Instructions: Make sure your questions and answers are on different pages. Do not include your name or any other identifying information. I will know that information from Canvas.

Question 1: What are the tradeoffs between using gradient descent vs Newton's method to optimize a neural network?

Question 2: Why is gradient descent guaranteed to converge to the unique global optimum for convex functions?

Question 3: What is one difference between a convex function and a strongly convex function?

Answer Question 1: Gradient descent may converge more slowly than Newton's method, since it relies on only the first order gradient of the function while Newton's method utilizes the second order gradient as well in its optimization. However, Newton's method must compute the entire Hessian matrix of second partial derivatives, which can be very computationally and spatially complex with large datasets especially.

Answer Question 2: By definition, a convex function's second order gradient will be strictly greater than or equal to zero, such as in the case of f(a,b) = a^2 + b^2. This kind of function can never have more than one local minimum, as a negative gradient would be required at some point between two local minima. Since gradient descent is already guaranteed to find a local minimum, and a convex function's global minimum is its only local minimum, gradient descent will be able to find it regardless of the starting point.

Answer Question 3: The Hessian of either a convex or strongly convex function is a square matrix of that function's second order gradients. For a convex function this Hessian must be positive semidefinite, meaning that its eigenvalues must be greater than or equal to zeros. A distinction between the two is that the Hessian of a strongly convex function must be positive definite, meaning that its eigenvalues must be greater than zero. In other words, a convex function can have second order partial derivatives equal to zero while a strongly convex function must have all of its second order partial derivatives explicitly greater than zero.