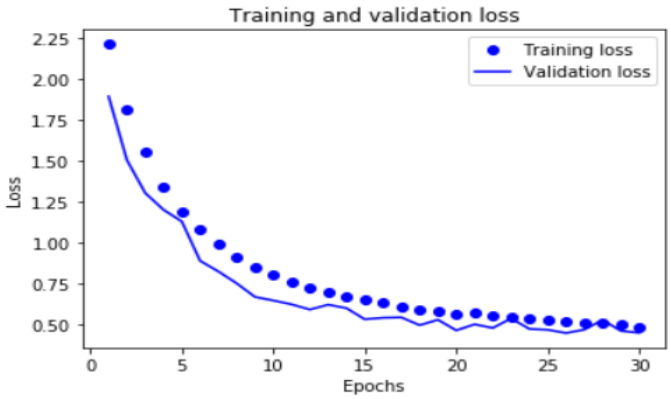
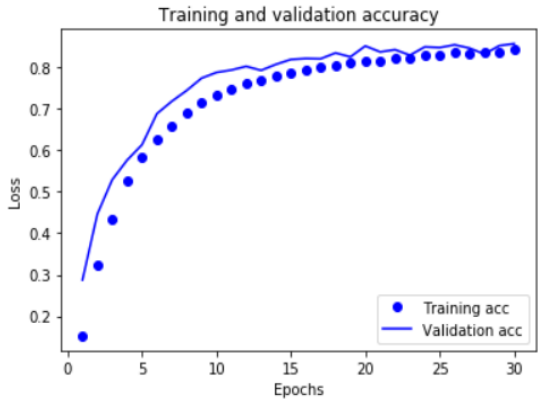
Part 1: Convolutional Neural Network

Dataset – CIFAR-10

Our task for the assignment was to do a comparative study of the different types of Optimizers used in improvising the performance of a ML Algorithm. We have applied a handful of optimizers and validated our choice as below:

1) **ADADELTA**





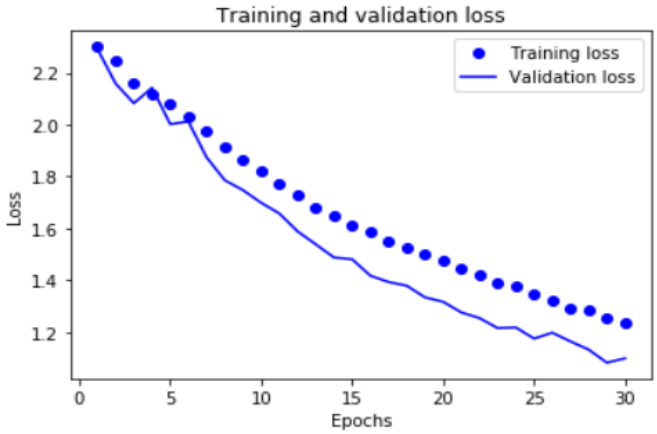

```
accuracy(x_test, y_test, model)
```

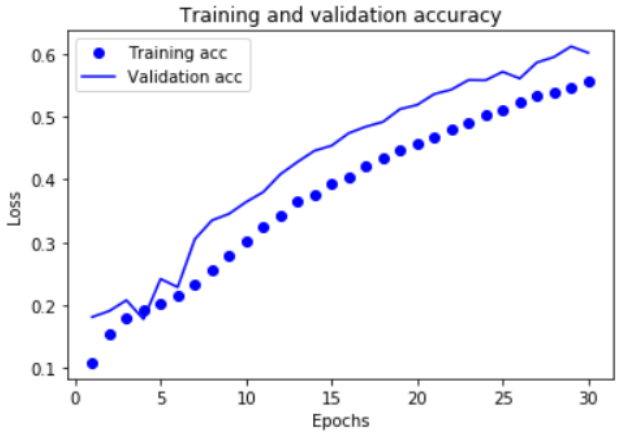
```
85.57000000000001
```

```
accuracy(x_train, y_train, model)
```

```
89.684
```

2) SGD





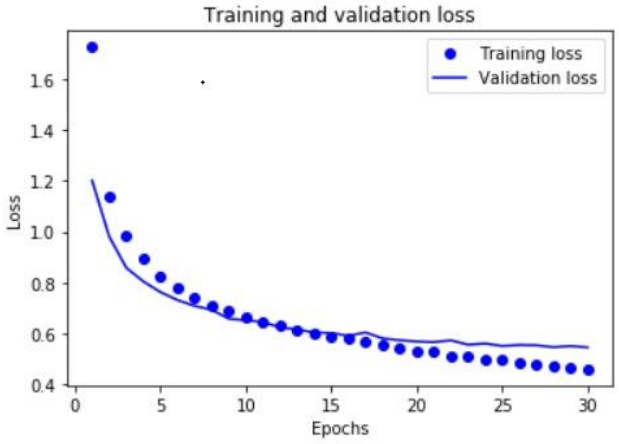

```
accuracy(x_test, y_test, model)
```

```
60.19
```

```
accuracy(x_train, y_train, model)
```

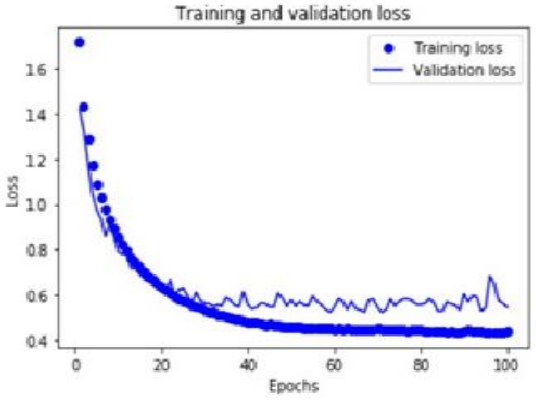
```
60.756
```

3) ADAGRAD



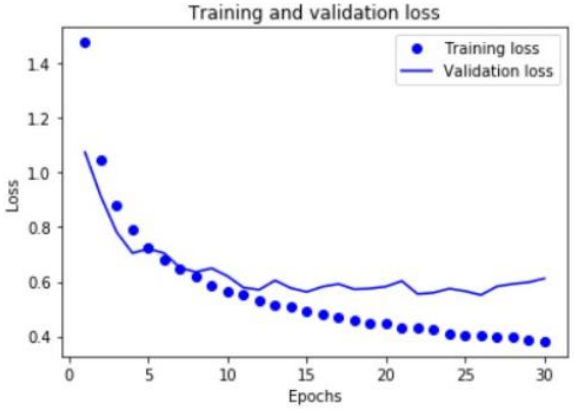

```
10000/10000 [==============================
Test loss: 0.5453300515174866
Test accuracy: 0.8153
```

4) RMSPROP




```
Saved trained model at /home/ubuntu/
10000/10000 [==============================
Test loss: 0.546309899520874
Test accuracy: 0.8221
```

5) ADAM




```
Epoch 30/30
1563/1563 [==============================] - 25
Saved trained model at /home/ubuntu/mynotebooks
10000/10000 [==============================] -
Test loss: 0.6117798939466477
Test accuracy: 0.812
```

From the above comparative study, we see that the performance of the model is better when the chosen optimizer is AdaDelta (Accuracy -) as compared to Stochastic Gradient Descent (Accuracy -) being lowest.

Reasoning: AdaDelta is an extension of AdaGrad. AdaDelta trains deeper and larger nets better, which means the learning rate is faster with sparsely represented data. This is necessary to converge data quickly (needed for images since they are represented sparsely).

In contrast to the above, the performance of the Stochastic Gradient Descent was poor for the network. A comparative theoretical study between SGD and AdaDelta clarifies that AdaDelta is a better performer and more suitable than SGD owing to it per-parameter learning capability.

Conclusion: AdaDelta is the best optimizer for the current dataset.