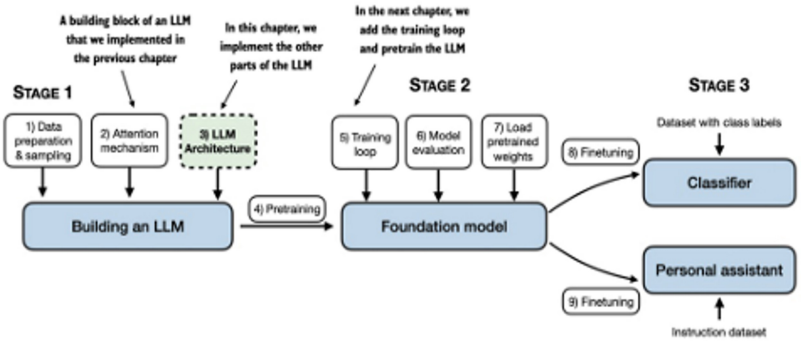
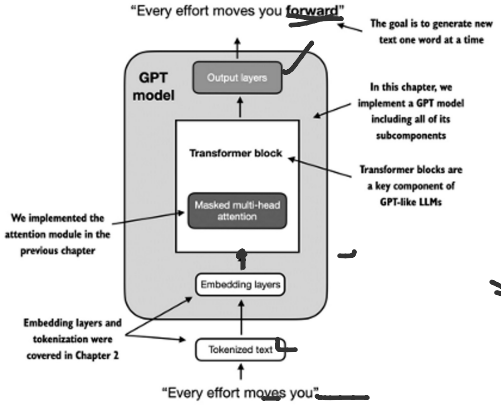
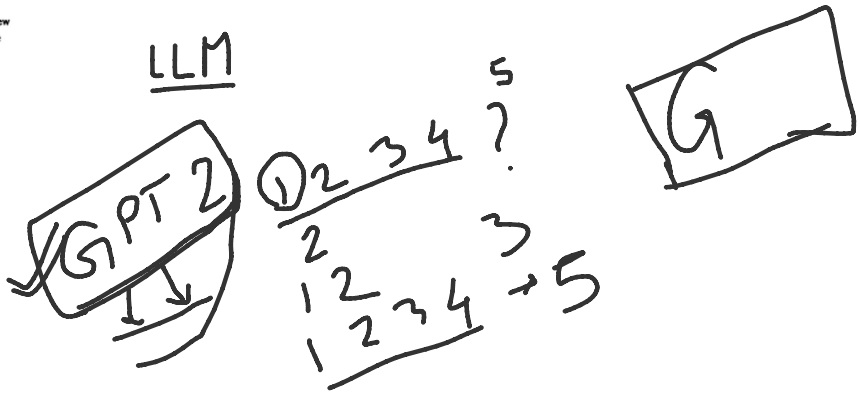
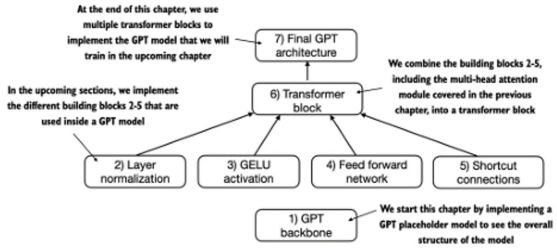
LLM Architecture (GPT Model)

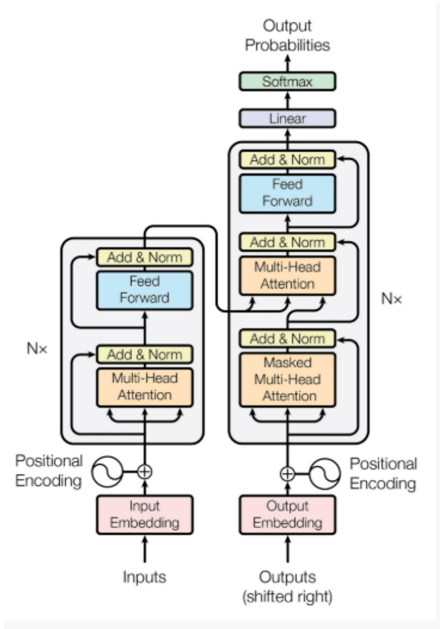
25 फरवरी 2025 22:22

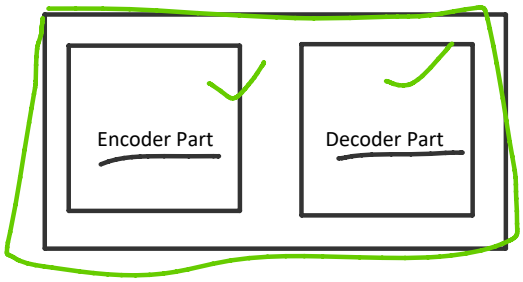




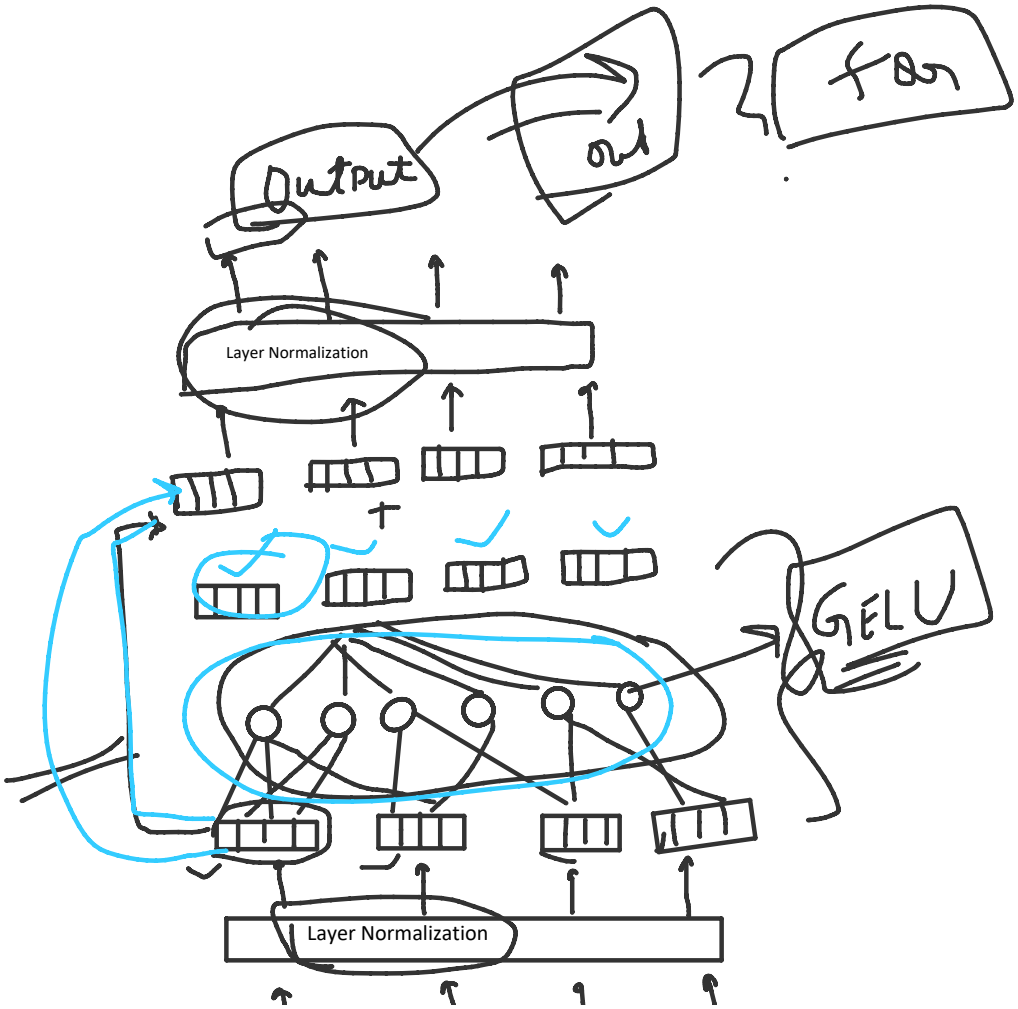


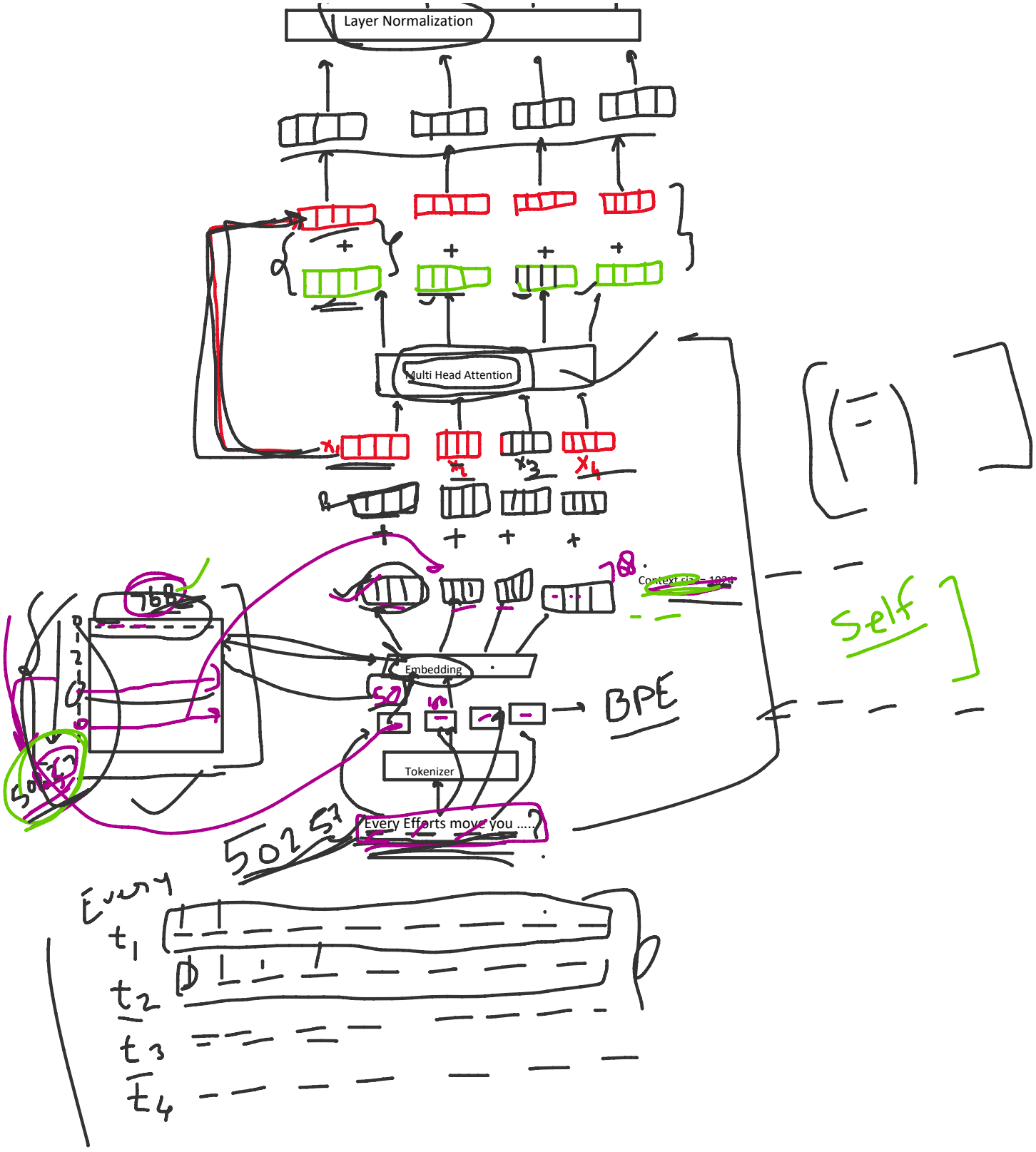



Output Probabilities
Softmax
Linear
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
Add & Norm
Masked Multi-Head Attention
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
Nx
Nx
Positional Encoding
Positional Encoding
Input Embedding
Output Embedding
Inputs
Outputs (shifted right)


Encoder Part
Decoder Part


ENCODER1
Decoder1
layers
6

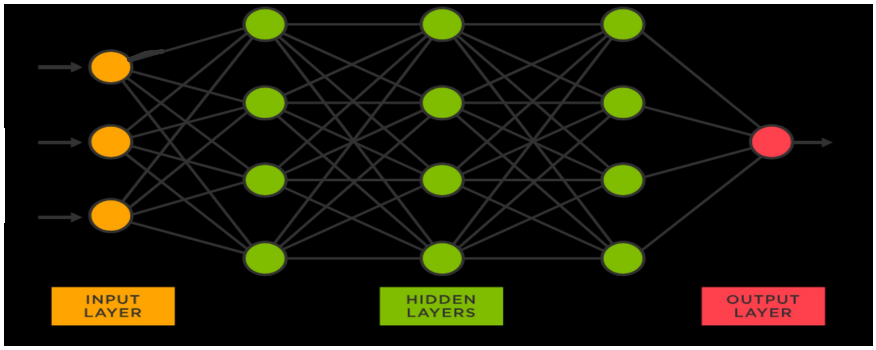

Output
out
for
Layer Normalization
GELU
Layer Normalization



Agenda : LLM Architecture

Understanding Layer Normalization and Code

Why Normalization



$$W_{new} = W_{old} - \eta \frac{\partial L}{\partial W}$$

f_1	f_2
1	100
2	220
3	350

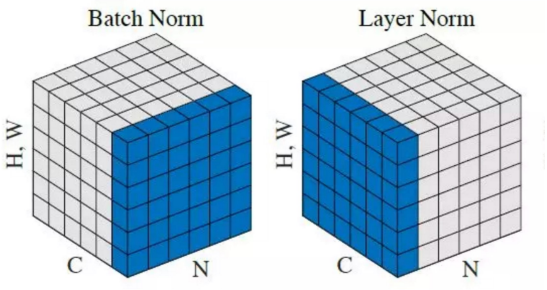
Stable Learning : Challenges of Vanishing and exploding gradient.

Faster Training : Fast convergence to minima.

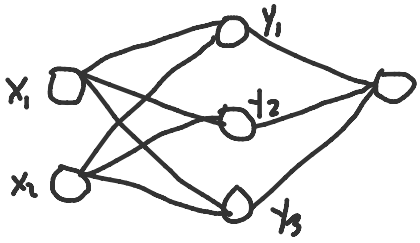
Internal Covariate issue.

$\mu = 0 \quad \sigma = 1$

Types of Normalization



Batch Normalization



X1	X2	Z1	Z2	Z3
2	6	2	3	2
3	7	5	4	5
1	5	8	5	6
5	8	9	6	8
4	4	2	7	9

$\mu_1 \quad \sigma_1 \quad \mu_2 \quad \sigma_2 \quad \mu_3 \quad \sigma_3$

$$\hat{z}_{11} = \frac{z_{11} - \mu_1}{\sigma_1}$$

$$\hat{z}_{12} = \frac{z_{12} - \mu_1}{\sigma_1}$$

$$z'_{11} = \gamma \hat{z}_{11} + \beta$$

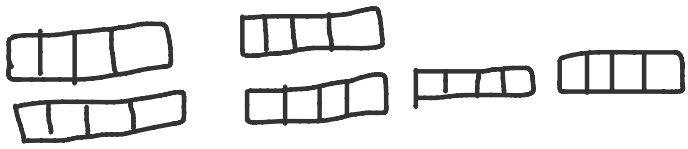
$\gamma = 1$

$\beta = 0$

Issue With Batch Normalization

Batch size 2

Hello Team
I Love Machine Learning



Emb Dimension 4

												Padding							
Hi There	2.5	3.2	1.3	4.5	1	3.2	5.4	-0.9	0	0	0	0	0	0	0	0			
I Love Machine Learning	1.1	-2.3	3.4	4.2	3.2	0.9	0.8	9.1	2.1	11	2.3	3.1	1.2	1.1	0.9	0.7			

Matrix Form

Hi	2.5	3.2	1.3	4.5
There	1	3.2	5.4	-0.9
Padding	0	0	0	0
Padding	0	0	0	0

I	1.1	-2.3	3.4	4.2
Love	3.2	0.9	0.8	9.1
Machine	2.1	11	2.3	3.1
Learning	1.2	1.1	0.9	0.7

Lets Code

Batch 2