Question 1: What is one fundamental difference between gradient descent and Newton's method?

Question 2: How do you define convex functions?

Question 3: Name one loss function that results in a convex function to optimize for linear regression. Hint: it is one of the simplest loss functions.

Answer 1: Newton's method uses a matrix of second order derivatives while gradient descent only uses first. This has the effect of of causing Newton's method to converge quicker as it is a second order optimization.

Answer 2:

Convex functions. A function $f : \mathbb{R}^d \to \mathbb{R}$ is convex if for all $x, y \in \mathbb{R}^d$, and all $\eta \in [0, 1]$

$$f(\eta x + (1 - \eta)y) \leq \eta f(x) + (1 - \eta)f(y).$$

In plain English it is really just a U-shaped function in which if you drew a line between any to points the line will lie above the function.

Answer 3: Least squared error:
It is convex for linear regression but not for logistic regression.