1.

A) In the initial given code we had 2 layers in LSTM Neural Network, with a dropout of 0.05 and softmax activation function.

A screenshot of a computer program

Description automatically generated

b) Result of using different parameters in the LSTM model.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Number of Layers | Dropout | Activation Function | Batch Size | Number of Epochs | Training Accuracy | Testing Accuracy | Validation Accuracy | Unlabeled Accuracy | Prediction Accuracy |
| 2 | 0.05 | SoftMax | 32 | 100 | 73 | 68 | 83 | 78 | 69 |
| 2 | 0.1 | SoftMax | 32 | 100 | 72 | 69 | 81 | 79 | 69 |
| 2 | 0.3 | SoftMax | 32 | 100 | 71 | 68 | 79 | 79 | 68 |
| 2 | 0.5 | SoftMax | 32 | 100 | 70 | 64 | 77 | 77 | 64 |
| 2 | 0..05 | SoftMax | 32 | 200 | 72 | 68 | 83 | 80 | 68 |
| 2 | 0.1 | SoftMax | 32 | 200 | 72 | 69 | 85 | 78 | 69 |
| 2 | 0.3 | SoftMax | 32 | 200 | 71 | 68 | 77 | 78 | 68 |
| 2 | 0.5 | SoftMax | 32 | 200 | 71 | 68 | 77 | 78 | 68 |
| 3 | 0.05 | SoftMax | 32 | 100 | 73 | 66 | 83 | 78 | 68 |
| 3 | 0.1 | SoftMax | 32 | 100 | 74 | 71 | 83 | 73 | 71 |
| 3 | 0.3 | SoftMax | 32 | 100 | 71 | 68 | 79 | 79 | 68 |
| 3 | 0.5 | SoftMax | 32 | 100 | 72 | 68 | 79 | 78 | 68 |
| 3 | 0.05 | SoftMax | 32 | 200 | 74 | 66 | 83 | 76 | 66 |
| 3 | 0.1 | SoftMax | 32 | 200 | 73 | 68 | 81 | 76 | 66 |
| 3 | 0.3 | SoftMax | 32 | 200 | 71 | 68 | 81 | 79 | 68 |
| 3 | 0.5 | SoftMax | 32 | 200 | 71 | 68 | 81 | 79 | 68 |
| 2 | 0.05 | Sigmoid | 32 | 100 | 71 | 64 | 81 | 79 | 64 |
| 2 | 0.1 | Sigmoid | 32 | 100 | 71 | 69 | 81 | 79 | 69 |
| 2 | 0.3 | Sigmoid | 32 | 100 | 71 | 68 | 83 | 82 | 68 |
| 2 | 0.5 | Sigmoid | 32 | 100 | 69 | 68 | 85 | 83 | 68 |
| 2 | 0.05 | Sigmoid | 32 | 200 | 73 | 68 | 83 | 79 | 68 |
| 2 | 0.1 | Sigmoid | 32 | 200 | 72 | 68 | 81 | 79 | 68 |
| 2 | 0.3 | Sigmoid | 32 | 200 | 71 | 68 | 79 | 79 | 68 |
| 2 | 0.5 | Sigmoid | 32 | 200 | 70 | 66 | 81 | 77 | 66 |
| 3 | 0.05 | Sigmoid | 32 | 100 | 75 | 69 | 85 | 72 | 68 |
| 3 | 0.1 | Sigmoid | 32 | 100 | 73 | 68 | 83 | 78 | 68 |
| 3 | 0.3 | Sigmoid | 32 | 100 | 72 | 66 | 83 | 77 | 66 |
| 3 | 0.5 | Sigmoid | 32 | 100 | 71 | 68 | 81 | 79 | 68 |
| 3 | 0.05 | Sigmoid | 32 | 200 | 74 | 69 | 81 | 72 | 69 |
| 3 | 0.1 | Sigmoid | 32 | 200 | 73 | 68 | 81 | 78 | 68 |
| 3 | 0.3 | Sigmoid | 32 | 200 | 71 | 68 | 81 | 79 | 68 |
| 3 | 0.5 | Sigmoid | 32 | 200 | 71 | 68 | 77 | 79 | 68 |
| 2 | 0.05 | ReLU | 32 | 100 | 64 | 59 | 72 | 80 | 59 |
| 2 | 0.1 | ReLU | 32 | 100 | 69 | 68 | 79 | 81 | 68 |
| 2 | 0.3 | ReLU | 32 | 100 | 42 | 29 | 36 | 0 | 68 |
| 2 | 0.5 | ReLU | 32 | 100 | 42 | 29 | 36 | 0 | 68 |
| 2 | 0.05 | ReLU | 32 | 200 | 42 | 29 | 36 | 0 | 68 |
| 2 | 0.1 | ReLU | 32 | 200 | 42 | 29 | 36 | 0 | 68 |
| 2 | 0.3 | ReLU | 32 | 200 | 42 | 29 | 36 | 0 | 68 |
| 2 | 0.5 | ReLU | 32 | 200 | 42 | 29 | 36 | 0 | 68 |
| 3 | 0.05 | ReLU | 32 | 100 | 42 | 29 | 36 | 0 | 68 |
| 3 | 0.1 | ReLU | 32 | 100 | 42 | 29 | 36 | 0 | 68 |
| 3 | 0.3 | ReLU | 32 | 100 | 42 | 29 | 36 | 0 | 68 |
| 3 | 0.5 | ReLU | 32 | 100 | 42 | 29 | 36 | 0 | 68 |
| 3 | 0.05 | ReLU | 32 | 200 | 42 | 29 | 36 | 0 | 68 |
| 3 | 0.1 | ReLU | 32 | 200 | 42 | 29 | 36 | 0 | 68 |
| 3 | 0.3 | ReLU | 32 | 200 | 42 | 29 | 36 | 0 | 68 |
| 3 | 0.3 | ReLU | 32 | 200 | 42 | 29 | 36 | 0 | 68 |

The highlighted cell is having the highest percentage of test and prediction accuracy so I am considering those parameters and using the same to compare RNN, GRU and LSTM models in 2nd Question.

A screenshot of a computer

Description automatically generated

2.

LSTM Model

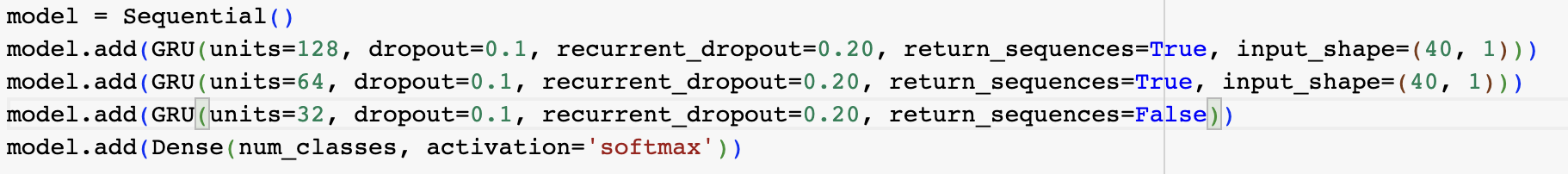
|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Number of Layers | Dropout | Activation Function | Batch Size | Number of Epochs | Training Accuracy | Testing Accuracy | Validation Accuracy | Unlabeled Accuracy | Prediction Accuracy |
| 3 | 0.1 | SoftMax | 32 | 100 | 74 | 71 | 83 | 73 | 71 |

A screenshot of a computer

Description automatically generated

GRU Model

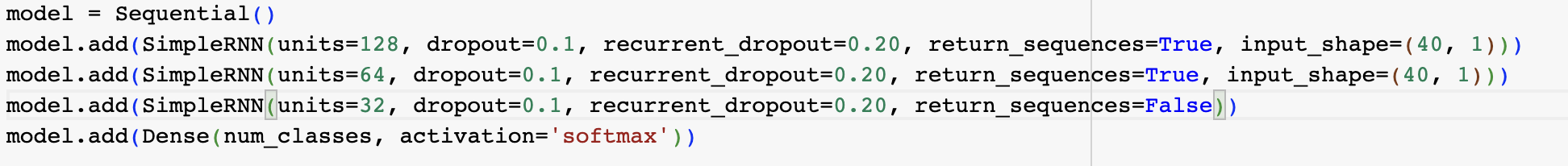
|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Number of Layers | Dropout | Activation Function | Batch Size | Number of Epochs | Training Accuracy | Testing Accuracy | Validation Accuracy | Unlabeled Accuracy | Prediction Accuracy |
| 3 | 0.1 | SoftMax | 32 | 100 | 77 | 71 | 75 | 71 | 69 |



Simple RNN Model

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Number of Layers | Dropout | Activation Function | Batch Size | Number of Epochs | Training Accuracy | Testing Accuracy | Validation Accuracy | Unlabeled Accuracy | Prediction Accuracy |
| 3 | 0.1 | SoftMax | 32 | 100 | 69 | 69 | 77 | 81 | 69 |

Out of all the 3 models LSTM model has good test and prediction accuracy.



3.

LSTM Model without any regularization

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Number of Layers | Dropout | Activation Function | Batch Size | Number of Epochs | Training Accuracy | Testing Accuracy | Validation Accuracy | Unlabeled Accuracy | Prediction Accuracy |
| 3 | 0.1 | SoftMax | 32 | 100 | 74 | 71 | 83 | 73 | 71 |

A screenshot of a computer program

Description automatically generated

With Early Stopping

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Number of Layers | Dropout | Activation Function | Batch Size | Number of Epochs | Training Accuracy | Testing Accuracy | Validation Accuracy | Unlabeled Accuracy | Prediction Accuracy |
| 3 | 0.1 | SoftMax | 32 | 100 | 71 | 68 | 83 | 80 | 68 |

A screenshot of a computer

Description automatically generated

A computer screen shot of a computer code

Description automatically generated