Machine Translation Sabrina Brändle, Eirini Valkana

**MT Exercise 3  
Recurrent networks**

**Link to repository:** [**https://github.com/sabrinabraendle/mt-exercise-3**](https://github.com/sabrinabraendle/mt-exercise-3)

**Training perplexity after each epoch**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Train. perplexity** | **dropout 0** | **dropout 0.3** | **dropout 0.5** | **dropout 0.7** | **dropout 1** |
| **Epoch 1** | 257.39 | 260.83 | 275.33 | 308.2 | 464.75 |
| **Epoch 2** | 114.17 | 129.76 | 148.67 | 189.31 | 428.07 |
| **Epoch 3** | 83.85 | 99.61 | 116.31 | 158.08 | 418.49 |
| **Epoch 4** | 67.89 | 84.45 | 100.64 | 140.97 | 414.37 |
| **Epoch 5** | 57.46 | 74.67 | 91.06 | 131.23 | 412.15 |
| **Epoch 6** | 49.25 | 67.75 | 85.02 | 123.69 | 410.8 |
| **Epoch 7** | 41.63 | 62.18 | 79.72 | 118.65 | 409.91 |
| **Epoch 8** | 35.29 | 57.64 | 76.53 | 114.47 | 409.29 |
| **Epoch 9** | 30.26 | 54.07 | 72.94 | 111.49 | 408.85 |
| **Epoch 10** | 26.32 | 51.49 | 70.38 | 108.48 | 405.55 |
| **Epoch 11** | 23.02 | 48.78 | 68.72 | 106.8 | 405.55 |
| **Epoch 12** | 20.51 | 46.59 | 66.66 | 104.77 | 405.55 |
| **Epoch 13** | 18.51 | 44.59 | 64.94 | 103.48 | 405.54 |
| **Epoch 14** | 16.94 | 43.04 | 63.7 | 100.96 | 405.51 |
| **Epoch 15** | 15.63 | 41.52 | 62.8 | 100.62 | 405.48 |
| **Epoch 16** | 14.64 | 40.18 | 61.23 | 99.33 | 404.71 |
| **Epoch 17** | 13.77 | 39.06 | 60.52 | 98.24 | 404.73 |
| **Epoch 18** | 13.01 | 37.81 | 59.5 | 97.71 | 404.45 |
| **Epoch 19** | 12.37 | 37.11 | 58.33 | 96.61 | 404.47 |
| **Epoch 20** | 11.81 | 36.14 | 57.64 | 96.17 | 404.49 |
| **Epoch 21** | 11.36 | 35.45 | 56.69 | 95.8 | 404.5 |
| **Epoch 22** | 10.98 | 34.81 | 56.57 | 95.14 | 404.51 |
| **Epoch 23** | 10.69 | 34.04 | 55.57 | 94.69 | 404.41 |
| **Epoch 24** | 10.38 | 33.5 | 54.68 | 93.99 | 404.41 |
| **Epoch 25** | 10.18 | 32.89 | 54.53 | 93.48 | 404.42 |
| **Epoch 26** | 9.81 | 32.36 | 54.09 | 92.78 | 404.42 |
| **Epoch 27** | 9.61 | 31.82 | 53.74 | 91.81 | 404.43 |
| **Epoch 28** | 9.52 | 31.4 | 53.02 | 91.19 | 404.43 |
| **Epoch 29** | 9.21 | 31.11 | 52.5 | 91.27 | 404.43 |
| **Epoch 30** | 9.06 | 30.76 | 52.4 | 90.89 | 404.44 |
| **Epoch 31** | 8.92 | 30.53 | 51.66 | 89.88 | 404.44 |
| **Epoch 32** | 8.63 | 29.99 | 51 | 90.8 | 404.44 |
| **Epoch 33** | 8.61 | 29.85 | 51.25 | 89.76 | 404.44 |
| **Epoch 34** | 8.51 | 24.92 | 51.05 | 89.26 | 404.44 |
| **Epoch 35** | 8.33 | 24.12 | 50.52 | 89.2 | 404.45 |
| **Epoch 36** | 8.07 | 23.63 | 50.14 | 82.03 | 404.45 |
| **Epoch 37** | 8.02 | 23.04 | 49.75 | 80.32 | 404.45 |
| **Epoch 38** | 5.43 | 22.8 | 49.44 | 78.78 | 404.45 |
| **Epoch 39** | 4.31 | 22.58 | 49.27 | 78.06 | 404.42 |
| **Epoch 40** | 3.73 | 22.36 | 49.2 | 77.19 | 404.42 |

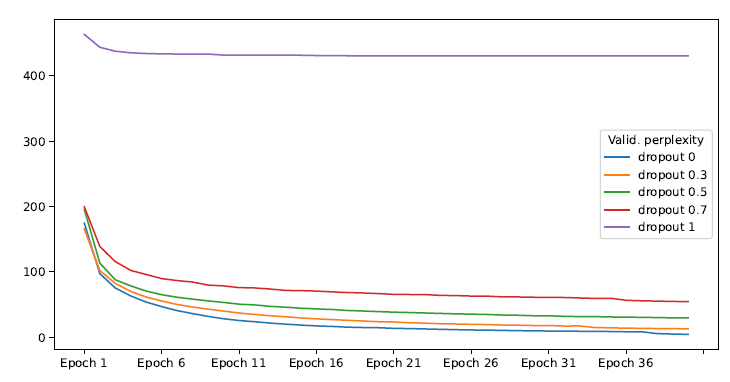
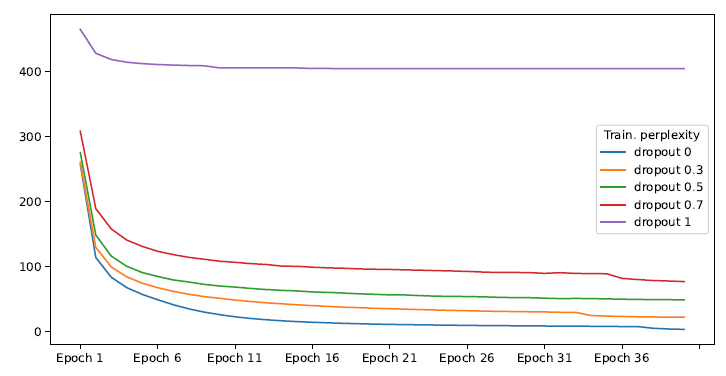
**Validation perplexity on the validation set after each epoch**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Valid. perplexity** | **dropout 0** | **dropout 0.3** | **dropout 0.5** | **dropout 0.7** | **dropout 1** |
| **Epoch 1** | 174.27 | 165.24 | 195.47 | 199.31 | 463.2 |
| **Epoch 2** | 97.09 | 101.9 | 113.23 | 138.69 | 443.49 |
| **Epoch 3** | 75.08 | 82.12 | 87.6 | 115.48 | 437.53 |
| **Epoch 4** | 62.87 | 69.79 | 78.29 | 101.9 | 435.07 |
| **Epoch 5** | 53.47 | 61.21 | 70.35 | 95.68 | 433.94 |
| **Epoch 6** | 46.53 | 55.14 | 64.74 | 89.39 | 433.4 |
| **Epoch 7** | 40.44 | 49.76 | 60.86 | 86.39 | 433.17 |
| **Epoch 8** | 35.8 | 45.93 | 57.92 | 84.12 | 433.1 |
| **Epoch 9** | 31.57 | 42.63 | 55.27 | 79.51 | 433.13 |
| **Epoch 10** | 27.95 | 39.69 | 52.91 | 78.25 | 431.48 |
| **Epoch 11** | 25.27 | 36.72 | 50.26 | 75.59 | 431.4 |
| **Epoch 12** | 23.4 | 34.54 | 49.11 | 75.11 | 431.36 |
| **Epoch 13** | 21.31 | 32.56 | 46.95 | 73.48 | 431.34 |
| **Epoch 14** | 19.69 | 30.99 | 45.67 | 71.33 | 431.33 |
| **Epoch 15** | 18.26 | 29.15 | 44.1 | 70.97 | 431.33 |
| **Epoch 16** | 16.97 | 27.81 | 43.01 | 70.32 | 430.86 |
| **Epoch 17** | 16.22 | 26.67 | 42.15 | 69.07 | 430.86 |
| **Epoch 18** | 15.09 | 25.56 | 40.63 | 68.07 | 430.52 |
| **Epoch 19** | 14.54 | 24.51 | 39.81 | 67.41 | 430.43 |
| **Epoch 20** | 14.22 | 23.37 | 38.86 | 66.44 | 430.4 |
| **Epoch 21** | 13.29 | 22.98 | 38.07 | 65.32 | 430.4 |
| **Epoch 22** | 12.91 | 21.78 | 37.46 | 65.08 | 430.4 |
| **Epoch 23** | 12.33 | 21.29 | 36.79 | 64.88 | 430.37 |
| **Epoch 24** | 11.8 | 20.27 | 36.09 | 63.78 | 430.34 |
| **Epoch 25** | 11.19 | 19.92 | 35.44 | 63.36 | 430.32 |
| **Epoch 26** | 10.79 | 19.49 | 35.02 | 62.69 | 430.3 |
| **Epoch 27** | 10.56 | 19.02 | 34.48 | 62.65 | 430.29 |
| **Epoch 28** | 10.18 | 18.36 | 33.65 | 61.35 | 430.28 |
| **Epoch 29** | 9.67 | 18.09 | 33.33 | 61.31 | 430.28 |
| **Epoch 30** | 9.46 | 17.6 | 32.54 | 60.9 | 430.27 |
| **Epoch 31** | 9.14 | 17.47 | 32.42 | 60.56 | 430.27 |
| **Epoch 32** | 8.99 | 16.84 | 31.68 | 60.49 | 430.27 |
| **Epoch 33** | 8.82 | 16.9 | 31.22 | 59.75 | 430.27 |
| **Epoch 34** | 8.79 | 14.53 | 31.22 | 59.1 | 430.26 |
| **Epoch 35** | 8.4 | 14 | 30.69 | 59.34 | 430.26 |
| **Epoch 36** | 8.03 | 13.56 | 30.37 | 56.2 | 430.26 |
| **Epoch 37** | 8.04 | 13.3 | 30.2 | 55.41 | 430.26 |
| **Epoch 38** | 5.22 | 12.98 | 29.72 | 55 | 430.26 |
| **Epoch 39** | 4.51 | 12.74 | 29.33 | 54.49 | 430.26 |
| **Epoch 40** | 4.06 | 12.52 | 29.21 | 54.04 | 430.26 |

**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** |
| **Epoch 40** | 4.32 | 13.37 | 30.62 | 55.39 | 415.02 |

**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?**

* Dropout 0 because best test perplexity

**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?**

The vocabulary is matched quite well, and some parts of the generated text make sense syntactically. Nevertheless, the sentences as a whole do not have meaning yet. The vocabulary and style therefore resembles the training data, but the quality is not met with this model.