

# AMS 691 PP2

Zhiwei Lin

December 2023

## Q1

dimension of  $L$  is  $|V| \times d$   
dimension of  $H$  is  $D \times D$   
dimension of  $I$  is  $d \times D$   
dimension of  $U$  is  $D \times |V|$   
dimension of  $\mathbf{b}_1$  is  $1 \times D$   
dimension of  $\mathbf{b}_2$  is  $1 \times |V|$

## Q2

$$PP^{(t)}(y^{(T)}, \hat{y}^{(t)}) = \frac{1}{\bar{P}(x_{t+1}^{pred} = x_{t+1} | x_t, \dots, x_1)} = \frac{1}{\sum_{j=1}^{|V|} y_j^{(t)} \hat{y}_j^{(t)}} = \frac{1}{\hat{y}_z^{(t)}}$$

and

$$CE(y^{(T)}, \hat{y}^{(t)}) = - \sum_{i=1}^{|V|} y_i^{(T)} \cdot \log(\hat{y}_i^{(t)}) = \log\left(\frac{1}{\hat{y}_z^{(t)}}\right)$$

Hence, Cross entropy is log of perplexity.

### Q3

the best parameters I have are:

```
batch_size = 64
embed_size = 512
hidden_size = 1024
num_steps = 10
max_epochs = 15
early_stopping = 15
dropout = 0.1
lr = 0.01
vocab_size = 0
```

Test perplexity: 225.5712890625. Increase number of epoches, embed size and hidden size will likely help to get better test perplexity.

If I started with "in palo alto", I get in palo alto it fall credit-card party and on new york stock exchange stems from the differences <eos>

if I started with "in the future", I get in the future with his james secretary nicholas brady <eos>

If I started with "in light of", I get in light of futures contracts but some economists men to come down the <unk> <unk> <unk> from investors at an offer has long considered all of a claim of britain attracted filled last year to \$ N or \$ N million or N taxpayers is held this quake about N of \$ N in spending parents reported out the account that is currently held another soft magazine publishing critics helpful <unk> set off the agency steps underwriters for N ky. on drugs income rose on sales were also brothers will continue to capitalize as <unk> fur seeds <eos>

As observed, all of the sentences generate by the model make no sense. The model generate words based on previous words to ensure grammar correct, but if combining words generated by the model, it makes no sense. It seems if I give different starting text, the length of sentence varies. If you look into test data, it's about stock market and is written in old early style, so then model will generate more response to "in light of" than "in the future".