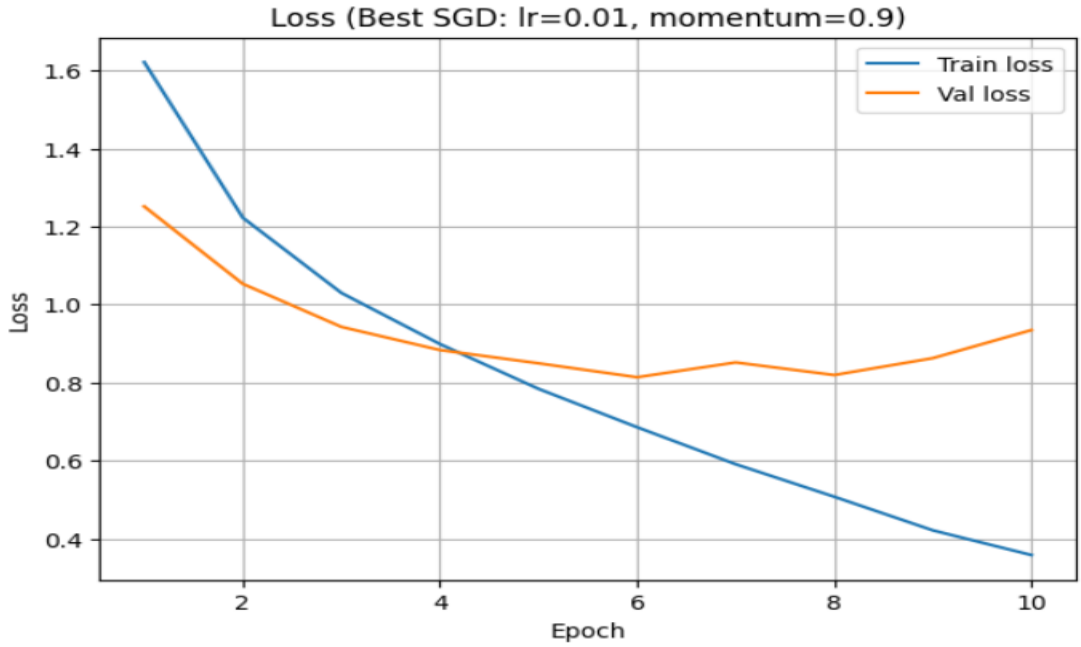
**Qamar Uddin**

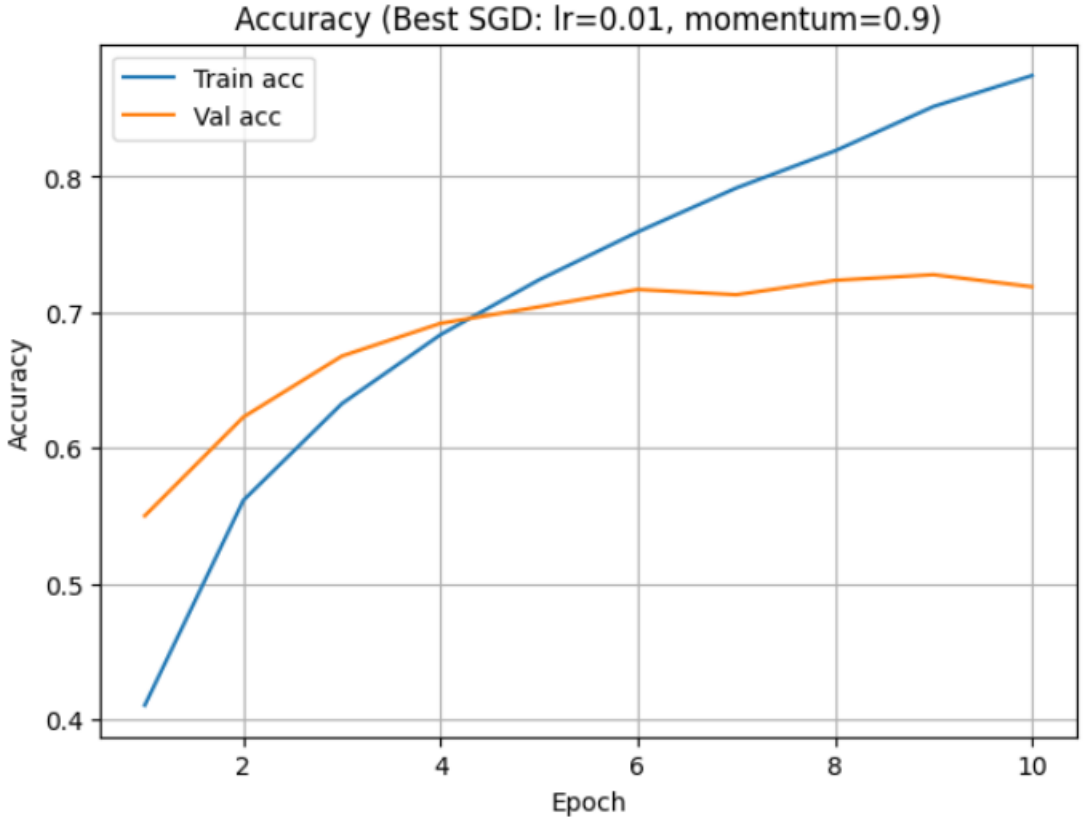
**Exercise 2 Report.**

In this experiment, I tuned the learning rate and momentum of the SGD optimizer using the same CNN architecture from Exercise 1 (fixed ReLU activation). I tested several (learning rate, momentum) combinations and evaluated their effect on convergence and validation accuracy over 10 epochs.

The results show that both hyperparameters strongly influence training stability and speed. With **zero momentum**, the best configuration was **lr = 0.1**, achieving **71.35% validation accuracy**. However, adding momentum significantly improved performance. The best overall configuration was **lr = 0.01** with **momentum = 0.9**, reaching **72.77% validation accuracy**.

Higher momentum produced smoother and faster convergence compared to plain SGD. Too small learning rates led to slow learning, while too large learning rates combined with high momentum caused some instability. Overall, the experiment shows that tuning learning rate and momentum is essential for achieving strong performance with SGD.


Loss (Best SGD: lr=0.01, momentum=0.9)
Train loss
Val loss
Loss
Epoch
1.6
1.4
1.2
1.0
0.8
0.6
0.4
2
4
6
8
10


Accuracy (Best SGD: lr=0.01, momentum=0.9)
Train acc
Val acc
Accuracy
Epoch
0.8
0.7
0.6
0.5
0.4
2
4
6
8
10

**Second part of the question.**

I have also applied RMSProp, AdaGrad and Adam optimizer on the same CNN model with 5 epoch and the result is shown in the pictures below.
1. **RMSProp** performed very badly. The performance was constant. I.e. **10% percent** for each value of lr [0.1, 0.01, 0.001] with 0.9 momentum value; I didn't see any differences there in the results.
2. **Adam** performs well with lr value of **0.001** with no momentum value providing **70.60%.**

3. **AdaGrad** perform a little bit better with lr 0.01 and no momentum value gives **67.04%**

The results summery are shown below with the graph of all optimizers

**Note!** Overall, through the whole experiment SGD with momentum performs the best.

```
Summary of best validation accuracy per optimizer:
SGD     : best lr=0.01, best val acc=70.57%
RMSPROP: best lr=0.1, best val acc=10.00%
ADAGRAD: best lr=0.01, best val acc=67.04%
ADAM    : best lr=0.001, best val acc=70.60%
```