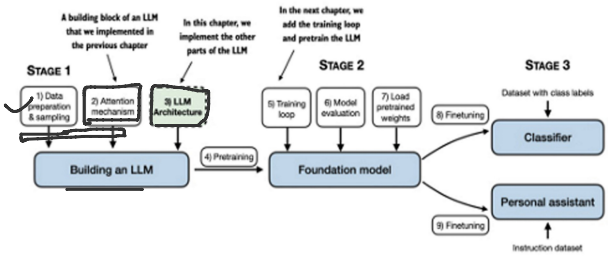
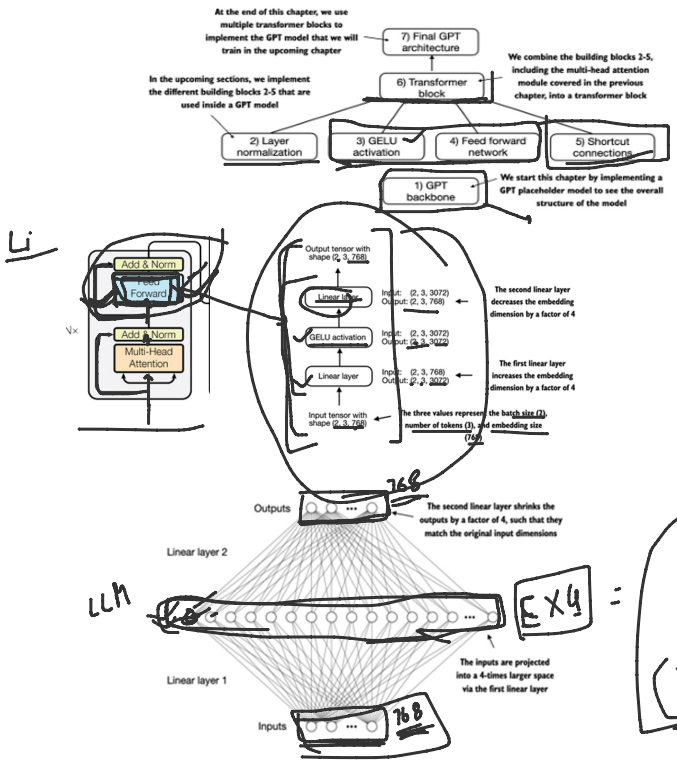
LLM Architecture (GPT Model)-Feed Forward Network

25 फरवरी 2025 22:22





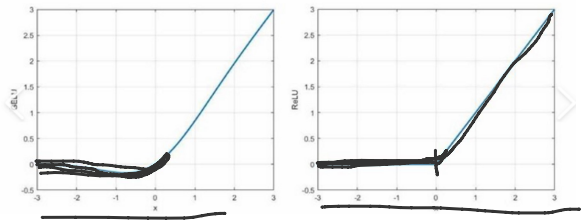
GELU Activation Function

$$GELU = x \cdot \Phi(x)$$

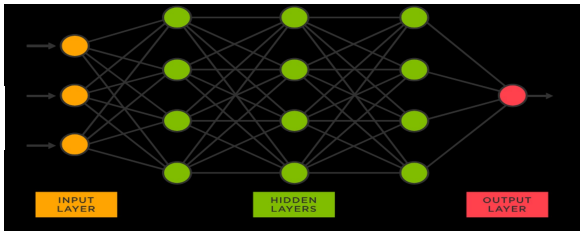
CDF of Standard Gaussian distribution

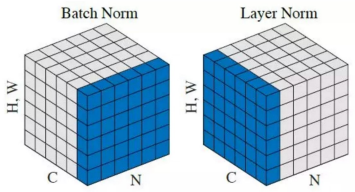
GELU Activation approximation used in GPT2

$$GELU(x) \approx 0.5 \cdot x \cdot (1 + \tanh[\sqrt{(2/\pi)} \cdot (x + 0.044715 \cdot x^3)])$$



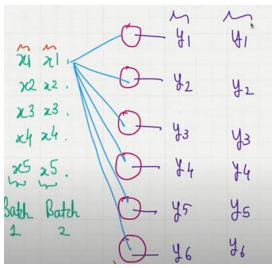
$$W_{new} = W_{old} - \eta \frac{\partial L}{\partial W}$$





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

									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

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