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 (SGD/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 |
| 500 | 100 | - | SGD | - | 0.001 | 11.96 |

Here, I tested different rates of dropout

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **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 |
| 500 | 100 | p=0.8 | Adamax | - | 0.001 | 9.43 |

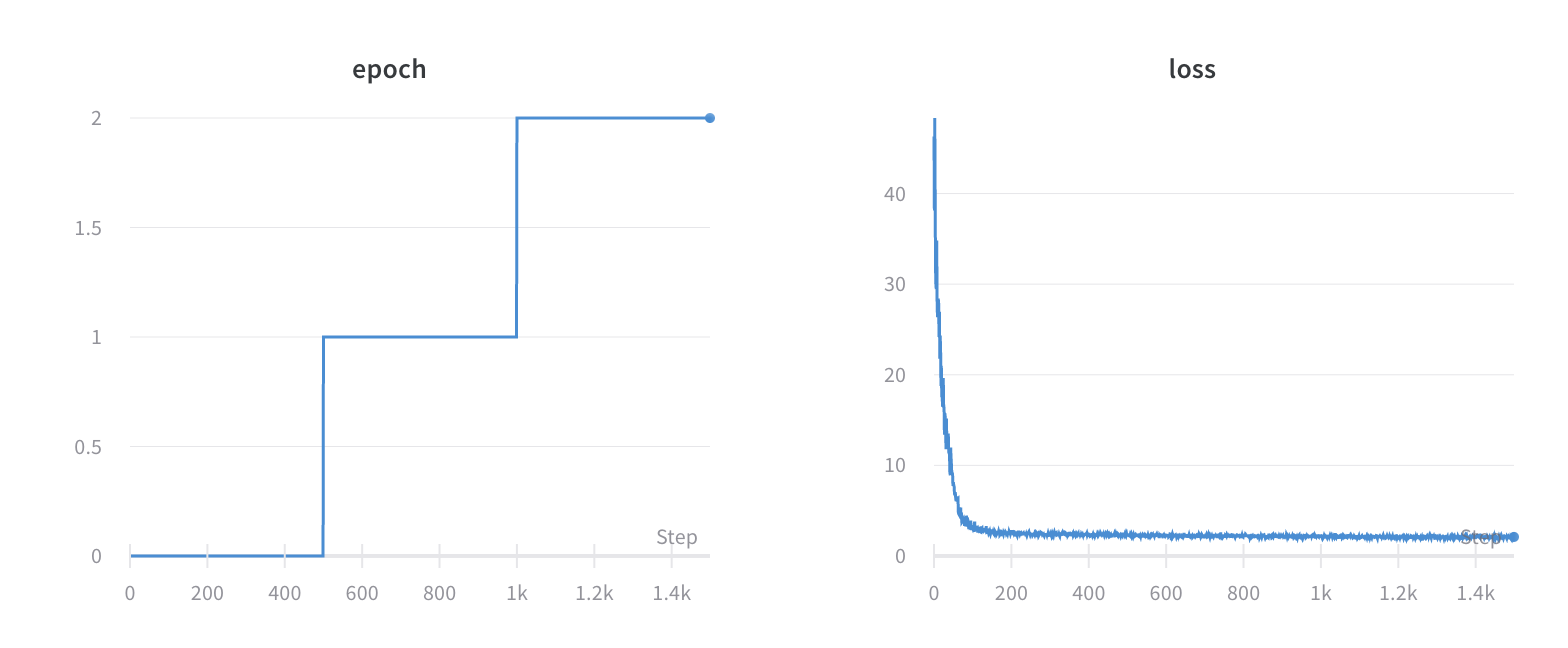
Here, I tested different weight\_decay ‘s for the optimizer

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **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 |
| 500 | 100 | - | Adamax | 0.55 | 0.001 | 29.8 |

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 | 0.0002 | 19.39 |
| 500 | 100 | - | Adamax | 0.5 | 0.00006 | 24.22 |

Charts from wandb tool:



This was for my last run. The final loss was 2.1061 on the training set.

**Part III**

Fine Tuning: 80.06%

Freezing: 65.1%

I calculated all of these accuracy’s based on the test set.