

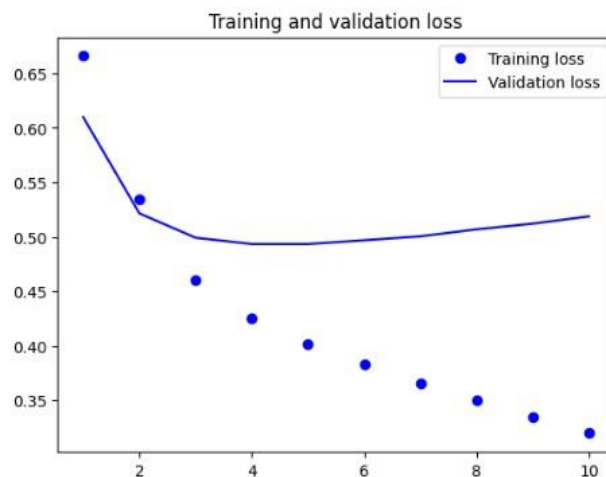
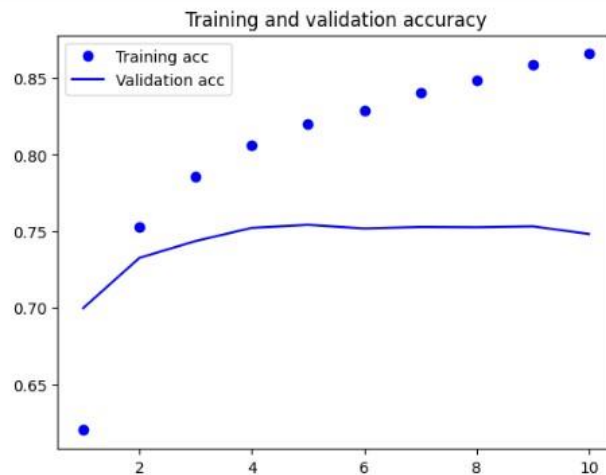
## REPORT ON ASSIGNMENT 3

### WORD EMBEDDING USING EMBEDDING LAYERS AND PRE-TRAINED MODEL

The IMDB dataset is used in our base model with embedding layers from keras with maximum features of 10000 and cut off words of reviews as 20. Here training and validation samples are set to 25000 each.

Cut-off Words	Training Sample	Validation Sample	Word Count	Validation Accuracy
20	25000	25000	10000	75.48

```
Epoch 1/10  
625/625 [=====] - 2s 2ms/step - loss: 0.6659 - acc: 0.6201 - val_loss: 0.6098 - val_acc: 0.6998  
Epoch 2/10  
625/625 [=====] - 1s 2ms/step - loss: 0.5343 - acc: 0.7531 - val_loss: 0.5216 - val_acc: 0.7326  
Epoch 3/10  
625/625 [=====] - 1s 2ms/step - loss: 0.4608 - acc: 0.7857 - val_loss: 0.4992 - val_acc: 0.7436  
Epoch 4/10  
625/625 [=====] - 1s 2ms/step - loss: 0.4251 - acc: 0.8059 - val_loss: 0.4935 - val_acc: 0.7522  
Epoch 5/10  
625/625 [=====] - 1s 2ms/step - loss: 0.4019 - acc: 0.8203 - val_loss: 0.4935 - val_acc: 0.7542  
Epoch 6/10  
625/625 [=====] - 1s 2ms/step - loss: 0.3827 - acc: 0.8289 - val_loss: 0.4969 - val_acc: 0.7518  
Epoch 7/10  
625/625 [=====] - 1s 2ms/step - loss: 0.3655 - acc: 0.8403 - val_loss: 0.5006 - val_acc: 0.7528  
Epoch 8/10  
625/625 [=====] - 1s 2ms/step - loss: 0.3497 - acc: 0.8488 - val_loss: 0.5070 - val_acc: 0.7526  
Epoch 9/10  
625/625 [=====] - 1s 2ms/step - loss: 0.3347 - acc: 0.8586 - val_loss: 0.5122 - val_acc: 0.7532  
Epoch 10/10  
625/625 [=====] - 1s 1ms/step - loss: 0.3198 - acc: 0.8658 - val_loss: 0.5188 - val_acc: 0.7482
```



Maximum Validation Accuracy is obtained when number of epochs=5.

Optimal number of epochs =5.

### Considering the embedding layer in model from scratch:

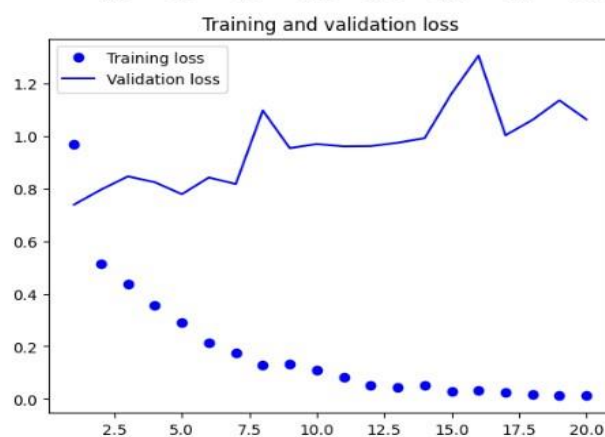
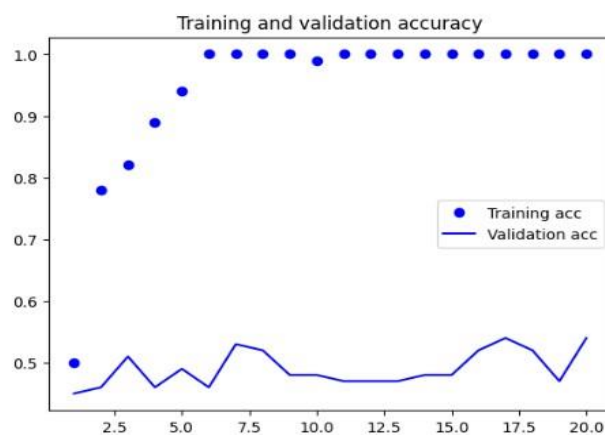
Number of epochs is set to 5 (optimal) and Re-running the example modifying the following:

1. Cut-off reviews after 150 words.
2. Restrict training samples to 100.
3. Validate on 10,000 samples.
4. Consider only the top 10,000 words.

Cut-off Words	Training Sample	Validation Sample	Word Count	Validation Accuracy
20	100	100	10000	52.4
150	100	10000	10000	54.3
150	1000	1000	10000	62
150	10000	10000	10000	86.53
150	15000	10000	10000	87.71
150	20000	20000	10000	87.73

### Considering pre-trained model with glove embedding and freezing the embedding layer

Cut-off Words	Training Sample	Validation Sample	Word Count	Validation Accuracy
20	100	100	10000	54



As we can see that the model quickly overfits even with a small training sample of 100.

**Pre-trained networks are useful when we have very few sample training data to train the model.**

Training the same model without loading the pre-trained word embeddings and without freezing the embedding layer.

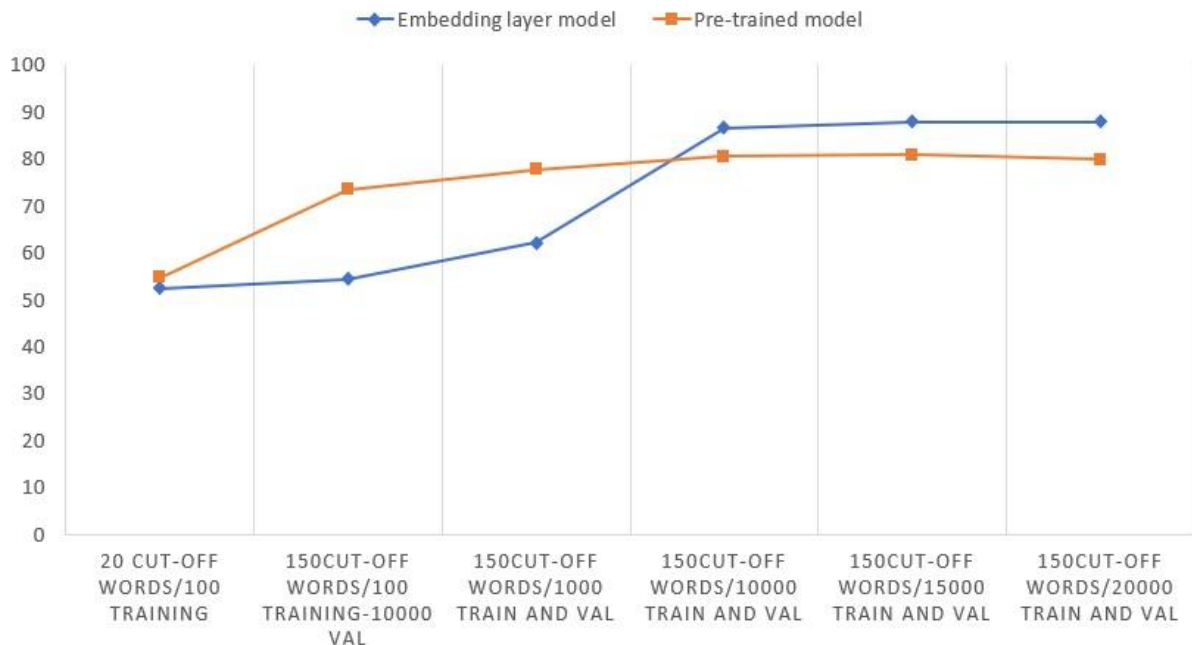
Cut-off Words	Training Sample	Validation Sample	Word Count	Validation Accuracy
20	100	100	10000	52.4

**The validation accuracy gets a little over 50% only which clearly proves that pre-trained network out-performs jointly learned embeddings for a given set of same training samples.**

Cut-off Words	Training Sample	Validation Sample	Word Count	Validation Accuracy
20	100	100	10000	54.62
150	100	10000	10000	73.29
150	1000	1000	10000	77.53
150	10000	10000	10000	80.46
150	15000	10000	10000	80.73
150	20000	20000	10000	79.81

## RESULTS & FINDINGS

### **VARIATION OF ACCURACY OF EMBEDDING LAYER AND PRE-TRAINED MODEL WITH VARYING CUT-OFF WORDS AND TRAINING SAMPLE**



- The pre-trained model was able to give better accuracy of 54.62 % than the model from scratch with embedding layer when we have less training samples of 100. Pre-trained models are loaded with pre-determined weights and that is why these models perform well where we have limited amount of training samples.
- By adjusting hyper-parameters like learning rate to 0.001, embedding vector dimensions and regularization techniques like drop-out of 0.2, the accuracy of the embedding layer model increased to a maximum of 87.73.
- As we can see from the above tables, when we increase the cut-off words from 20 to 150 there is a significant improvement in the accuracy of the model. The model can capture additional semantic relationship between words if the vocabulary is expanded, which could increase the precision of the word embeddings.
- The embedding layer model outperforms the pre-trained network model when we train the model on 10000 training and validation samples.
- Even when the cut-off words were set to 20, the base model was able to give an accuracy of 75.48 % since we used 25000 samples to train and validate the model.
- This clearly gives an outline that increasing the sample size can improve the accuracy of the model in word embedding text sequence models, but the model will eventually over-fit if we further increase the training size after a particular value. (Here in our case – 20000).