Loss with different optimizers (different learning rates)

(Patience set for early stopping = 3)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Epochs | SGD | | Adam(lr=0.001) | | Adam(lr=0.0001) | | Adam(lr=0.001, weight\_decay = 0.001) | |
| Training | Validation | Training | Validation | Training | Validation | Training | Validation |
| 1 | 2.12 | 2.25 | 1.76 | 1.48 | 1.81 | 1.6 | 1.76 | 1.68 |
| 2 | 1.81 | 1.71 | 1.33 | 1.28 | 1.52 | 1.46 | 1.42 | 1.31 |
| 3 | 1.57 | 1.53 | 1.01 | 1.15 | 1.34 | 1.34 | 1.27 | 1.29 |
| 4 | 1.39 | 1.62 | 0.658 | 1.31 | 1.2 | 1.3 | 1.16 | 1.24 |
| 5 | 1.24 | 1.54 | 0.368 | 1.57 | 1.07 | 1.21 | 1.07 | 1.32 |
| 6 | 1.09 | 1.48 | 0.174 | 2.16 | 0.947 | 1.17 | 0.961 | 1.24 |
| 7 | 0.906 | 1.69 |  |  | 0.823 | 1.16 | 0.839 | 1.43 |
| 8 | 0.693 | 1.71 |  |  | 0.71 | 1.26 |  |  |
| 9 | 0.461 | 1.87 |  |  | 0.598 | 1.21 |  |  |
| 10 |  |  |  |  | 0.486 | 1.26 |  |  |
| Accuracy: | 51.83% | | 61.87% | | 61.95% | | 58.14% | |

Red marks indicate the models saved before overfitting (i.e., the points at which best model is saved)