Consider the model:

$$y = \theta_1 e^{-\theta_2 x} + \theta_3$$

and the data points

$$\{(0,\ 2.300),\ (1,\ 1.036),\ (2,\ 0.571)\}$$

(a) Define your cost as the square of the l_2 norm and write a function which takes vectors x, y, and θ and returns the cost.

(b) Use an automatic differentiation library of your choice to take the gradient of your cost function

(c) Calculate the normalized negative gradient ($\hat{g}$) at $\theta = \begin{pmatrix} 1.5 \\ 0.6 \\ 0 \end{pmatrix}$

(d) Plot $L(\theta_0 + \alpha\hat{g})$ where L is your loss function, θ_0 is the theta vector given in part c, $\hat{g}$ is the normalized negative gradient from part c, and α is a scalar with values ranging from 0 to 1.

(e) Using the same y-axis, plot $L(\theta_0 + \alpha\hat{d})$ where $\hat{d} = \begin{pmatrix} 0.138 \\ 0.983 \\ 0.118 \end{pmatrix}$.

(f) Using your two plots, explain why the gradient is useful in the context of modeling.