## horizontal line

GRU (Gated Recurrent Unit)

28.05.2025

# GRUs

# LSTM Vs GRUs

* LSTM have 3 gates (forget , input , output) while GRUs have only 2 gates (reset gate, update gate).
* LSTM have more parameters while GRUs have less.
* LSTM have complex Architecture taking more training time while less in GRUs.
* LSTM has 2 states cell state and hidden state while GRUs have only one Hidden state.

# The Architecture

The Set up has vectors , pointwise operations and neural network layers same as of LSTMs.

Rule is same for GRUs as well.

Here the vectors are Ht-1 , Xt , Ht , rt , zt and ht\_bar.

Xt can be of any dimension while all others are of same dim.

Main Steps in the Processing :

1. Calculate rt (reset gate)
2. Calculate ht\_bar (Candidate hidden state)
3. Calculate zt (update gate)
4. Calculate ht (current hidden state)

# What exactly is hidden state ?

Ht-1 —---> rt —----> ht\_bar —----> zt —----> ht.

Hidden state (STM) is memory of the system that takes context either as some words like as we did for the story [Power , Conflict , Tragedy , Revenge].

How GRUs maintain this…Initially we feed only 3 lines on what basis Ht-1 is [0.5 0.6 0.7 0.1]. While giving Xt as next line , rt is used to update it [0.8 0.4 0.5 0.3] as ht\_bar.

Because it found to be inclined more towards Xt , So to make balance between them zt is used.

If Xt is important we prefer more Ht\_bar and if not we prefer more Ht-1.

# Calculating the Reset gate

rt = (Wr [Ht-1, Xt] + Br)

rt = (3X7)(7X1) + (3X1)

rt = (3X1) + (3X1)

rt = (3X1)

Rest gate basically takes probability of what percent of context should be maintained in Ht-1.

E.g rt = [0.8 0.2 0.1 0.9] means 80% of Power , 20% of conflict and so on. This forms resetted Hidden State.

Ht\_bar = tanh( Wh [Ht-1 rt , Xt] + Bh )

Ht\_bar = (3X7)(7X1) + (3X1)

Ht\_bar = (3X1) + (3X1)

Ht\_bar = (3X1)

# Calculating the Update Gate

zt = (Wz [Ht-1, Xt] + Bz)

zt = (3X7)(7X1) + (3X1)

zt = (3X1) + (3X1)

zt = (3X1)

Zt updates the ht-1 on the basis of how much of percent Xt is important.

Ht = ((1-Zt) Ht-1 ) ( Zt Ht\_bar )

Ht = (3X1)