1 True or False

Circle True or False if the statement is, well, true or false.

1. (True / False) : For any given (valid) model, data, loss function, and initial parameters, there exists a learning rate which will cause gradient descent to reach the global minimum loss.
2. (True / False) : It is possible for two identical models, trained on the same data, but using different loss functions, to find the same optimal parameter values.
3. (True / False) : The number of learning steps (Gradient Descent steps) a model takes to reach an optimal solution is affected by the initial values of its parameters.

2 Written response

1. Imagine we have data that has very high and very low magnitude examples (i.e. $x^{(1)}$, $y^{(1)} = 0.001, 0.02$ and $x^{(2)}$, $y^{(2)} = 180, 540$). Normal loss functions would disproportionately learn from the higher magnitude data, but would ignore the low magnitude data. Come up with a loss function such that a model that minimizes the loss will have low *proportional* difference in predictions and labels in both cases:

$LOSS(\hat{y}, y) =$

3 Computation Graph and MATH

1. Draw the computation graph of the following function : $y = \frac{x-\mu}{\sigma}$ (this is normalization) where x has two features (i.e x = $[x_1, x_2]$) and :

$$\mu = E[x] = \frac{1}{N}\sum_{i=0}^{N} x^{(i)} \quad \text{and} \quad \sigma = std(x) = \frac{1}{N}\sum_{i=0}^{N}(x^{(i)} - \mu)^2$$

2. (BONUS) Use your computation graph to calculate dy/dx (this is hard I know, but give it a try if you have time) :