Basic Mathematics for Machine Learning

In Machine Learning we do predictions

You have a set of inputs and you get an output

All machine learning models can be represented as functions

$$f(x) = y$$

Real-world relationships are:

- Non-linear
- High-dimensional
- Complex

Neural networks are stacked functions

Each layer transforms the data

Role of Activation Functions

Without activation functions, the network becomes a big linear function

Linear models cannot learn complex decision boundaries

Activation functions introduce non-linearity

Activation functions add intelligence to the model

Types of Activation Functions

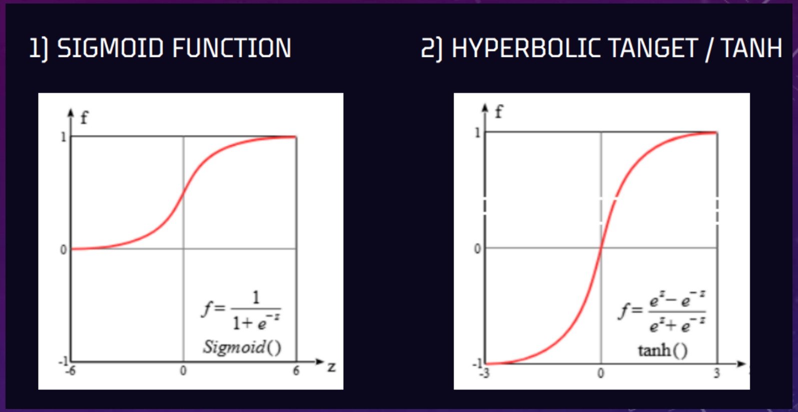
Sigmoid Function

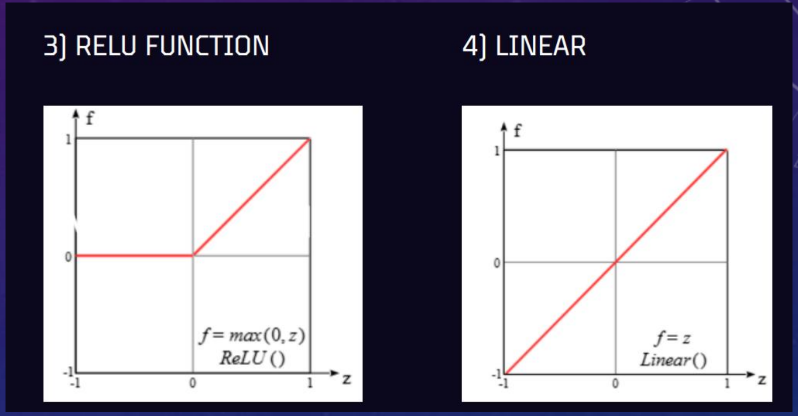
Hyperbolic Tangent (Tanh)

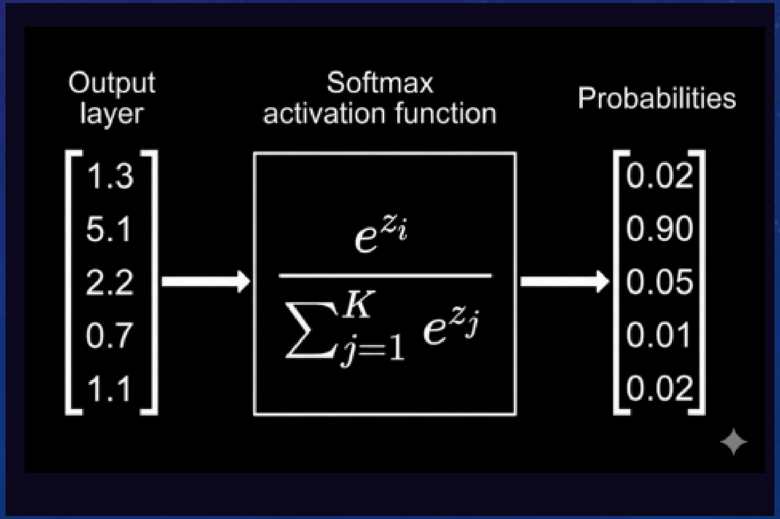
ReLU Function

Linear Function

Softmax Function







Loss Functions

Prediction is denoted as $\hat{y}$

Objective is to minimize the difference between $y$and $\hat{y}$

Regression Losses

Mean Absolute Error (MAE)

Mean Squared Error (MSE)

MAE:

- Uniform gradient
- Gradient does not exist at zero

MSE:

- Big errors lead to large loss
- Smooth function
- Gradient exists everywhere

Classification Losses

Binary Cross Entropy

Categorical Cross Entropy

Used when output is probability

BINARY CROSS ENTROPY

$$L = -[y\log(p) + (1-y)\log(1-p)]$$

CATEGORIAL CROSS ENTROPY

$$L = -\sum_{i=1}^{K} y_i \log(p_i)$$

Yi =1 when i is the coreect class and zero when it is wrong

Type	Prediction	Best Loss	Why
Regression	Numbers	MSE, MAE	Measures numeric distance
Binary Class	0/1 probability	BCE	Models uncertainty
Multi-Class	Prob. distribution	CCE	Counts probability error

Convolutional Neural Networks (CNN)

CNN stands for Convolutional Neural Networks

Fully connected networks flatten input into 1D vector

Large images require huge number of parameters

Convolution reduces number of parameters

Properties of Convolution

Translation Invariance

Locality

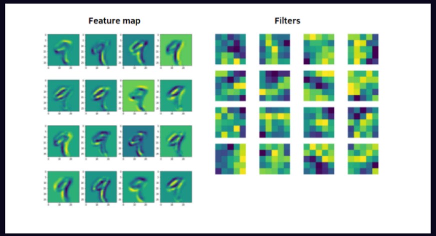
Convolution Definition

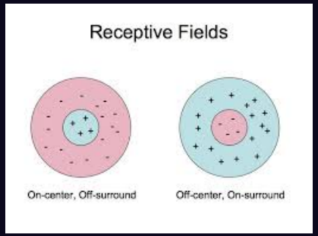
Convolution measures overlap as one function is flipped and shifted

2D convolution is used for images

Feature map: A feature map is the spatial output of a convolutional layer, where each channel represents the activation of a learned feature across image locations.

Receptive field: The receptive field of a neuron is the region of the input (across all previous layers) that can influence that neuron's activation.





Convolution Terminologies

Feature Map:

Spatial output of convolutional layer

Receptive Field:

Region of input influencing a neuron

Padding:

Used to maintain output size

Stride:

Number of pixels kernel moves

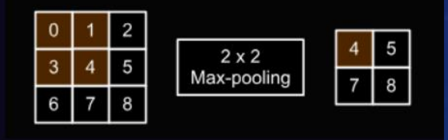
Pooling:

Used to remove unnecessary information

Max Pooling

Average Pooling

Max Pooling : $Y_{i,j} = \max(X_{i:i+p,\, j:j+p})$



Average Pooling : $Y_{i,j} = \text{mean}(X_{i:i+p,\, j:j+p})$

Sequence Modeling

Sequence data is not independent

Output depends on previous inputs

Examples:

- Language modeling
- Stock prices
- Medical records

Auto-Regressive Models

Prediction depends on past observations

T-th order Markov Model considers last T values

Recurrent Neural Networks (RNN)

RNNs process sequences step by step

RNNs maintain hidden states

Hidden state stores past information

RNN equation:

$$h_t = f(x_t, h_{t-1})$$

Problems in RNN

Vanishing gradient

Exploding gradient

Modern RNNs

Long Short-Term Memory (LSTM)

Modified form of RNN

Designed to handle long-term dependencies

Uses memory cell

Uses gates:

- Forget Gate
- Input Gate
- Output Gate

Gated Recurrent Unit (GRU)

Simplified version of LSTM

Fewer parameters

Uses:

- Update gate
- Reset gate

Bidirectional RNN

Processes sequence forward and backward

Uses past and future context

Encoder–Decoder Architecture

Used for sequence-to-sequence tasks

Encoder processes input sequence

Decoder generates output sequence

Used in machine translation