Model summary is followed:

```
Layer (type)                    Output Shape               Param #
=================================================================
word_embedding_layer (Embedd (None, 23, 300)             27914700
_____
flatten_2 (Flatten)             (None, 6900)               0
_____
hidden_layer (Dense)            (None, 128)                883328
_____
output_layer (Dense)            (None, 2)                  258
=================================================================
Total params: 28,798,286
Trainable params: 883,586
Non-trainable params: 27,914,700
```
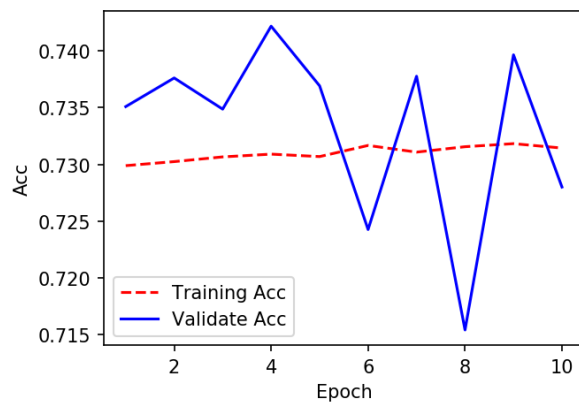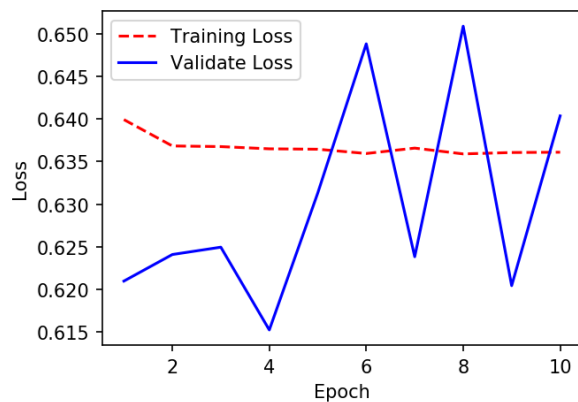
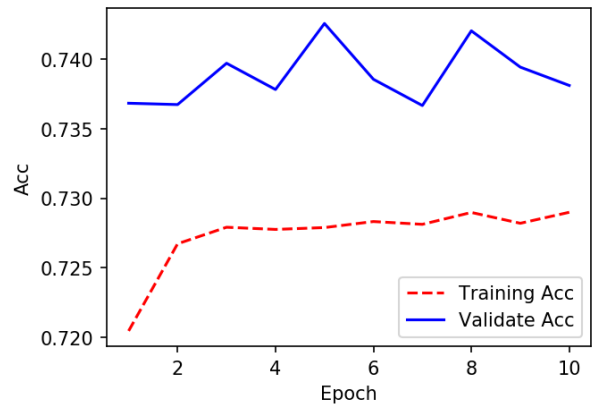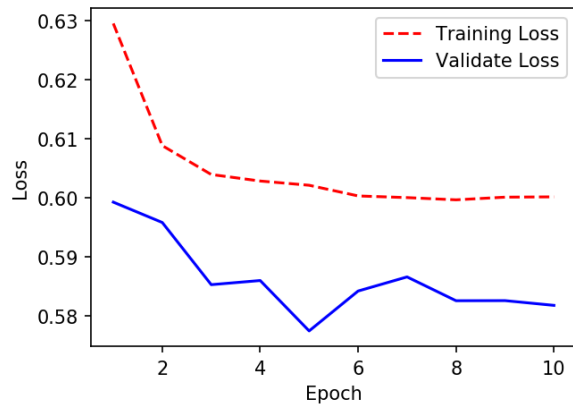**Dropout rate is 0.3 and. L2-norm regularization is applied. Trained models as below.**

relu activition dropout_0.3 and L2-norm regularization
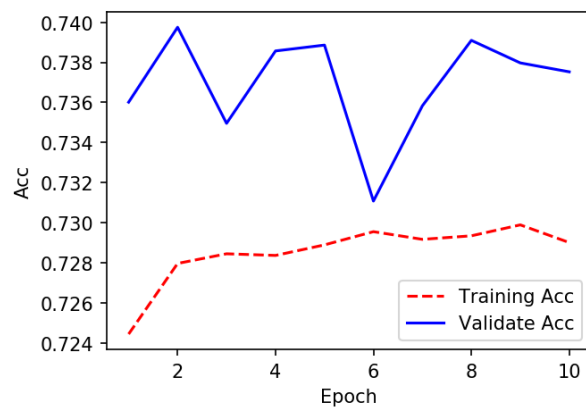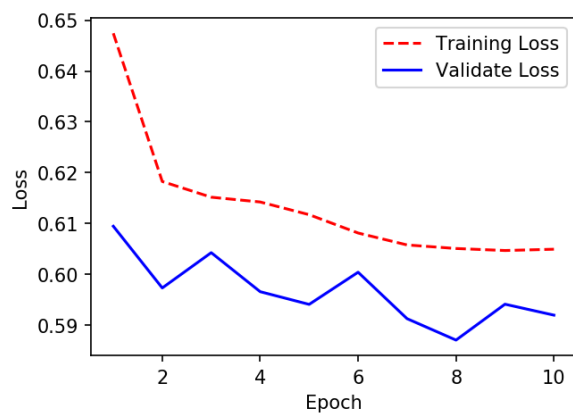
Tanh Activition dropout_0.3 and L2

Sigmoid Activition dropout_0.3 and L2-norm Regularization

| Activation Function	Accuracy (%)
Tanh	76.33
Sigmoid	77.64
Relu	77.65

ReLu is less computationally expensive than tanh and sigmoid because it involves simpler mathematical operations. That is a good point to consider when we are designing deep neural nets. From the results, the activation functions have not influent the accuracy significantly.

**What effect does addition of L2-norm regularization have on the results?**

Without L2-norm regularization, the accuracy of validate dataset is going up, and then drop down, which means that the model is overfitted. As figures followed show that the L2 regularized model has become much more resistant to overfitting than the baseline model, even

though both models have the same number of parameters. `l2(0.001)` means that every coefficient in the weight matrix of the layer will add `0.001 *` `weight_coefficient_value**2` to the total loss of the network.
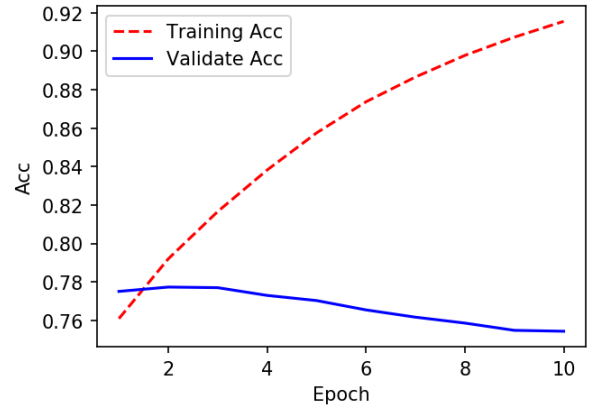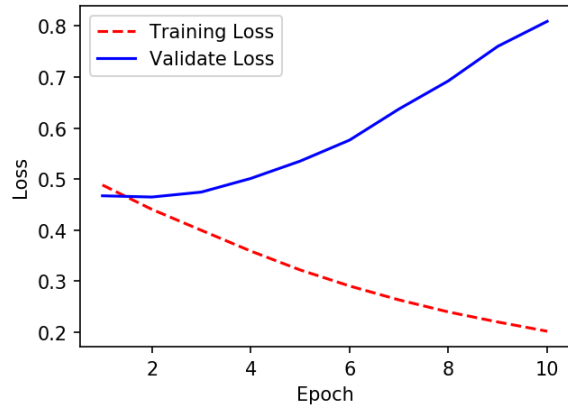
relu activition
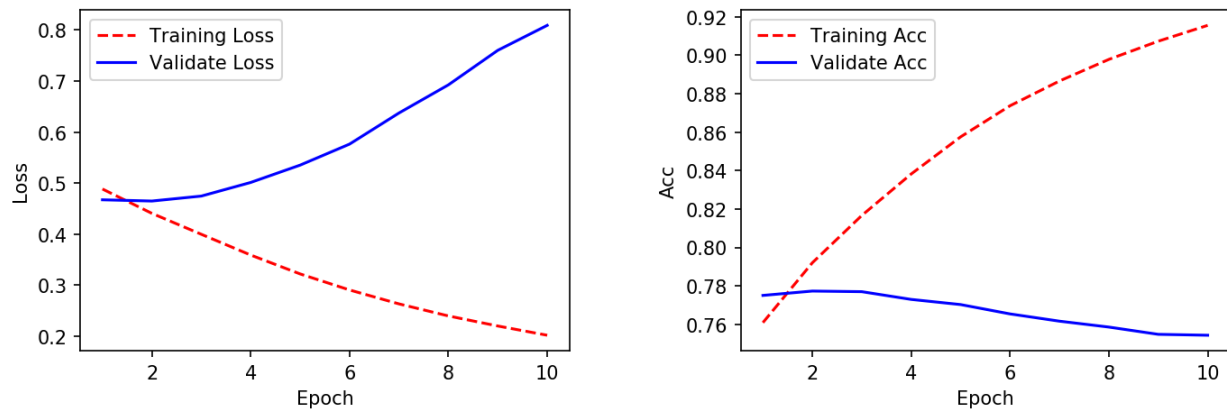

Relu Activition L2 norm Regularization
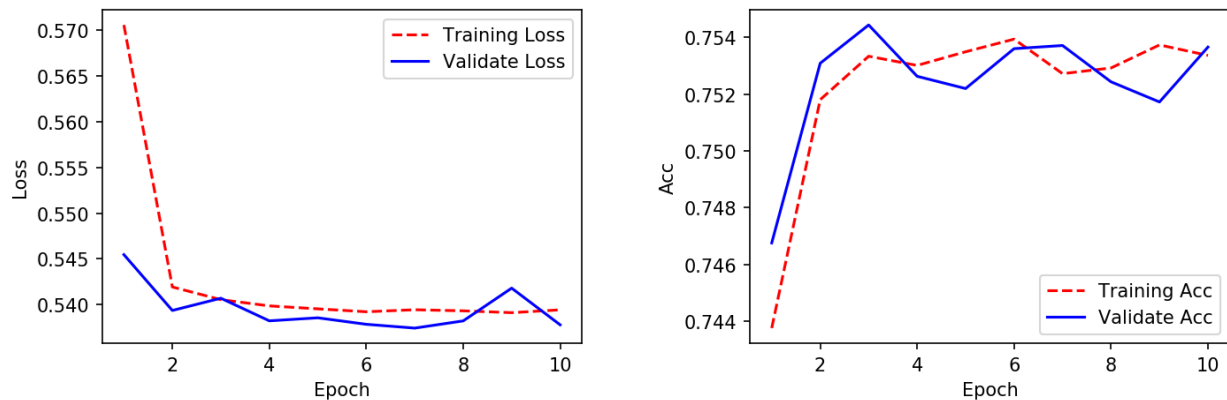
What effect does dropout have on the results?

Dropout is also one of the most effective and most commonly used regularization techniques for neural networks. After applying dropout, the overfitting is prevented by the appropriate dropout rate, which is the fraction of the features that are being zeroed-out; it is usually set between 0.2 and 0.5. From the figures below, when the dropout rate is 0.1, the overfitting cannot be prevented, while the dropout rate of 0.3 and 0.6 work well.
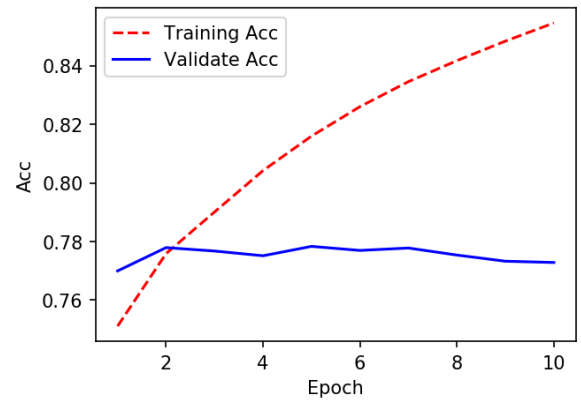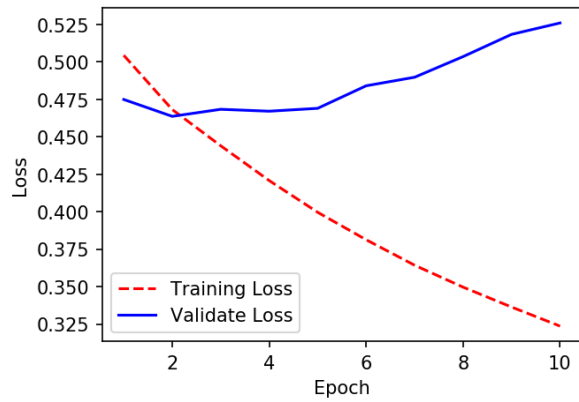
relu activition

Relu Activition dropout_0.1
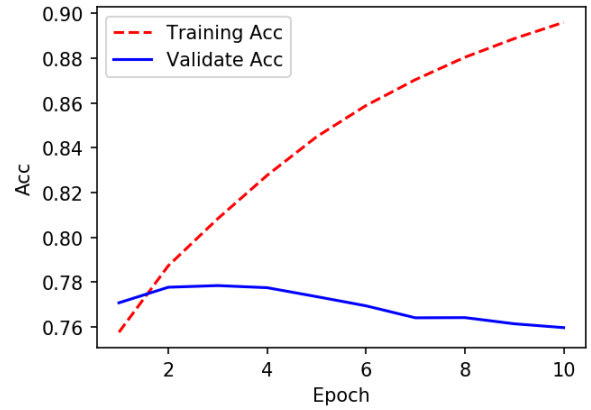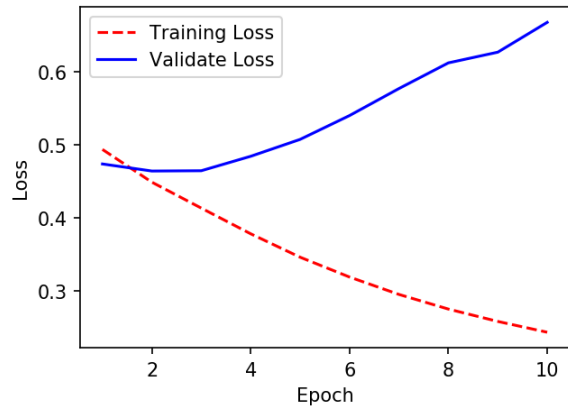
relu activition dropout_0.3

Relu Activition dropout_0.6