Language Modelling - Lab 4

Marco Prosperi (257857)

University of Trento

marco.prosperi@studenti.unitn.it

1. Introduction (approx. 100 words)

The project aimed to improve the baseline RNN performance by incrementally adding features and dynamically adjusting hyperparameters (learning rate, hidden units, embedding size). The model was trained on the Penn Treebank dataset, using perplexity as the main evaluation metric. In the second part, advanced techniques were incorporated to further enhance the language model, including Weight Tying to reduce parameters, Variational Dropout (excluding DropConnect) for regularization, and Averaged Stochastic Gradient Descent (AvSGD) for better optimization. These improvements aimed to enhance generalization and convergence while maintaining low perplexity (less than 250).

2. Implementation details (max approx. 200-300 words)

Do not explain the backbone deep neural network (e.g. RNN or BERT). Instead, focus on what you did on top of it. **Add references if you take inspiration from the code of others**

3. Results

Add tables and explain how you evaluated your model. Tables and images of plots or confusion matrices do not count in the page limit.

Model	PPL	LR	Hidden	Emb
RNN	173.22	0.1	100	100
LSTM	137.31	2	300	300
LSTM + Var Dropout	123.14	2	300	300
LSTM + Var Dropout + AdamW	109.43	0.001	400	400

Table 1: *Perplexity and hyperparameters of the models.*

4. References

[1] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[1]