

67658 Natural Language Processing – Fall 2021

Exercise (3)

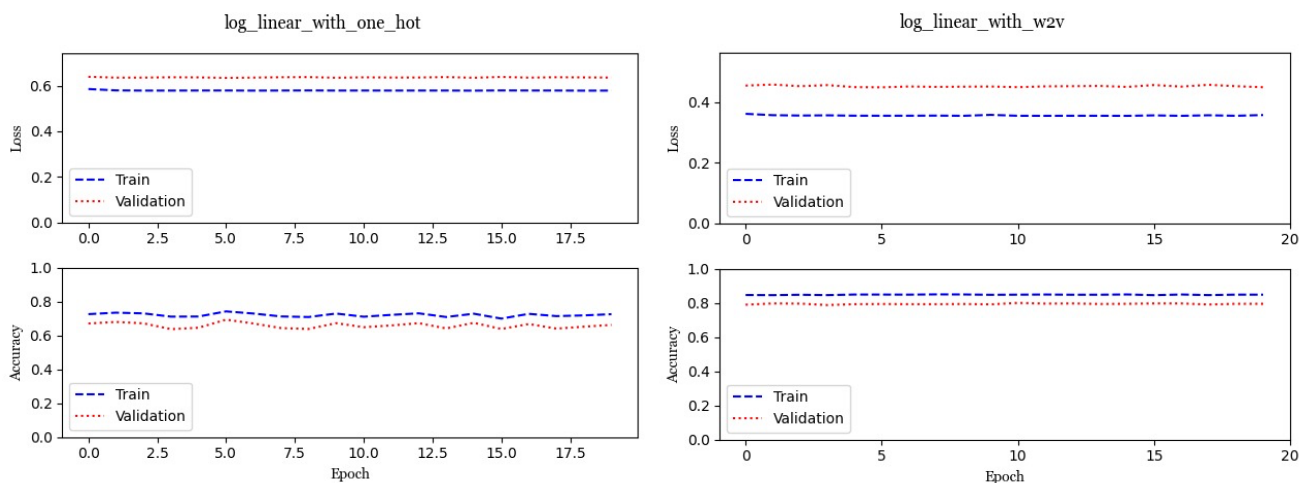
Waseem Abu Leil & Muaz Abdeen

1. Compare Simple log linear model & Word2Vec log linear model:

With the following parameters:

learning rate = 0.01, training epochs = 20, batch size = 64,

and weight decay = 0.001.



Accuracy values

	simple	W2V
Validation set	0.64	0.80
Test set	0.66	0.82

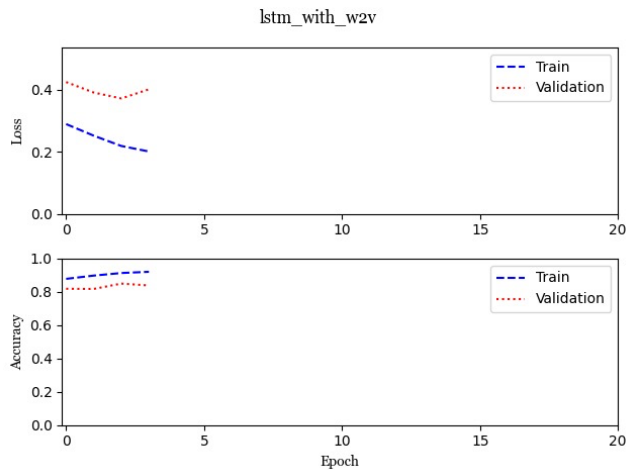
One can notice that the simple model failed to learn and improve the results, its accuracy stays almost the same on the validation set ~64%, while in the whole test set, it is ~66%. For the Word2Vec model its validation accuracy oscillates around ~80%, and the test accuracy is a bit greater ~82%.

We think that the simplicity of the net is the main cause for relative stability of the results, and the model failed to learn and improve. Also, it is clear that Word2Vec model is better than the simple one, because it was able to capture more complicated aspects of the text, in other words, it encodes semantic properties of the sentence which correspond to the task, where average-one-hot encodes only vocabulary so its semantical encoding is lacking.

2. Comparing with LSTM model:

We run LSTM with the following parameters:

learning rate = 0.001, training epochs = 4, batch size = 64, weight decay = 0.001, and dropout probability = 0.5.



Accuracy values

	Validation set	Test set
Simple log linear	0.64	0.66
W2V log linear	0.80	0.82
LSTM with W2V	0.83	0.85

Here the results are much better than the simple log linear model and slightly better than the word2vec model. We think that this is because it can encapsulate past words in the sentence and find relations between different words of the input sentence.

3. Comparing the results of all the models on the two special subsets:

Accuracy values

	Simple log linear	W2V log linear	LSTM with W2V
Negated polarity	0.56	0.56	0.69
Rare words	0.46	0.80	0.74

We notice that simple log linear and w2v log linear have almost the same accuracy on the negated polarity, and the LSTM overcomes both. We think that this is because the construction of the negation, which makes it captured better by a sequential method.

Regarding the rare words, W2V log linear and LSTM have almost the same accuracy, and both overcome the simple model.