Problem 1(d)

Linear Activation

|  |  |  |
| --- | --- | --- |
| TimeTaken | Architecture | Accuracy |
| 5.750703096 | 50,2 | 83.62% |
| 4.812809944 | 50,50,2 | 83.84% |
| 6.165680885 | 50,50,50,2 | 84.10% |
| 7.49960804 | 50,50,50,50,2 | 84.32% |

Best Architecture = [50,50,50,50,2]

Total Time taken in training : 24.219 sec

Accuracy increase with increase in complexity/layers but not significantly

|  |  |  |
| --- | --- | --- |
| TimeTaken | Architecture | Accuracy |
| 6.623776197 | 50,50,2 | 84.18% |
| 9.731425047 | 50,500,2 | 83.80% |
| 29.18974996 | 50,500,300,2 | 84.58% |
| 71.31847501 | 50,800,500,300,2 | 84.72% |
| 145.3290951 | 50,800,800,500,300,2 | 84.87% |

Total Time taken : 262.19 sec and base architecture is [50,800,800,500,300,2]

Here again complexity does not increase significantly with the increase in layers. Hence in case of Linear activation we can say that accuracy does not significantly with increase of layers. and time taken increase is approx. exponential with the more layers in case of Linear activation

Problem 1(e) Sigmoid Activation

|  |  |  |
| --- | --- | --- |
| TimeTaken | Architecture | Accuracy |
| 10.68606091 | 50,50,2 | 73.91% |
| 27.88713598 | 50,500,2 | 76.57% |
| 55.71131706 | 50,500,300,2 | 71.74% |
| 134.364043 | 50,800,500,300,2 | 71.74% |
| 297.4557581 | 50,800,800,500,300,2 | 71.74% |

Best Architecture is [50, 500, 2]

Total Time Taken is 526.10 sec

**Time Taken (Linear vs Sigmoid Acitvation)**

We can see that time taken for the same layes is double in case of Sigmoid activation.

**Accuracy (Linear Vs. Sigmoid Activation)**

Accuracy decreases in case of Sigmoid when compared to Linear activation but remains almost same when compared to sigmoid activation

why this trend is different from that of linear activations

Problem 1(f) Relu Activation

|  |  |  |
| --- | --- | --- |
| TimeTaken | Architecture | Accuracy |
| 11.2527442 | 50,50,2 | 80.26% |
| 13.09005499 | 50,500,2 | 82.32% |
| 31.39364719 | 50,500,300,2 | 81.70% |
| 73.61942387 | 50,800,500,300,2 | 80.48% |
| 145.824964 | 50,800,800,500,300,2 | 79.68% |
|  |  |  |

Total time : 275.18 sec

why this trend is di\_erent from that of linear activations.

Best architecture is [50, 500, 2]

**Time Taken (Linear vs Sigmoid Acitvation vs Relu Activation)**

Time taken in case of Linear and Relu is almost the same for same layer however the time taken in Relu or Linear is half as that of Sigmoid and when layers are increased then time for Sigmoid becomes double of Linear/Relu

**Accuracy (Linear vs Sigmoid Acitvation vs Relu)**

Accuracy for Linear/Relu is almost comparable for similar layers however accuracy of Relu remains lowers than both sigmoid/Linear model. In Relu activation, accuracy decreases with increasing the number of layers.

Problem 1(g) L2-Regularization

TimeTaken Architecture Accuracy Regularization

80.53353 50,800,500,300,2 79.72% 10^-7

157.2322819 50,800,500,300,2 80.61% 5 \* 10^-7

245.332484 50,800,500,300,2 80.73% 10^-6

331.523109 50,800,500,300,2 81.26% 5 \* 10^-6

427.041451 50,800,500,300,2 80.77% 10^-5

Best value regularization = 5 \* 10^-6

Total Time Taken : 1241.66 sec

**Accuracy wrt Regularization value keeping Architecture same**

As can be observed keeping the same value of architecture and increasing regularization does not changes accuracy trend with a great difference. Although we can see an increasing trend in general but the difference or accuracy gain is not very significant.

**Time wrt Regularization value keeping Architecture same**

As clear from the trend that time taken in training increases with decrease of regularization value keeping layers same. The training time increases approc by one and half time with each time regularization value is changed

Problem 1(h) Early stopping and L2-Regularization

explain the trend of observations. Report the best value of the regularization hyperparameter this time.Is it the same as with only L2-regularization? Did early stopping help?

TimeTaken Architecture Accuracy Regularization

89.02349806 50,800,500,300,2 78.72% 10^-7

109.2510221 50,800,500,300,2 75.70% 5\*10^-7

178.7063761 50,800,500,300,2 79.23% 10^-6

248.2455261 50,800,500,300,2 78.48% 5\*10^-6

322.592227 50,800,500,300,2 80.12% 10^-5

Total Time Taken : 322 sec

Best Value : 10^-5

As we can see that time taken to model for the same layer and same regularization value decreases because of early stopping.

No, the value is not the same as with L2-regularization without early stopping as in previous case.

Yes, it helps as we can see that accuracy percentage does not change significantly for the same value of regularization

Problem 1(i)

TimeTaken Architecture Accuracy Decay

289.094275 50,800,500,300,2 72.69% 10^-5

570.8413081 50,800,500,300,2 75.90% 5\*10^-5

872.6115921 50,800,500,300,2 72.41% 10^-4

1143.773129 50,800,500,300,2 45.29% 3\*10^-4

1427.690883 50,800,500,300,2 37.15% 7 \*10^-4

1695.150869 50,800,500,300,2 52.35% 10^-3

Best Model is with decay 5\*10^-5

Problem 1(j)

TimeTaken Architecture Accuracy Momentum

141.929801 50,800,500,300,2 85.39% 0.99

280.1011569 50,800,500,300,2 80.62% 0.98

421.0175378 50,800,500,300,2 77.62% 0.95

561.1003819 50,800,500,300,2 74.63% 0.90

695.980783 50,800,500,300,2 71.45% 0.85

Best value of momentum is 0.99

Problem 1(k)

Optimal regularization Coefficient: 5\*(10^-6)

Optimal decay: 5\*(10^-5)

Optimal momentum coefficient: 0.99

|  |  |  |
| --- | --- | --- |
| TimeTaken | Architecture | Accuracy |
| 30.1751771 | 50,800,500,300,2 | 76.69% |