CS 1699

HW 3

Mathew Varughese

**Hyperparameters Explored:**

Dropout Rate

Optimizer

Regularization (Weight Decay)

Learning Rate

All accuracy values are the performance on the test data set.

Here, I tested different optimizers (Adam/Adamax)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Hidden Size** | **Batch Size** | **Dropout** | **Optimizer** | **Weight Decay** | **Learning Rate** | **Accuracy** |
| 500 | 100 | - | Adam | - | 0.001 | 17.45 |
| 500 | 100 | - | Adamax | - | 0.001 | 29.17 |

Here, I tested different rates of dropout (p=0.2 and p=0.5)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Hidden Size** | **Batch Size** | **Dropout** | **Optimizer** | **Weight Decay** | **Learning Rate** | **Accuracy** |
| 500 | 100 | p=0.2 | Adamax | - | 0.001 | 17.06 |
| 500 | 100 | p=0.5 | Adamax | - | 0.001 | 10.07 |

Here, I tested different weight\_decay ‘s for the optimizer

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Hidden Size** | **Batch Size** | **Dropout** | **Optimizer** | **Weight Decay** | **Learning Rate** | **Accuracy** |
| 500 | 100 | - | Adamax | 0.5 | 0.001 | 30.7 |
| 500 | 100 | - | Adamax | 0.2 | 0.001 | 23.73 |

Here, I tested different learning rates

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Hidden Size** | **Batch Size** | **Dropout** | **Optimizer** | **Weight Decay** | **Learning Rate** | **Accuracy** |
| 500 | 100 | - | Adamax | 0.5 | 0.00001 | 30.74 |
| 500 | 100 | - | Adamax | 0.5 | 0.00006 | 33.54 |