Deep Learning

**Assignment 2 Part 2**

short line

Pramil Panjawani (PhD19008)

Reshan Faraz (PhD19006)

Date : 26-Feb-2021

**Q1) Best optimizer find the ReLU function**

1. **Optimizer**

|  |  |
| --- | --- |
| **Optimizer** | **Train accuracy** |
| Standard GD | 0.7745 |
| GD + Momentum | 0.9027 |
| NAG | 0.913 |
| Adagrad | 0.8351 |
| RMSprop | 0.778 |
| ADAM | 0.7978 |

1. **Initialization Technique**

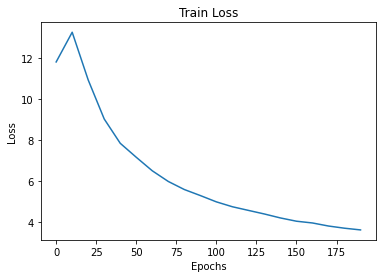
|  |  |
| --- | --- |
|  | **Train accuracy** |
| He | 0.938 |
| Xavier | 0.9374 |

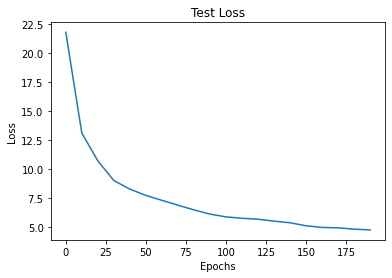
1. **Regularization**

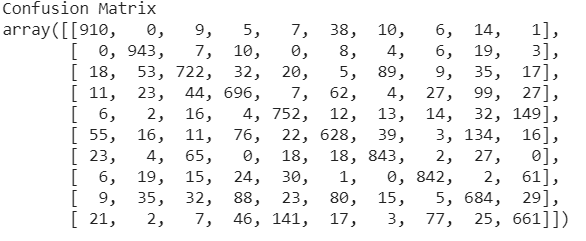
|  |  |
| --- | --- |
|  | **Train accuracy** |
| L1 | 0.9367 |
| L2 | 0.9384 |
| Dropout | 0.9416 |

1. Train and Test loss vs Epochs

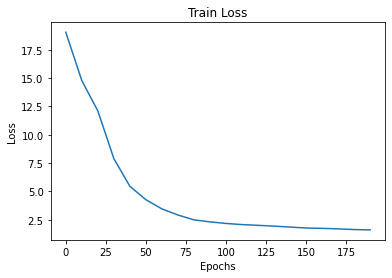
i) Gradient Descent

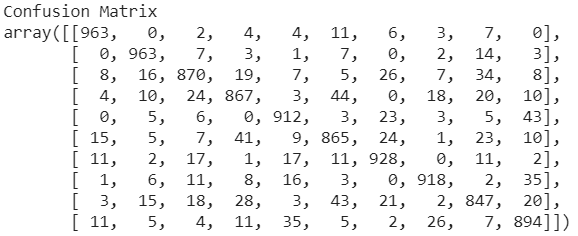
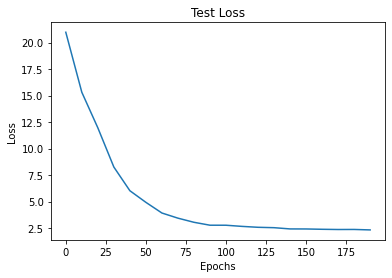




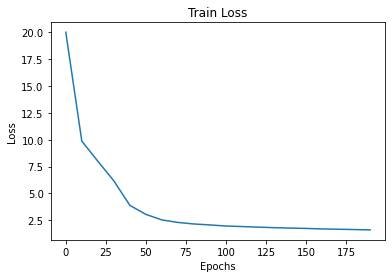
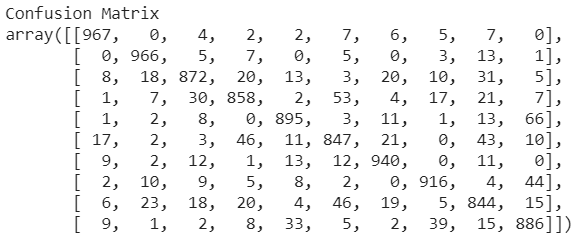
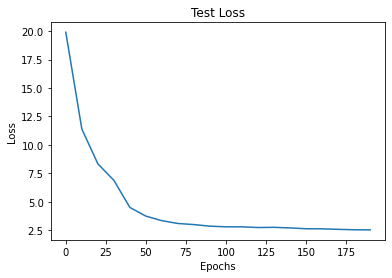


ii) Gradient Descent + Momentum

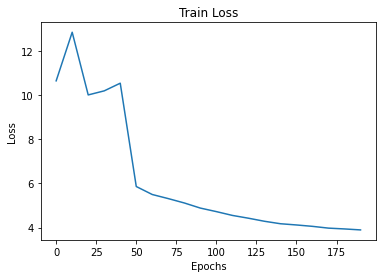
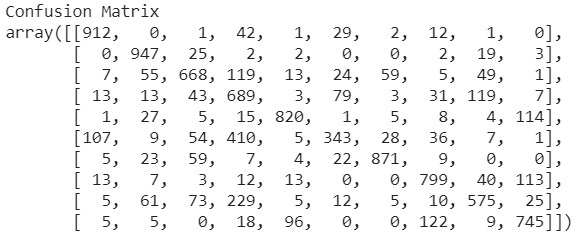
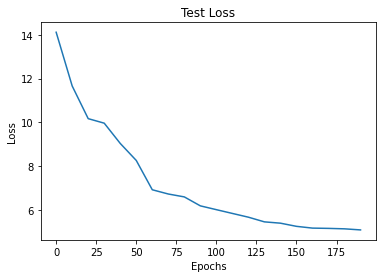




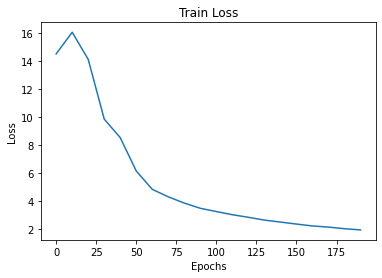
iii) NAG

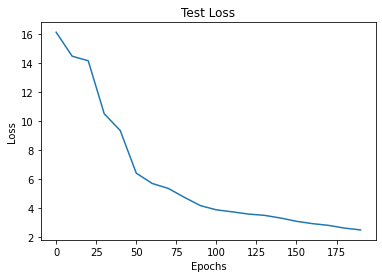
 

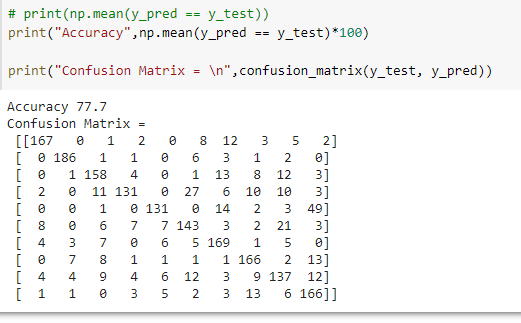
iv) Adagrad

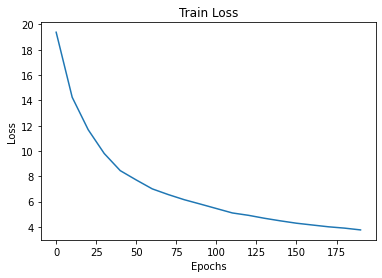
v) RMSprop

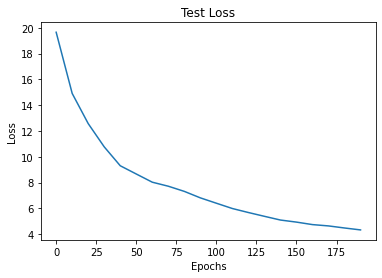
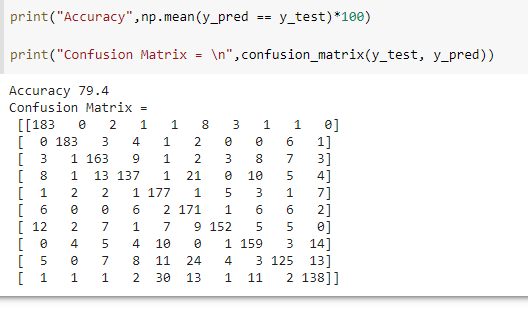




****

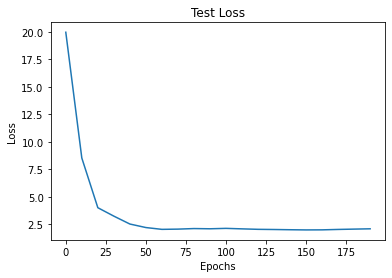
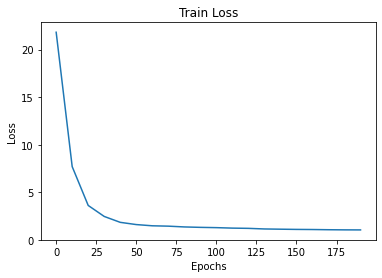
vi) Adam

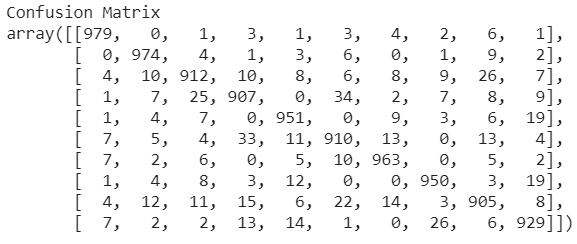


****

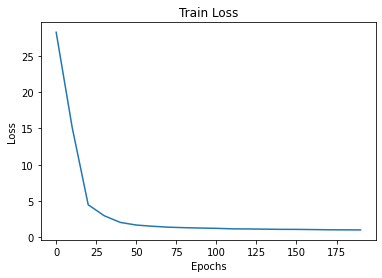
**Q2 ) Results of Weight Initialization Techniques:**We observe NAG with ReLU gives the best accuracy

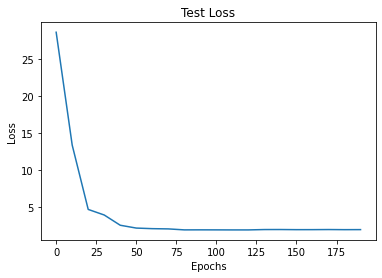
1. **He initialization**

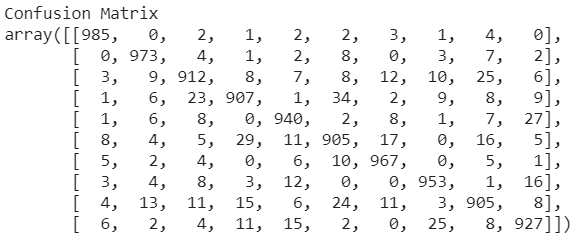
****

****

1. **Xavier Initialization**

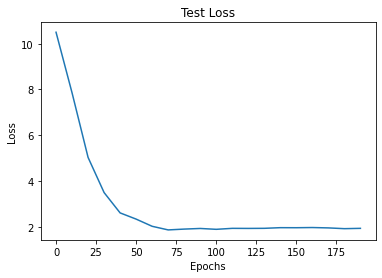
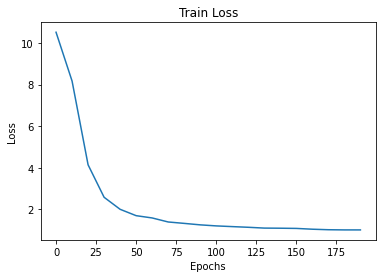
****

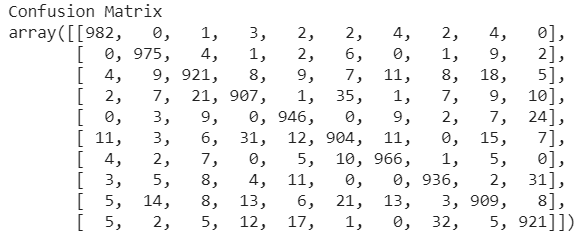
****

****

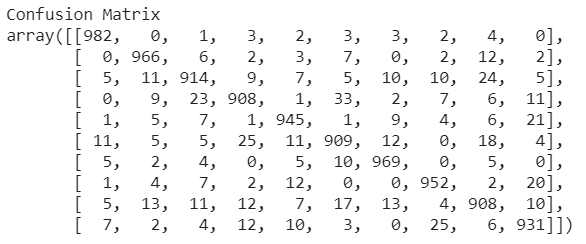
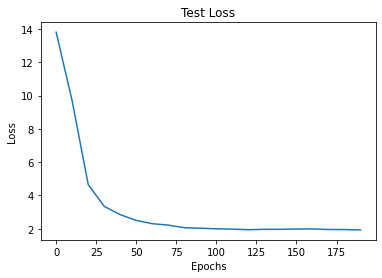
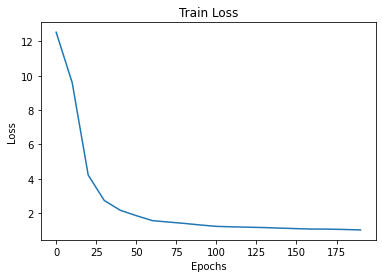
**Q3 ) Results of Regularization Techniques:**We observe both He and Xavier Initialization giving similar results, but since we are using ReLU activation we will proceed with

1. **L1**

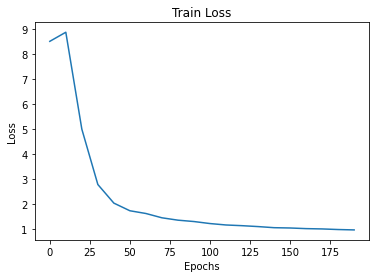
****

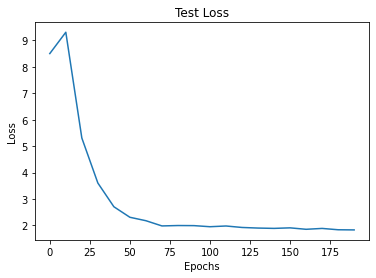
****

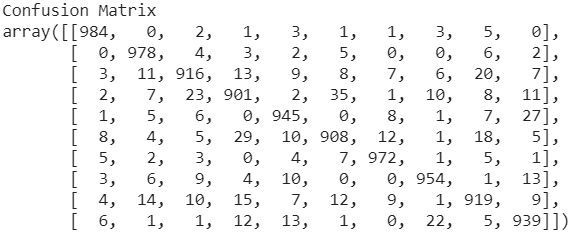
1. **L2**

****

1. **Dropout**

****

****

****

**Conclusion**

**We observe that for optimizers Nesterov’s gradient descent performs best for both test and train accuracy and takes somewhere around 25-30 epochs to converge. Followed by gradient descent with momentum and adagrad. Adam and RMSprop don’t perform as well but still outperform standard gradient descent.**

**For weight initialization, both He and Xavier perform equally well for all practical purposes. Both show a slight improvement over random initialization.**

**Dropout regularization performs slightly better than L1 and L2 regularisation, both of which perform equally well.**

**With standard training, it takes up to 500 epochs to reach 85% accuracy, which we outperform significantly in only 25-30 epochs when using optimizers and regularization.**