# Language Modelling

## Forward LM

| Epochs | Batch Size | Embedding Size | Hidden Size | Learning Rate | Perplexity | Loss | Fraction of Dataset |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 10 | 5 | 100 | 100 | 0.0002 | 3866 | 8.26 | 1/100 |
| 10 | 5 | 100 | 100 | 0.0001 | 3482 | 8.14 | 1/100 |
| 10 | 5 | 100 | 100 | 0.00005 | 2049 | 7.62 | 1/25 |
| 10 | 10 | 100 | 100 | 0.00005 | 2082 | 7.62 | 1/25 |
| 10 | 7 | 100 | 100 | 0.00001 | 2191 | 7.69 | 1/25 |

- I expect the Perpexity to decrease with increase in the fraction of dataset. 1/100 --> 1/25 we improvement 1000 perplexity.
- Unfortunately, the tranning time is huge which makes is infeasible in our personal laptops.

## Backward LM

- Only one experiment because intuitively forward and backward LM should have the same complexity.
- So the hyperparamters that work for Forward LM should work for Backward LM.
- So we use the the hyperparamters which we got the lowest perplexity for Forward LM.

| Epochs | Batch Size | Embedding Size | Hidden Size | Learning Rate | Perplexity | Loss | Faction of Dataset |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 10 | 10 | 100 | 100 | 0.00001 | 1997 | 7.25 | 1/25 |

# DownStream

- Unable to achive high accuracy to a badly trained ELMov due to computation restriction.
- Experiments using glove.

| Epochs | Batch Size | Embedding Size | Hidden Size | Learning Rate | Accuracy |
| --- | --- | --- | --- | --- | --- |
| 10 | 100 | 100 | 200, 30, 4 | 0.001 | 0.9126 |
| 10 | 100 | 100 | 200, 30, 4 | 0.0009 | 0.9153 |