[COM4513-6513] Lab 5: Neural Language Modelling

Registration Number : **180128033**

1 Introduction

In this lab assignment, we will implement a neural language model. A network involving a large number of processors operating in parallel and arranged in layers is known as Neural network. The Inner most or first layer is loaded with raw input information just like analogous to optic nerves. For next layer which is known as hidden layer, preceding layer output works as input.Finally, last layer i.e output layer act as the output for the system. All of these layers have different dimension.

	First	Second	Third
Layer	Inner Layer	Hidden Layer	Output Layer
Dimension	1 x 20	1 x 128	1 x 17

2 Implementation

In the Implementation of neural network, we have used the given word-embedding code.As asked in the question, test sentences given in the code are replaced by the sentences given to use.After replacing the sentences, we have created a list of sentences with START and END, where each sentence is represented as a list of tokens.Further, a dictionary of unique word is generated from the list of sentences. In the code given, class NGramLanguageModeler is having 2 function init and forward.In the function init, EMBEDDING_DIM = 10, CONTEXT_SIZE = 2 and vocab size are three parameters passed.By multiplying EMBEDDING_DIM = 10 and CONTEXT_SIZE = 2, we get the output for the inner layer.Hidden layer or the second layer is having 128 neurons and the outer layer is 17 vectors which provide them required dimension. Similarly there are 17 unique words fetched from 5 given sentences.From function Forward, we are getting logProbs and embedding which is the change done to predict cosine similarity.

2.1 Equation

: In the below given mathematical equation, R is know as the Relu Function.It is used for activating function.

$$Hid_i(\mathbf{x}, \theta_i, b_i) = R_i(x_1 * \theta_{i1} + \cdots + x_j * \theta_{ij} + \cdots + x_{20} * \theta_{i20} + b_i)(1)$$

In the above equation, i tends from (1 to 20) specifying 20 neurons in the inner most layer whereas j tends from (1 to 128) specifying 128 neuron in the hidden layer.The final and the outermost layer consist of 17 neurons which aligned to 128 neurons in the previous layer(hidden layer) by a function known as Softmax Activation function given below.

$$Out_k(\mathbf{x}, \Theta, \omega_k, \mathbf{b}, c_k) = S_k(\omega_{k1} * Hid_1(\mathbf{x}, \theta_1, b_1) + \omega_{ki} * Hid_i(\mathbf{x}, \theta_i, b_i) + \cdots + \omega_{k}17 * Hid_1(\mathbf{x}, \theta_1 7, b_1 7 + c_k)(2$$

In the above, equation j ranges from 1 to 128(neurons in the hidden layer) whereas k ranges from 1 to 17(neurons in the outermost layer)

3 Sanity Check

In the sanity check, we have chosen sentence "The mathematician ran to the store." from the prepossessed 5 sentences. On changing the hyper-parameter learning rate to 0.01 and taking 15 epochs, model was able to predict "mathematician" every time for 5 consecutive times.
Model was not able to predict “physicist" because in the training set "The mathematician" has occurred more number of times as the context in comparison to the "The Physicist".

4 Test

Yes, model is predicting that the embedding for “physicist” and “mathematician” are closer together than the embedding for “philosopher” and “mathematician”. Although value of embedding changes every time, we run the code.

Embeddings	Cosine Similarity
Physicist and Mathematician	0.0778
Philosopher and Mathematician	-0.0696