NeuralNetwork

lucassauvage

2024-

Contents

Definitions

Machine Learning : Machine learning is a field of study in artificial intelligence that aims to give machines the ability to "learn" from data, via mathematical models. More precisely, it is the process by which relevant information is drawn from a set of training data.

Loss Function : A loss function is a mathematical function that measures the difference between the predicted output of a machine learning model and the actual output.

1 Introduction

This document summarizes all the mathematics used in my Neural Network project. This project is an AI currently trained to recognize numbers. Coding was not the hardest part of this project, as there is a lot of code already available online. The challenge lies in understanding the mathematics required to build a program that can learn on its own.

2 Gradient

2.1 Definition

Gradient is a key concept in optimization and machine learning, particularly in neural network training. It represents the direction and rate of change of a function relative to its inputs

2.2 Why are we using it

In machine learning, particularly in neural networks, the aim is to minimize a loss function $L(\theta)$ where θ represents the model parameters (weight and bias). The gradient provides essential information for adjusting parameters to reduce loss. The gradient of the loss function $\Delta L(\theta)$ indicates how the loss changes in relation to small variations in θ By following the negative gradient (the direction of fastest descent), we can adjust θ to minimize the loss.

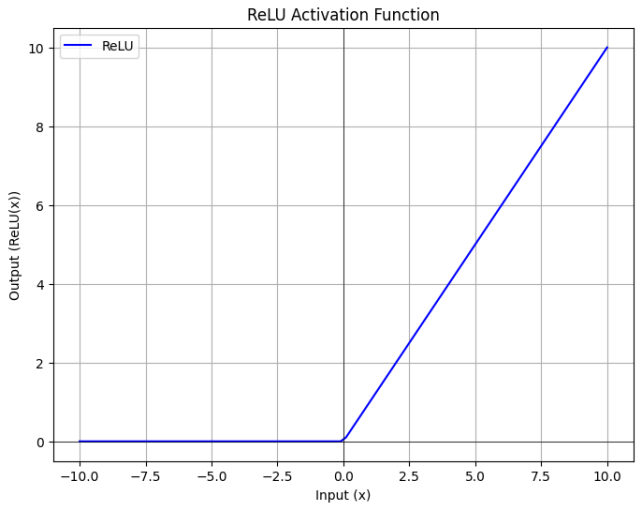
3 ReLu Function

3.1 Definition

The ReLu (Rectified Linear Unit) function is a piecewise linear function that will output the input directly if it is positive, otherwise, it will output zero. This function introduces the property of nonlinearity to a deep learning model and solves the vanishing gradients issue. The formula is :

$$\text{ReLU}(x) = \max(0, x)$$

Here's the graph for the ReLu function:



3.2 Why are we using it ?

- Simple function : it returns the input value if it is positive and 0 otherwise. Unlike other activation functions such as sigmoid or hyperbolic tangent, ReLU does not involve expensive exponential calculations.
- ReLU activates only a portion of neurons because, for all negative inputs, the output is 0. This introduces a form of sparsity in the activations, which improves computational efficiency and reduces the risk of overfitting.

Despite its advantages, ReLu function has some drawbacks :

- Dead neurons: When a neuron consistently receives negative values as input, its gradient becomes 0, and it stops learning
- Non-differentiable at 0: The point $x = 0$ is a discontinuity.

4 Cross Entropy

The cross entropy between two probability distribution p and q over the same underlying set of events, measures the average number of bits needed to identify an event drawn from the set when the coding scheme used for the set is optimized for an estimated probability distribution q, rather than the true distribution p

4.1 For discrete probability

For discrete probability distribution p and q this means :

$$H(p, q) = -\sum_{x} p(x) \log q(x)$$

$H(p, q)$ is often used for the cross entropy and the joint entropy of p and q.

4.2 Cross-entropy minimization

Cross-entropy is often used in the maximum likelihood framework. In my project, cross entropy's used to for the loss function. Using the cross-entropy formulation between the real (or "true") data distribution p and the estimated distribution q, we actually minimize the cross-entropy to find the optimal parameters. Minimize cross entropy means maximising model likelihood.

5 Sources

ReLu :https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/

Cross Entropy : https://en.wikipedia.org/wiki/Cross-Entropy