Local Minima Problem

1. Introduction about local minima problem
2. How local minima confuse the ANN

1. Introduction about Local Minima problem:

- The local minima point in the loss function is one of the problems while updating the weights. The local minima points confuse ANN and ANN will treat the local minima as a global minima.
- Till now we consider the loss function graph as a normal bell curve as shown in the fig-1. This graph has only one global minima point.
- But in real world scenario the curve looks like fig - 2 with local minima, local maxima, global minima and global maxima points.

2. How Local Minima Confuse the ANN:

Now let's understand how ANN treat the local minima as global minima due to tangent confusion:

Consider weight updating formula.

$$\omega_{new} = \omega_{old} - \alpha \frac{\partial L}{\partial \omega_{old}}$$

Here the loss function is.

$$L = \sum_{i=1}^{k} (y_i - \hat{y}_i)^2$$

- The meaning of loss derivatization $(\frac{\partial L}{\partial \omega_{old}})$ is, calculating the tangent of the loss function. This tangent gives the slope of loss function at a particular point.
- Which means if the graph doesn't have any slope at any point that place treated as a global minima. This is actual working of gradient descent optimizer.
- But if we observe the slope of local minima points also zero, Due to this reason ANN takes this local minima as a global minima.

At local minima point 1 the slope is zero means $\frac{\partial L}{\partial \omega_{old}} = 0$.

Then $\omega_{new} = \omega_{old} - \alpha\,(0)$

$\omega_{new} = \omega_{old}$

- ANN stops the training and considering global minima got reached whenever the weights ω_{new} and ω_{old} are same.

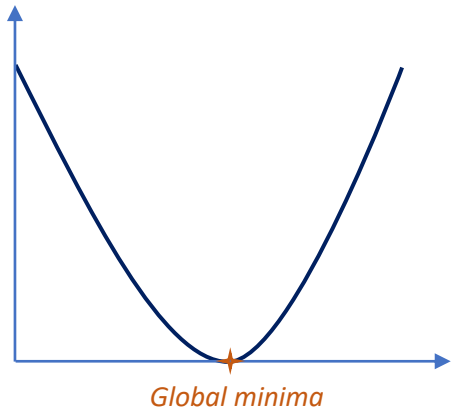

Figure 1 - Convex function

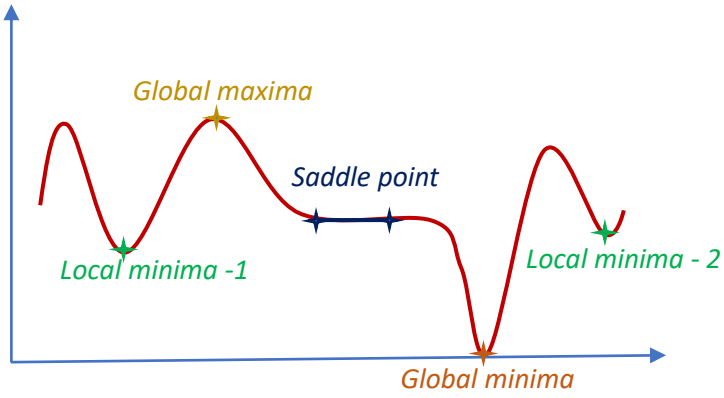

Figure 2 – Non-Convex function

- The saddle point is also a one of the tangent problems, we can see it in fig-2. The ANN treats this saddle point as global minima, because tangent of saddle point is 0.