

1. Why does Newton's method converge faster than gradient descent?
2. Given a constant  $L > 0$  for all  $x$ , what constraint does the step size  $\alpha$  have in order to guarantee to decrease at each iteration?
3. What does it mean for a function to be convex?

1. Newton's method is a second-order optimization method which gradient descent is a first-order optimization method.
2. Step size  $\alpha$  must be small enough such that  $1 \geq \alpha L$ .
3. A function is convex if a line segment between any two points in the function will lie above the graph of the function.