Question 1: Difference between Newton's method and gradient descent?
Question 2: How does gradient descent work?
Question 3: What is the condition number of the problem?

Answer 1: Newton's method converges faster than gradient descent. This is because Newton's method uses the second derivative of f while gradient descent uses the first derivative.

Answer 2: Gradient descent follows the direction of the steepest descent of the function, which means it moves towards the minimum of the function. At each iteration, it updates the parameters in the opposite direction of the gradient of the function with respect to the parameters.

Answer 3: The condition number of the problem "encodes how hard a strongly convex problem is to solve." Large number means it is hard to solve.