Preliminary Results

CS4681- Advanced Machine Learning

Name - Kumarasekara G.K.

Index Number - 210314E

Project Code - NLP009

So far, I have experimented on the impact of the choice of activation function on the performance of PEGASUS-X using three activation functions and four benchmark datasets. The three activations functions I have experimented with are

- Rectified Linear Unit (ReLU)
- Gaussian Error Linear Unit (GELU)
- Sigmoid Linear Unit (SiLU)

The four benchmark datasets I have used are

- CNN/DailyMail dataset
- XSum dataset
- SummScreen dataset
- GovReport dataset

The following results were obtained from experiments done using those activation functions and datasets.

Dataset	ReLU	GELU	SiLU
SummScreen	27.2	35.0	18.8
GovReport	41.5	59.3	31.8
CNN/DailyMail	34.8	43.4	25.1
XSum	37.4	45.8	22.9

Table 1: ROUGE-1 scores across different activation functions

Dataset	ReLU	GELU	SiLU
SummScreen	4.8	8.9	2.3
GovReport	20.1	29.3	13.2
CNN/DailyMail	18.8	21.2	10.2
XSum	17.2	22.8	10.4

Table 2: ROUGE-2 scores across different activation functions

Overall, the GELU activation function consistently achieved higher ROUGE-1 and ROUGE-2 scores across all datasets, suggesting that its smooth, probabilistic gating function better captures non-linear relationships in summarisation tasks compared to ReLU and SiLU.