Deep Learning - DEI

Group 30: Alice Mota ist1102500 | Francisco Leitão ist103898

### Contribution:

Question 1 was solved by both members, with Francisco having done exercises 1.1 and 1.3, and Alice doing 1.2. Question 2 was done by Francisco, while Question 3 was done by Alice.

# Homework 2:

## Question 3

1. In the training set (Plot 1) the performance, with error rate as the chosen metric was of 0.3154. In the test set, the vanilla character-level encoder-decoder displayed a Character Error Rate (CER) of 0.3087 and a Word Error Rate (WER) of 0.8100. The small gap between validation and test CER indicates strong generalization. However, the high test WER highlights the model’s difficulty in achieving exact word-level matches despite accurate character-level predictions, suggesting the need for more sophisticated mechanisms.



Plot 1 – Vanilla character-level encoder-decoder error rate in function of the epoch number

1. In the training set (Plot 2) the performance, with error rate as the chosen metric was of 0.2166. In the test set, the vanilla character-level encoder-decoder displayed a Character Error Rate (CER) of 0.2132 and a Word Error Rate (WER) of 0.7190. While CER is close to the best validation rate, WER can likely be improved through further optimization or fine-tuning.



Plot 2 – Vanilla character-level encoder-decoder error rate in function of the epoch number

1. The results obtained after running the implemented function and evaluating the model on the test set are as follows:|
   1. Character Error Rate (CER): 0.2265
   2. Word Error Rate (WER): 0.7400
   3. WER@3: 0.6510, which means that approximately 65% of examples lack a correct prediction among the 3 samples, highlighting potential areas for model improvement.

Nucleous sampling enhanced text generation by balancing quality and diversity. It introduced variability in predictions as well, as illustrated by the obtained examples:

| **Ground Truth** | **Predictions** |
| --- | --- |
| remainder | remander, romander, remandere |
| Jingchuan | Jingquine, Zinghthwan, Jingtuan |
| scare | scar, skare |
| glossal | glossel, glossal |
| touchedness | tutchidness, tutchedness, tutchidenis |
| alkaloses | alcolousis, alcolocise, alcolosise |
| companions | companniens, compannians, Kempanians |
| Fingal | fingaall, fingaal |
| grievances | greevences, greavences, greevencis |
| Bronx | Branks, Branx, brongs |

Although predictions vary, some examples demonstrate challenges in producing accurate outputs, particularly for less common words (e.g., "Jingchuan").

The results indicate a trade-off between generating varied predictions and maintaining accuracy. Further refinement of the sampling strategy could include adjusting the pp or combining beam search.