Aaliya Hussain
2nd Period
5/5/20

**Backpropagation Classwork**

1. Backpropagation is an algorithm for training perceptron networks based on minimizing an error function through repeated gradient descent. The error function can be represented as a multivariable function of the weights and biases of the perceptrons in the network, which allows us to take the partial derivative of each variable (i.e. weight/bias) in succession, starting from the layer just before the output layer. By applying the process to all input/output pairs in the training set, the network will eventually come to accurately model the function based on the data given.
2. Just as we can train individual perceptrons by continuously changing their weights and biases until they classify all inputs correctly, we can train networks by altering all the weights and biases in every layer using backpropagation. The error of a network can be modelled as a multivariable function of all the different weights and biases, which allows backpropagation to minimize it through a series of gradient descents.
3. A network is like a mathematical function that uses many variables and calculations to "answer" a "yes-or-no" question about a certain input. However, we do not know which variable values would make the network as accurate as possible. Backpropagation is the process we use to find those ideal values, and it depends on making the error of the network as small as possible in the shortest possible time.
4. I strongly agree with her. Backpropagation is essentially repeated derivation represented in matrix form. Its complicated, intricate nature does not come from its underlying mathematical operations. Rather, it is complex due to its repeated, multilayered structure and its dual forms as both gradient descents and matrix operations. To me, the greatest challenges presented by backpropagation are remembering to fully apply the chain rule, keeping track of all the different symbols used, and understanding matrix equivalents to calculus operations. The underlying math of derivations, gradient descents, and minimization is actually fairly simple .
5. Gradient descent is the central mathematical operation involved in backpropagation. In multivariable calculus, gradient descent allows for minimization of a function through calculation of partial derivatives. Because the error function of a network can be represented as a multivariable function of all the weights and biases of the individual perceptrons, repeated gradient descent can be used to minimize the error of the network by adjusting each weight and bias.
6. The transpose operation, which "flips" a matrix in terms of rows and columns, accounts for the differences in matrix sizes throughout the network. Each layer is represented with different sizes of weight and bias matrices because the number of rows and columns depends on the number of nodes in the current and next layers, respectively. As backpropagation proceeds from the layer closest to the output layer, the differing sizes do

not always allow for straightforward calculation of dot products. Rows and columns must be switched through transpose operations to ensure proper matrix sizes.

7. $\Delta_4 = A`(dot_4) * (\Delta_7 * w_{4,7} + \Delta_8 * w_{4,8} + \Delta_9 * w_{4,9})$