Machine Translation Sabrina Brändle, Eirini Valkana

**MT Exercise 3  
Recurrent networks**

**training perplexity after each epoch**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Training perplexity | Dropout 0 | Dropout 0.3 | Dropout 0.5 | Dropout 0.7 | Dropout 1 |
| **Epoch 1** |  |  | 275.33 |  |  |
| **Epoch 2** |  |  | 148.67 |  |  |
| **Epoch 3** |  |  | 116.31 |  |  |
| **Epoch 4** |  |  | 100.64 |  |  |
| **Epoch 5** |  |  | 91.06 |  |  |
| **Epoch 6** |  |  | 85.02 |  |  |
| **Epoch 7** |  |  | 79.72 |  |  |
| **Epoch 8** |  |  | 76.53 |  |  |
| **Epoch 9** |  |  | 72.94 |  |  |
| **Epoch 10** |  |  | 70.38 |  |  |
| **Epoch 11** |  |  | 68.72 |  |  |
| **Epoch 12** |  |  | 66.66 |  |  |
| **Epoch 13** |  |  | 64.94 |  |  |
| **Epoch 14** |  |  | 63.70 |  |  |
| **Epoch 15** |  |  | 62.80 |  |  |
| **Epoch 16** |  |  | 61.23 |  |  |
| **Epoch 17** |  |  | 60.52 |  |  |
| **Epoch 18** |  |  | 59.50 |  |  |
| **Epoch 19** |  |  | 58.33 |  |  |
| **Epoch 20** |  |  | 57.64 |  |  |
| **Epoch 21** |  |  | 56.69 |  |  |
| **Epoch 22** |  |  | 56.57 |  |  |
| **Epoch 23** |  |  | 55.57 |  |  |
| **Epoch 24** |  |  | 54.68 |  |  |
| **Epoch 25** |  |  | 54.53 |  |  |
| **Epoch 26** |  |  | 54.09 |  |  |
| **Epoch 27** |  |  | 53.74 |  |  |
| **Epoch 28** |  |  | 53.02 |  |  |
| **Epoch 29** |  |  | 52.50 |  |  |
| **Epoch 30** |  |  | 52.40 |  |  |
| **Epoch 31** |  |  | 51.66 |  |  |
| **Epoch 32** |  |  | 51.00 |  |  |
| **Epoch 33** |  |  | 51.25 |  |  |
| **Epoch 34** |  |  | 51.05 |  |  |
| **Epoch 35** |  |  | 50.52 |  |  |
| **Epoch 36** |  |  | 50.14 |  |  |
| **Epoch 37** |  |  | 49.75 |  |  |
| **Epoch 38** |  |  | 49.44 |  |  |
| **Epoch 39** |  |  | 49.27 |  |  |
| **Epoch 40** |  |  | 49.20 |  |  |

**validation perplexity on the valiation set after each epoch**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Valid. perplexity | Dropout 0 | Dropout 0.3 | Dropout 0.5 | Dropout 0.7 | Dropout 1 |
| **Epoch 1** |  |  |  |  |  |
| **Epoch 2** |  |  |  |  |  |
| **Epoch 3** |  |  |  |  |  |
| **Epoch 4** |  |  |  |  |  |
| **Epoch 5** |  |  |  |  |  |
| **Epoch 6** |  |  |  |  |  |
| **Epoch 7** |  |  |  |  |  |
| **Epoch 8** |  |  |  |  |  |
| **Epoch 9** |  |  |  |  |  |
| **Epoch 10** |  |  |  |  |  |
| **Epoch 11** |  |  |  |  |  |
| **Epoch 12** |  |  |  |  |  |
| **Epoch 13** |  |  |  |  |  |
| **Epoch 14** |  |  |  |  |  |
| **Epoch 15** |  |  |  |  |  |
| **Epoch 16** |  |  |  |  |  |
| **Epoch 17** |  |  |  |  |  |
| **Epoch 18** |  |  |  |  |  |
| **Epoch 19** |  |  |  |  |  |
| **Epoch 20** |  |  |  |  |  |
| **Epoch 21** |  |  |  |  |  |
| **Epoch 22** |  |  |  |  |  |
| **Epoch 23** |  |  |  |  |  |
| **Epoch 24** |  |  |  |  |  |
| **Epoch 25** |  |  |  |  |  |
| **Epoch 26** |  |  |  |  |  |
| **Epoch 27** |  |  |  |  |  |
| **Epoch 28** |  |  |  |  |  |
| **Epoch 29** |  |  |  |  |  |
| **Epoch 30** |  |  |  |  |  |
| **Epoch 31** |  |  |  |  |  |
| **Epoch 32** |  |  |  |  |  |
| **Epoch 33** |  |  |  |  |  |
| **Epoch 34** |  |  |  |  |  |
| **Epoch 35** |  |  |  |  |  |
| **Epoch 36** |  |  |  |  |  |
| **Epoch 37** |  |  |  |  |  |
| **Epoch 38** |  |  |  |  |  |
| **Epoch 39** |  |  |  |  |  |
| **Epoch 40** |  |  |  |  |  |

**test perplexity on the test set as the very last step in training**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Test perplexity | Dropout 0 | Dropout 0.3 | Dropout 0.5 | Dropout 0.7 | Dropout 1 |
| **ppl** |  |  |  |  |  |

**Create a line chart each for the training and the validation perplexity to visualize the results.**

**Can you see a connection between the training, validation and test perplexity? Based on your results, which dropout setting do you think is the best and why?**

**Sample some text from your best model (the one that obtains the lowest test perplexity), for instance by changing the script scripts/generate.sh. What do you think of its quality? Does it resemble the original training data?**