Assignment 1: Recognition of Handwritten Digits by Multilayer Perceptrons

Learning rate = input learning rate \* batch size \* ( 0.8 \*\* epoch times),即經過一次epoch 就將learning rate 乘上0.8

Version 1: neuron layers [16, 100, 100, 10], batch size = 1, epoch = 18, initial learning rate = 0.0085

Version 2: neuron layers [16, 100, 100, 10], batch size = 2, epoch = 19, initial learning rate = 0.0085

Version 3: neuron layers [16, 100, 100, 10], batch size = 6, epoch = 20, initial learning rate = 0.0085

Version 4: neuron layers [16, 100, 100, 10], batch size = 1, epoch = 12, initial learning rate = 0.0075

Version 5: neuron layers [16, 100, 10], batch size = 1, epoch = 13, initial learning rate = 0.0085

Accuracy:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Version 1 | Version 2 | Version 3 | Version 4 | Version 5 |
| Accuracy | 93.33% | 86.99% | 84.01% | 93.37% | 83.61% |

Confusion matrix:

˙ Version 1:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 344 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 16 | 0 |
| 0 | 301 | 56 | 1 | 1 | 4 | 0 | 1 | 0 | 0 |
| 0 | 3 | 360 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 0 | 3 | 0 | 330 | 0 | 0 | 0 | 0 | 0 | 3 |
| 0 | 0 | 0 | 0 | 352 | 1 | 4 | 0 | 0 | 7 |
| 0 | 0 | 0 | 13 | 1 | 302 | 0 | 0 | 0 | 19 |
| 0 | 2 | 0 | 0 | 0 | 0 | 334 | 0 | 0 | 0 |
| 2 | 53 | 2 | 2 | 0 | 0 | 0 | 289 | 16 | 0 |
| 0 | 0 | 1 | 0 | 0 | 5 | 0 | 0 | 329 | 1 |
| 0 | 7 | 0 | 3 | 0 | 0 | 0 | 1 | 1 | 324 |

˙ Version 2:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 333 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 30 | 0 |
| 0 | 187 | 109 | 1 | 1 | 3 | 52 | 10 | 0 | 1 |
| 0 | 11 | 345 | 0 | 0 | 0 | 0 | 8 | 0 | 0 |
| 0 | 3 | 0 | 331 | 0 | 0 | 0 | 1 | 0 | 1 |
| 0 | 0 | 0 | 1 | 335 | 1 | 7 | 0 | 0 | 20 |
| 0 | 0 | 0 | 28 | 0 | 246 | 5 | 0 | 0 | 56 |
| 0 | 0 | 0 | 0 | 1 | 0 | 331 | 0 | 4 | 0 |
| 0 | 38 | 6 | 10 | 14 | 0 | 0 | 294 | 2 | 0 |
| 7 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 327 | 0 |
| 0 | 3 | 0 | 10 | 3 | 3 | 0 | 2 | 1 | 314 |

˙ Version 3:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 331 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 30 | 0 |
| 0 | 152 | 136 | 0 | 1 | 2 | 46 | 23 | 0 | 4 |
| 0 | 4 | 343 | 0 | 0 | 0 | 0 | 17 | 0 | 0 |
| 0 | 5 | 0 | 326 | 2 | 0 | 0 | 1 | 0 | 2 |
| 0 | 0 | 0 | 0 | 345 | 0 | 10 | 1 | 0 | 8 |
| 0 | 0 | 0 | 65 | 0 | 200 | 14 | 0 | 0 | 56 |
| 1 | 0 | 5 | 0 | 0 | 0 | 330 | 0 | 0 | 0 |
| 0 | 35 | 17 | 16 | 3 | 2 | 0 | 277 | 2 | 12 |
| 8 | 0 | 0 | 0 | 0 | 10 | 0 | 0 | 317 | 1 |
| 0 | 8 | 0 | 7 | 0 | 1 | 0 | 1 | 1 | 318 |

˙ Version 4:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 342 | 0 | 1 | 4 | 0 | 0 | 0 | 0 | 12 | 4 |
| 0 | 308 | 49 | 0 | 1 | 0 | 5 | 1 | 0 | 0 |
| 0 | 8 | 354 | 0 | 0 | 0 | 0 | 2 | 0 | 0 |
| 0 | 3 | 0 | 331 | 0 | 0 | 0 | 0 | 0 | 2 |
| 0 | 0 | 1 | 0 | 345 | 4 | 11 | 0 | 0 | 3 |
| 0 | 0 | 0 | 13 | 0 | 314 | 5 | 0 | 0 | 3 |
| 1 | 0 | 0 | 0 | 0 | 0 | 333 | 0 | 2 | 0 |
| 0 | 50 | 0 | 5 | 0 | 0 | 0 | 302 | 3 | 4 |
| 0 | 0 | 0 | 0 | 0 | 1 | 0 | 2 | 333 | 0 |
| 0 | 4 | 0 | 8 | 0 | 3 | 0 | 1 | 1 | 319 |

˙ Version 5:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 328 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 33 | 0 |
| 0 | 179 | 112 | 0 | 1 | 47 | 0 | 10 | 0 | 15 |
| 0 | 9 | 348 | 0 | 1 | 0 | 0 | 6 | 0 | 0 |
| 0 | 3 | 0 | 328 | 3 | 0 | 0 | 1 | 0 | 1 |
| 0 | 0 | 1 | 0 | 347 | 1 | 8 | 1 | 0 | 6 |
| 0 | 0 | 0 | 62 | 0 | 213 | 5 | 1 | 0 | 54 |
| 0 | 0 | 9 | 0 | 0 | 1 | 324 | 0 | 2 | 0 |
| 0 | 36 | 13 | 5 | 15 | 2 | 0 | 292 | 1 | 0 |
| 15 | 2 | 0 | 0 | 0 | 12 | 0 | 0 | 306 | 1 |
| 0 | 33 | 0 | 27 | 10 | 5 | 0 | 0 | 1 | 260 |

Conclusion:

mini-batch size大小對於數字辨識的效果,size越大,準確度越低

learning rate的改變,在epoch的次數中,有不錯的效果,但是起始的learning rate不是那麼容易去設定,以及其改變的大小在這裡使用較簡單的方法去實作,並沒有辦法使效果更快顯現出來

神經網路的層數對於效果的影響,則是較容易顯現出來,另外tanh函數對神經網路的效果比sigmoid函數的效果好很多

此次作業中,我覺得learning rate的設定與調整,可以再做更多改良,達到對神經網路的更好學習速度及效果提升