lab5 report: Neural Language Modelling

Question 1: mathematical equations of neural network language model

N-grams language model, the representation of probability of t th word is $P(w_t|w_{t-n+1},\dots w_{t-1})$, using a fixed context of size $n-1$.

For neural language model, at first, get features as vector for each word by embedding word, i.e. d-dimensional vector C_{w_t} for each word.

I put $n-1$ word context's features into a matrix C, which every column is w_{t-i}'s features. So, vector C_k contains the learned features for word k. Let vector x denote the concatenation of these $n-1$ feature vectors:

$$x = (C_{w_{t-n+1},1},\dots,C_{w_{t-n+1},d},C_{w_{t-n+2},1},\dots,C_{w_{t-2},d},C_{w_{t-1},1},\dots,C_{w_{t-1},d})$$

Then, using the standard artificial neural network structure, probability classification was conducted on the output unit using softmax activation function to get the probability prediction of the next word starting from x:

$$P(w_t = k|w_{t-n+1},\dots w_{t-1}) = \frac{e^{ak}}{\sum_{l=1}^{N} e^{al}}$$

where:

$$a_k = b_k + \sum_{i=1}^{h} W_{ki}\, tanh(c_i + \sum_{j=1}^{(n-1)d} V_{ij}\, x_{ij})$$

where the vectors b,c and matrices W,V are also parameters (in addition to matrix C). Let us denote θ for the concatenation of all the parameters. The capacity of the model is controlled by the number of hidden units h and by the number of learned word features d .

The neural network is trained using a gradient-based optimization algorithm to maximize the training set log-likelihood:

$$L(\Theta) = \sum_t l\, ogP(w_t|w_{t-n+1},\dots w_{t-1})$$

The gradient $\frac{dL(\Theta)}{d\Theta}$ can be computed using the error back-propagation algorithm.

Question 2 & 3: output

```
************************************************************
When embedding_dim = 3, epoch_times = 500, learning_rate = 0.001, hidden_size = 50
The prediction of words are : ['mathematician', 'ran', 'to', 'the', 'store', '.']
Correct prediction!

The score of physicist = 10.210327502681826
The score of philosopher = 2.5922873242517577
The higher score between physicist and philosopher is : physicist

The cosine similarity between mathematician and physicist = 0.3147851
The cosine similarity between mathematician and philosopher = -0.46233085
More similar two words are : mathematician and physicist
************************************************************
When embedding_dim = 5, epoch_times = 500, learning_rate = 0.01, hidden_size = 50
The prediction of words are : ['mathematician', 'ran', 'to', 'the', 'store', '.']
Correct prediction!

The score of physicist = 71116.65754944767
The score of philosopher = 15358.51601639328
The higher score between physicist and philosopher is : physicist

The cosine similarity between mathematician and physicist = -0.25097254
The cosine similarity between mathematician and philosopher = -0.2841034
More similar two words are : mathematician and physicist
************************************************************
When embedding_dim = 5, epoch_times = 500, learning_rate = 0.01, hidden_size = 100
The prediction of words are : ['mathematician', 'ran', 'to', 'the', 'store', '.']
Correct prediction!

The score of physicist = 176659.69900444476
The score of philosopher = 91118.68331627772
The higher score between physicist and philosopher is : physicist

The cosine similarity between mathematician and physicist = 0.42370993
The cosine similarity between mathematician and philosopher = -0.22234318
More similar two words are : mathematician and physicist
************************************************************
When embedding_dim = 8, epoch_times = 800, learning_rate = 0.001, hidden_size = 100
The prediction of words are : ['mathematician', 'ran', 'to', 'the', 'store', '.']
Correct prediction!

The score of physicist = 13706.46338664561
The score of philosopher = 4700.856918095318
The higher score between physicist and philosopher is : physicist

The cosine similarity between mathematician and physicist = 0.19291314
The cosine similarity between mathematician and philosopher = -0.094890215
More similar two words are : mathematician and physicist
************************************************************
When embedding_dim = 10, epoch_times = 800, learning_rate = 0.01, hidden_size = 100
The prediction of words are : ['mathematician', 'ran', 'to', 'the', 'store', '.']
Correct prediction!

The score of physicist = 99022.19100225132
The score of philosopher = 2151.598799761445
The higher score between physicist and philosopher is : physicist

The cosine similarity between mathematician and physicist = 0.596952
The cosine similarity between mathematician and philosopher = -0.18194295
More similar two words are : mathematician and physicist
```