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 most commonly used optimization methods?

Question 2: Why is Newton's method not as popular as gradient descent if it is more efficient?

Question 3: Are there techniques used to prevent a gradient descent algorithm from becoming stuck at a local minimum and never able to reach a global minimum?

Answer Question 1: The most popular iteration methods are Gradient Descent Algorithm and its variants, Stochastic Gradient Descent, and the MiniBatch Gradient Descent.

Answer Question 2: Newton's method is much more constrained as it relies on the presence of a defined second derivative. Many functions do not have this, leaving gradient descent to be "safer".

Answer Question 3: Stochastic gradient descent uses randomness and analyzes results for every observation in the training process, rather than waiting until the end. This helps the algorithm improve its performance by decreasing the likelihood of getting stuck in one of the local minimums.