Question 1: What are some examples of hypotheses and loss functions that result in a convex objective?

Question 2: What happens to our learning process using Gradient Descent when we decrease, or increase the learning rate?

Question 3: What is the main difference between gradient descent and Newton's method? Can we rewrite the gradient descent update rule to obtain Newton's update rule?

Answer Question 1:

The quadratic functions, including x^2, x^4, result in the convex objective. These functions have the shape of a cup, such that when drawing a line segment between any two distinct points, that line segment will lie above the function. Therefore, loss functions that can be represented as quadratic formulas, such as squared loss result in the convex objective

Other loss functions that result in the convex objectives are MSE, Huber Loss, and logarithmic loss

Answer Question 2:

When we increase the learning rate, our gradient descent converges faster. However, if the learning rate is too high, we may step over the global minimum of the objective function we try to minimize

When we decrease the learning rate, our gradient descent converges slower, but we are more likely to approach the global minimum of the objective function without stepping over it.

Answer Question 3:

Newton's method is a second-order optimization method (it uses the second derivative of f) whereas gradient descent is only a first-order method (it uses the first derivative of f). We can rewrite the gradient decent update rule to obtain Newton's update rule:

$$w_{t+1} = w_t - \alpha_t \nabla f(w_t) \quad where\ \alpha_t = \nabla^2 f(w_t)$$