Problem that can occur with complex deep neural network

### 1. Problem : Local Minima

- Problem: cannot find the lowest point; Gradient descent in this case cannot find the actual lowest point.
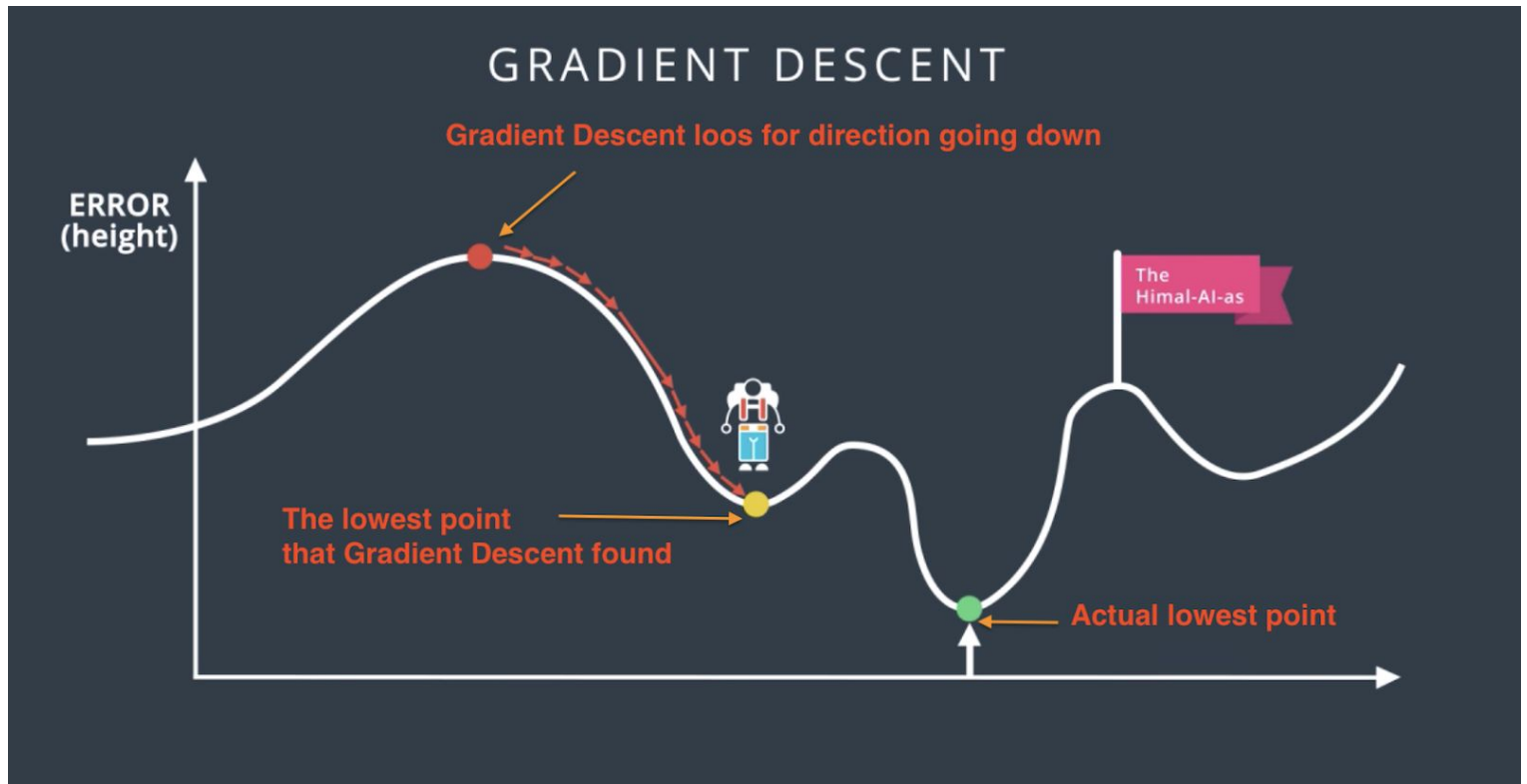
- What could be the solution ?

Random Restart
- One of the solutions to the local minima issue
- Choose a few different random places and do gradient descend from all of them, it helps increase probability of getting local minimum OR at least a pretty good local minimum
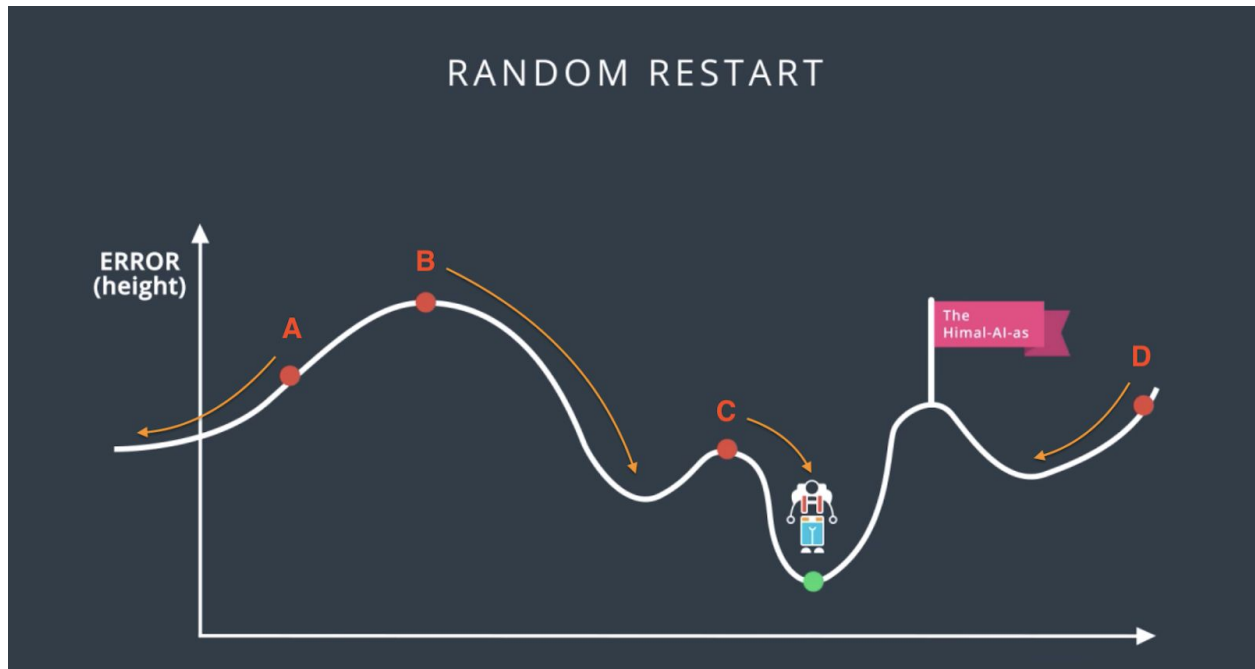

- A, B, C, D are the few points we chose to be random starts

2. Problem: Vanishing Gradient
- Sigmoid function has pretty curve flat so if we calculate point at way left or way right, this derivative is almost zero. Why is it an issue ?
- Derivative tell us the direction to move.
- In perceptron, it means the derivative sigmoid function is small; and the product of these small value leads to tiny values.
- This makes the training difficult because grading give us very tiny changes to make on the weights

- **Solution to vanishing gradient**
  - **Choose other activation functions**
    - Hyperbolic Tangent function
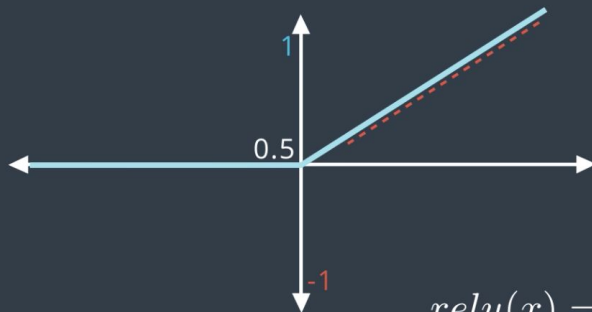

HYPERBOLIC TANGENT FUNCTION
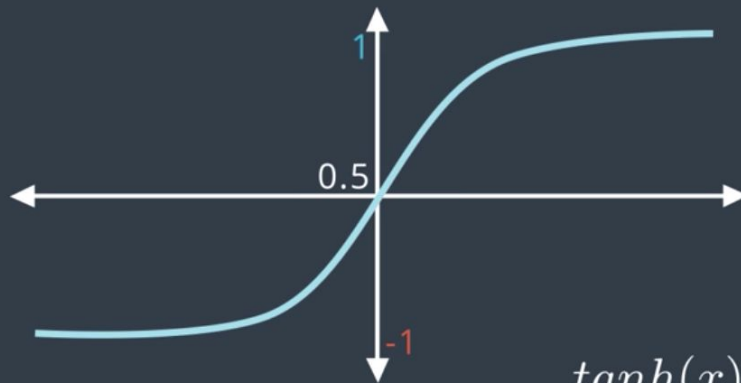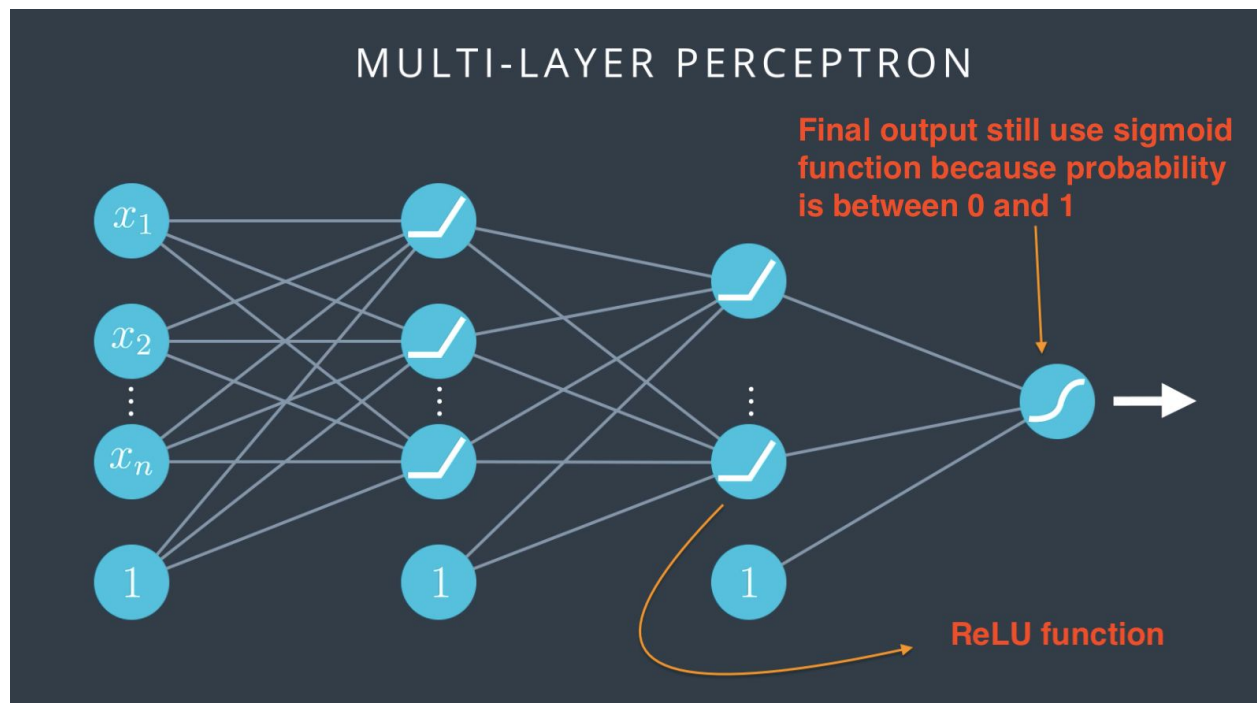
$$tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

The range is from -1 and 1, so the derivatives are larger
    - Rectified Linear Unit (ReLU)
- Return same value if the value is positive
- Return 0 if the value is negative

RECTIFIED LINEAR UNIT (ReLU)

$$relu(x) = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

==

Mini Summary

- Problems occur in deep neural network
    - Local minima
    - Vanishing gradient
- Solution for local minima: random start
- Solutions for vanishing gradient
    - Hyperbolic Tangent function
    - ReLU function