**2. programming**

1. Binary model

|  |  |  |
| --- | --- | --- |
| Max\_iters (Hidden\_size = 32) | Training Accuracy | Validation Accuracy |
| 40 | 0.921 | 0.475 |
| **60** | **0.940** | **0.970** |
| 80 | 0.941 | 0.480 |

|  |  |  |
| --- | --- | --- |
| Hidden\_size (Max\_iters = 60) | Training Accuracy | Validation Accuracy |
| 32 | 0.940 | 0.970 |
| 64 | 0.950 | 0.975 |
| **128** | **0.956** | **0.970** |

Findings:

For fixed hidden\_size, there are overfitting for max\_iter = 40 and max\_iter = 80. In the other word, when hidden\_size = 32 and max\_iters = 60, the model performs better than other two in the first table.

For fixed max\_iters, the performance of model will be improved by increasing hidden\_size, which is obvious in the second table.

It could conclude that when hidden\_size = 128 and max\_iters = 60, the binary model performs best.

1. Multi-class model

|  |  |  |
| --- | --- | --- |
| Max\_iters (Hidden\_size = 32) | Training Accuracy | Validation Accuracy |
| 300 | 0.951 | 0.850 |
| 400 | 0.951 | 0.852 |
| **500** | **0.962** | **0.859** |

|  |  |  |
| --- | --- | --- |
| Hidden\_size (Max\_iters = 400) | Training Accuracy | Validation Accuracy |
| 32 | 0.951 | 0.852 |
| 64 | 0.965 | 0.868 |
| **128** | **0.981** | **0.886** |

Findings:

For fixed hidden\_size, the gap between training accuracy and validation accuracy could be decreased by increasing max\_iters to some extent. Unfortunately, when max\_iters = 500, the gap becomes larger.

For fixed max\_iters, it is obvious that the gap between training accuracy and validation accuracy decreases for increasing hidden\_size.

It could conclude that to avoid overfitting, increasing hidden\_size has better effect.

1. Regularization model

|  |  |  |
| --- | --- | --- |
| Max\_iters (Hidden\_size = 32) | Training Accuracy | Validation Accuracy |
| 300 | 0.943 | 0.846 |
| 400 | 0.945 | 0.848 |
| 500 | 0.960 | 0.855 |

|  |  |  |
| --- | --- | --- |
| Hidden\_size (Max\_iters = 400) | Training Accuracy | Validation Accuracy |
| 32 | 0.945 | 0.848 |
| 64 | 0.957 | 0.847 |
| 128 | 0.978 | 0.877 |

To avoid overfitting, we could increase the size of training set or add some penalty term. I have tried the dropout for neural network, which randomly sets some neurons to zero. The probability of dropping is set as 0.5.

By comparing with multi-class model and regularization, we are expecting following two conclusions. One is that, with the same number of neurons, the gap between training accuracy and validation accuracy is decreased and the accuracy is improved by increasing times of iterations. The other is that, with the same max\_iters and hidden\_size, the accuracy is obviously advanced by employing randomly dropping. However, they could not be concluded by observed data. It might be a good idea to collect more training data.