Dataset: https://www.kaggle.com/uciml/sms-spam-collection-dataset use

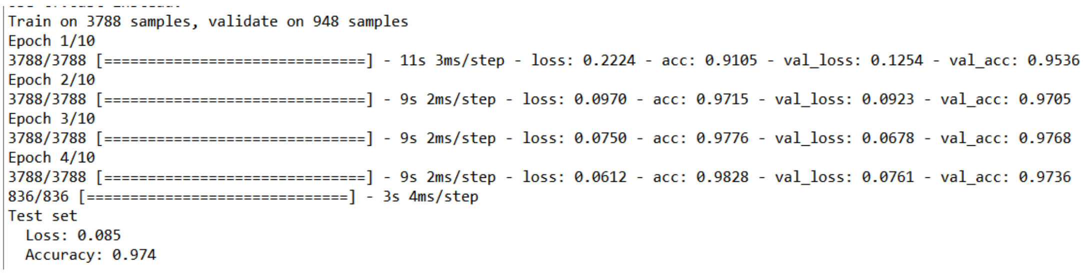
Download glove.6B.200d.txt from UCB for word semantic embedding

Task: Classification on whether the SMS message is spam or not

Approach: word embeddings, with LSTM model

Output:

Use 80% data to train, 20% to test with 97.4% accuracy



```
Train on 3788 samples, validate on 948 samples
Epoch 1/10
3788/3788 [==============================] - 11s 3ms/step - loss: 0.2224 - acc: 0.9105 - val_loss: 0.1254 - val_acc: 0.9536
Epoch 2/10
3788/3788 [==============================] - 9s 2ms/step - loss: 0.0970 - acc: 0.9715 - val_loss: 0.0923 - val_acc: 0.9705
Epoch 3/10
3788/3788 [==============================] - 9s 2ms/step - loss: 0.0750 - acc: 0.9776 - val_loss: 0.0678 - val_acc: 0.9768
Epoch 4/10
3788/3788 [==============================] - 9s 2ms/step - loss: 0.0612 - acc: 0.9828 - val_loss: 0.0761 - val_acc: 0.9736
836/836 [==============================] - 3s 4ms/step
Test set
  Loss: 0.085
  Accuracy: 0.974
```