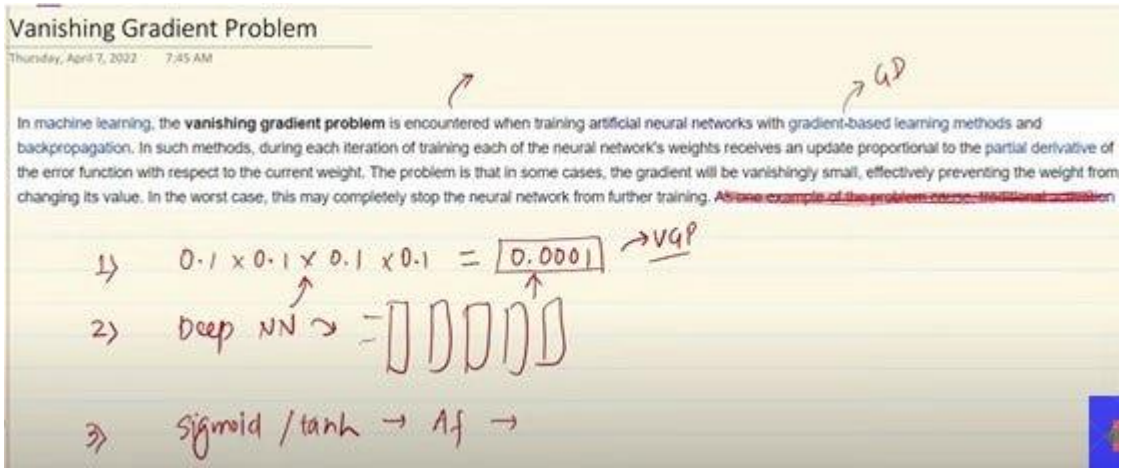
Vanishing Gradient Problem

eep Learning has rapidly become an integral part of modern AI applications, such as computer vision, natural language processing, and speech recognition. The success of deep learning is attributable to its ability to automatically learn complex patterns from large amounts of data without explicit programming. Gradient-based optimization, which relies on backpropagation, is the primary technique used to train deep neural networks (DNNs).

Understanding the Vanishing Gradient Problem

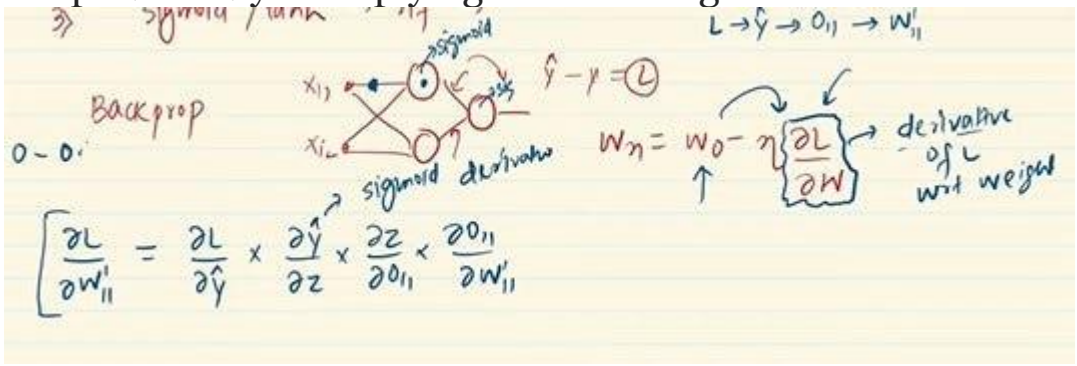
One of the major roadblocks in training DNNs is the vanishing gradient problem, which occurs when the gradients of the loss function with respect to the weights of the early layers become vanishingly small. As a result, the early layers receive little or no updated weight information during backpropagation, leading to slow convergence or even stagnation. The vanishing gradient problem is mostly attributed to the choice of activation functions and optimization methods in DNNs.



Vanishing gradient problems generally occurs when the value of partial derivative of loss function w.r.t. to weights are very small.The complexity of Deep Neural Networks also causes VD problem.

Role of Activation Functions and Backpropagation

Activation functions, such as sigmoid and hyperbolic tangent, are responsible for introducing non-linearity into the DNN model. However, these functions suffer from the saturation problem, where the gradients become close to zero for large or small inputs, contributing to the vanishing gradient problem. Backpropagation, which computes the gradients of the loss function with respect to the weights in a chain rule fashion, exacerbates the problem by multiplying these small gradients.



During backpropagation we calculate loss (y-y_hat) and update our weights using partial derivatives of loss function but it follows chain rule and to update w11 there will be a sequence of chain with multiplication of smaller values of gradient descent and learning rate which in result no change in weights .

How to recognise Vanishing Gradient Problem

1. Calculate loss using Keras and if its consistent during epochs that means Vanishing Gradient Problem.

2. Draw the graphs between weights and epochs and if it is constant that means weight has not changed and hence vanishing gradient problem.