaghela@stevens.edu

- **Read these instructions carefully** • Fill-in your personal info, as indicated above.

- You have 24 hours.
- There are three questions. Each question worths the same.
- Both computer-typed and hand-writing in the very clear form are accepted.
- This is an open-book test.
- You should work on the exam only by yourself.
- Submit your PDF/Doc/Pages **by 18:00 Oct 14th** on Canvas.

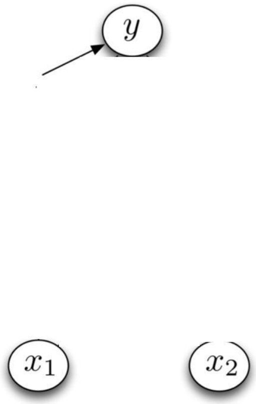
good luck!

1 Question

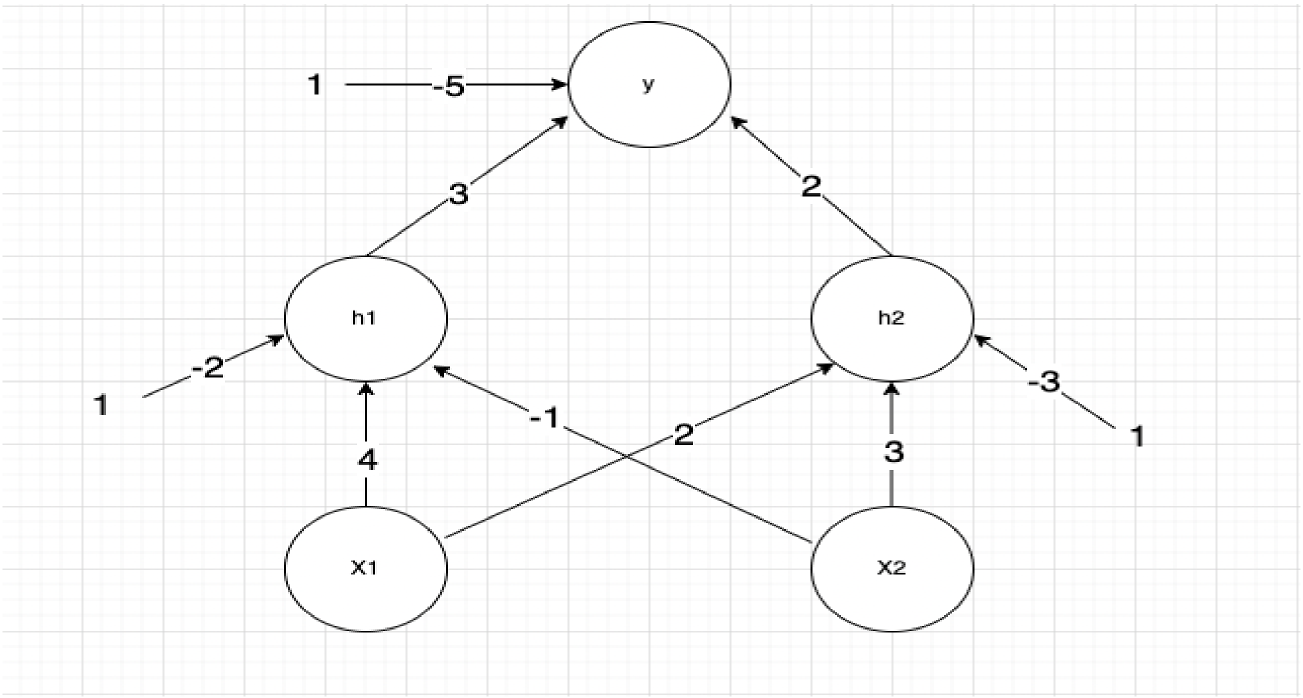
You are given one or several hidden nodes “h“, two inputs x_1, x_2, and the output y. Draw a neural network and assign the weights and bias that perform AND operation:

- if $x_1 = 0$, $x_2 = 0$, then $y = 0$
- if $x_1 = 1$, $x_2 = 0$, then $y = 0$
- if $x_1 = 0$, $x_2 = 1$, then $y = 0$
- if $x_1 = 1$, $x_2 = 1$, then $y = 1$

The activation function outputs 1 if the input is greater than zero and outputs 0 otherwise.



Explanation:



2 Question

- If we have a recurrent neural network (RNN), we can view it as a different type of network by ”unrolling it through time”. Briefly explain what that entails.

  Explanation:
  A Recurrent Neural Network (RNN) is typically a neural net architecture where all the nodes are connected in a different way when compared with a deep neural network or fully connected network. All nodes are connected in such a way that yields a series which computes taking time steps into consideration. This helps in retaining knowledge from the previous node which helps solve intricate problems like Language Translation, English to Spanish, and vice-versa etc.

  If we unroll an RNN through, we get a directed graph which computes the weights and bias on each time step. So each node acts as a time step in the series, and this helps in the retention of input as knowledge which will be used for computing the next nodes in the network. In a traditional neural network, each node is connected to the next nodes in the hidden layer or output layer. Here computing means a set of arithmetic matrix operations which are performed in an ideal implementation of RNN.

- Briefly explain how “unrolling through time” relates to “weight sharing” in convolutional networks.

  Explanation:
  Convolution neural networks (CNN) are a class of artificial neural networks, mostly used in applications featuring Image Classification, Object Detection in an image and also in Video/Gesture Classification. This type of neural net is based on the shared-weight architecture of the convolution kernels. Additionally, filters can also be applied that slide

along the input features, analogous to the Sliding Window technique, and provide a feature map. The shared-weight architecture uses the spacial structure of the image or a frame in a video to perform tasks like classification or object detection.

Unrolling across time can be seen in applications of Video Recognition where every frame acts as a time step and the frame's spacial structure can be used to classify the scene/action. In addition, "unrolling across time" can be also seen in the prediction of financial data which surprisingly CNN performs well. In this use case, the data vector for CNN should in 1-dimensional unlike in image classification (both grayscale, and RGB).

Weight sharing and unrolling across time are vital operations for CNN since weight sharing helps in learning intricate patterns and unrolling across time helps in retaining temporal features.

- In a deep neural network or a recurrent neural network, we can get vanishing or exploding gradients because the backward pass of back-propagation is linear, even for a network where all hidden units are logistic. Explain in what sense the backward pass is linear.

  Explanation:
  The goal of backpropagation is to optimize the weights so that the neural network can extract patterns and learns the features. A backpropagation method minimizes the loss function by adjusting the weights and biases of the neural network. In this method, the derivative of the activation function is calculated. This does not affect when there are few layers. But when there are more layers in the network, the value of the product of the derivative decreases until at some point the partial derivative of the loss function approaches a value close to zero, and the partial derivative vanishes. In simpler terms, the derivative of the activation function is a constant. In another case, when the constant is large, the gradient increments to a large value making the neural net unstable.

- Name one solution for the vanishing of the gradients and explain.

  Explanation:
  To avoid the problem of vanishing gradients we can use a different activation function such as Rectifier Linear Unit(ReLU). In this case, if the input value is positive it returns the value, if it

is negative then it returns 0.