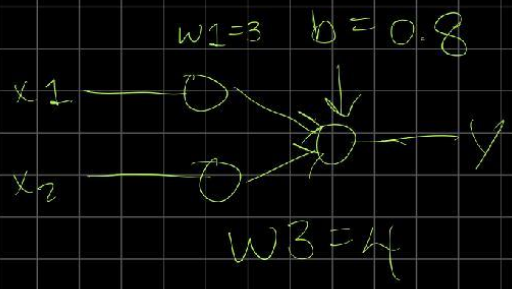
EXTRA TIME - CONFIRM WITH ACCESSIBILITY OFFICE

Question 1.
Which of the following are NOT parameters (based on the definition of a "parameter" from the text): weights, inputs, bias, constant e, learning rate

Question 2.
The logit is a transformation of the logarithm. A regression model, such as logistic regression, is defined as follows: A line is transformed into a logistic curve using the logit transformation. Logistic regression is used to fit a logistic curve to a collection of data in which the dependant variable can only take on the values 0 and 1, which is the case in most situations. It can be applied to ordinal data as well as continuous data.

Question 3.

	x_1	x_2	T
x_a	0.5	1.2	1
x_b	1.2	1	1
x_c	0.3	0.4	0



$$X = \sum_{i=1}^{2} w_i x_i + b = w1 \times 1 + w2 \times 2 + b$$

$$Y = F(X) = \frac{1}{1+e^{-x}}$$

$$y = F(x) = \frac{1}{1+e^{-1}} = 0.99$$

	x_1	x_2	T	Y
x_a	0.5	1.2	1	0.99
x_b	1.2	1	1	0.99
x_c	0.3	0.4	0	0.96

$$\sum_{i=1}^{3} (T_i - Y_i)^2$$

SSE = 0.9218

Question 4

The term "epoch" refers to the time spent training the neural network and all the training data in a single cycle. At some point, we use all of the information at once. Moving forward and backward simultaneously is accounted for as a single pass: The period is divided into one or more clusters, each of which contains a portion of the database used to train the neural network. The process of going through training examples is referred to as repetition. Periods of repetition are interspersed with periods of rest. Let us use a basic example in which we have 1000 data points to explain the idea. Epoch indicates the total number of training databases, algorithm learning machines, and learning machines that have been completed. In most cases, data sets are grouped (especially if the volume of data is excessive). One Epoch is a period during which all data is transferred back and forth solely through the neural network. Because a single season is too huge to be given to a computer simultaneously, we break it into numerous smaller groups of seasons. The number of epochs increases, as does the number of times the weights in the neural network are changed, and the transition from less than the right to a right curve becomes more frequent. This is the epoch during which the authentication error increases since the model is very similar to the training set and does not correctly integrate fresh input into the model. That is the point at which we must halt our training.

Question 5.

$x_a = [0.6, 0.4, 2, 0]$

$x_b = [1, 0.8, 1, 0]$

$x_c = [1.2, 0.7, 1.5, 1]$

$B = 0.8$

$W = [1, 1, 1, 1]$

$$X = \begin{bmatrix} 0.6 & 1 & 1.2 \\ 0.4 & 0.8 & 0.7 \\ 2 & 1 & 1.5 \\ 0 & 0 & 2 \end{bmatrix}_{4\times 3 = A_1 \times B_1}$$

$$W = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}_{4\times 1 = C_1 \times D_1}$$

$$B = \begin{bmatrix} 0.8 \\ 0.8 \\ 0.8 \\ 0.8 \end{bmatrix}_{A_1 \times D_1 = 4\times 1}$$

$$Y = XW + B$$

$$Y = \begin{bmatrix} 0.6 & 1 & 1.2 \\ 0.4 & 0.8 & 0.7 \\ 2 & 1 & 1.5 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0.8 \\ 0.8 \\ 0.8 \\ 0.8 \end{bmatrix}$$

$$Y = \begin{bmatrix} 0.6+1+1.2 \\ 0.4+0.8+0.7 \\ 2+1+1.5 \\ 0+0+1 \end{bmatrix} + \begin{bmatrix} 0.8 \\ 0.8 \\ 0.8 \\ 0.8 \end{bmatrix}$$

$$Y = \begin{bmatrix} 3.6 \\ 2.7 \\ 5.3 \\ 1.8 \end{bmatrix}$$

Question 6.

A computer perceives an image as a series of 0s and 1s. A pixel is the smallest unit of measurement in an image. A digital image is a 2D array of pixels shown on a computer screen. The (x, y) coordinates of each pixel and its value distinguish it from the others. A digital image is saved as a series of pixels arranged in a grid. Each pixel has a different number of channels than the previous one. An image with only one pixel is known as a grayscale image; on the other hand, an image with three channels is known as a colored image and has three channels: red, green, and blue. In an RGB image, the three color planes — Red, Green, and Blue — break the image into sections. There are a variety of color spaces in which images can be found, including grayscale, RGB, HSV, CMYK, and others. ConvNet condenses the images into a format that is easier to interpret while maintaining the crucial elements for making accurate predictions in the first place. When creating an architecture capable of learning features while simultaneously being scalable to large datasets, it is critical to consider this. The Pooling layer, like the Convolutional Layer, is responsible for lowering the spatial size of the Convolved Feature, which is similar to the Convolutional Layer. Dimensionality reduction must be implemented to reduce the number of computer resources necessary to process the data. Combined, the Convolutional Layer and the Pooling Layer make up Layer I of a Convolutional Neural Network, represented by the letter "i." Depending on the complexity of the images, the number of such layers may be expanded even higher to capture even more low-level information, but this will come at the expense of additional processing power.

Question 7

K-Means clustering is an unsupervised learning approach used to group data. In contrast to supervised learning, there is no labeled data for this grouping to work with. It is performed using K-Means clustering, which involves grouping things into clusters with similarities but distinct from the objects belonging to another cluster. Afterward, the K centroids are updated to reflect the assignment of all objects. Centroids migrate closer to the collection of formerly their own objects, reducing the distance between each point and its centroid even further.

Question 8.

The relative placements of data points do not change in this scenario. Principal components are linearly independent of one another. The primary components of the original dataset are linear combinations of the characteristics of the original dataset. Correlation is the statistical link that exists between two variables. A correlation can be positive, meaning that both variables move in the same direction, or negative, meaning that when the value of one variable grows, the values of the other variables fall. Statistical analysis (PCA) does not actually increase the variance in your data. Instead, it rotates the data set so that the directions in which it is most widely dispersed are aligned with the primary axes of the model. Using this technique, you can eliminate those dimensions along which the data is almost flat. Example is height and weight.

Question 9.

This type of representation is referred to as a one-hot encoding since there is only one index that has a non-zero value in this case. A vector is more commonly used to hold counts of the words in a larger chunk of text. This type of depiction is a "bag of words." A few examples of nonlinear systems under investigation today include fluid turbulence, combustion, and the biochemical interactions in living creatures, to name just a few of the numerous systems under investigation. An activation function aims to inject non-linearity into a neural network to improve performance. The outputs of neurons are controlled by the activation functions, which perform their purpose. Rather than learning a specific pattern, an activation function influences how a neural network learns various patterns in a given time period. Each

activation function has a different effect on the output than the others. Accordingly, avoiding several activation functions in a single layer would be more predictable in their effects.

Question 10.

The activities take place during the forward pass, which refers to calculating the values of the output layers based on the data received from the inputs. It's also making its way across all neurons, from the first to the last layer of the brain. In order to determine the loss function, the output values are multiplied together. The approach is used to efficiently train a neural network by using a technique known as the chain rule method. To put it another way, following each forward run through a network, backpropagation executes a backward transit across the network while modifying the model's parameters. The concept of function composition is at the heart of compositional learning, and it states that a single model cannot do all of the tasks. Deep neural networks did exceptionally well in a single task, such as classifying an image into a cat or dog or recognizing cancer.

Epoch - LogicPlum. https://logicplum.com/knowledge-base/epoch/
Implementation - Digital Design Standards.
http://digitaldesignstandards.com/standard/process/implementation/
How Does Computer Understand Images? | by Sabina Pokhrel ....
https://towardsdatascience.com/how-does-computer-understand-images-c1566d4537bf
Lab - Deploying a Predictive Model - Azure AI Gallery.
https://gallery.azure.ai/Experiment/Lab-Deploying-a-Predictive-Model-2
Three-Dimensional Separation and Characterization of ... - Frontiers in.
https://www.frontiersin.org/articles/10.3389/feart.2020.529263/full
5 Deep Learning Activation Functions You Need to Know.
https://builtin.com/machine-learning/activation-functions-deep-learning