Question 1: In graphical terms, what is the meaning of a convex function?

Question 2: What is the problem with starting with too high of a learning rate in gradient descent? What about too low?

Question 3: When discussing gradient descent, what do we take the “gradient” of and what does “gradient” mean?

Answer Question 1: A convex function is a function where if one where to draw a line segment between any two points in the function graph, this line segment would lie above the function.

Answer Question 2: Starting with too high of a learning rate may result in gradient descent skipping over more optimal minima. Whereas starting with too low of a learning rate may make gradient descent converge too slowly or get stuck at local minima.

Answer Question 3: In gradient descent we are taking the gradient of the loss function with respect to the model parameters. The gradient is simply a vector of partial derivatives of the model parameters (the weights in neural networks).