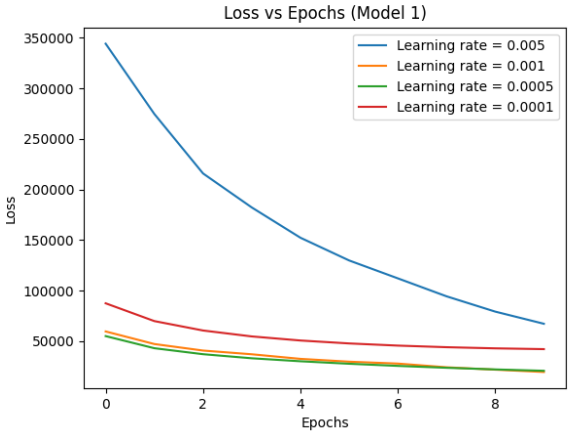
Machine Learning Assignment Report

Sk. Sabit Bin Mosaddek (1805106)

Model 1

- Layers:
  - Batch Normalization
  - Dense Layer with 512 Node (optimizer: Nesterov ADAM)
  - ReLU Layer
  - DropOut Layer with drop rate 0.5
  - Dense Layer with 26 Node (optimizer: Nesterov ADAM)
  - Softmax
- Mini Batch Gradient Descent (size = 64)
- 10 Epochs
- Cross Entropy Loss Function
- Xavier Initialization



Performance measures: (Training)

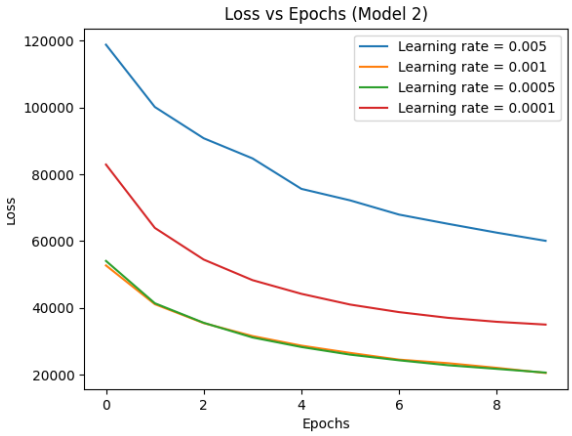
Performance measure	0.005	0.001	0.0005	0.0001
Accuracy (%)	92.2511	95.6354	95.5892	91.7694
F1-macro	0.9226	0.9565	0.9560	0.9177
Loss	38084.32	19499.01	20691.61	42103.74

Performance measures: (Validation)

Performance measure	0.005	0.001	0.0005	0.0001
Accuracy (%)	89.0224	89.2735	89.5566	88.4935
F1-macro	0.8899	0.8926	0.8954	0.8845
Loss	10640.85	11279.05	9730.20	10211.11

Model 2

- Layers:
    - Batch Normalization
    - Dense Layer with 512 Node (optimizer: ADAM)
    - ReLU Layer
    - DropOut Layer with drop rate 0.5
    - Dense Layer with 256 Node (optimizer: ADAM)
    - ReLU Layer
    - Batch Normalization
    - DropOut Layer with drop rate 0.3
    - Dense Layer with 128 Node (optimizer: ADAM)
    - ReLU Layer
    - Dense Layer with 26 Node (optimizer: ADAM)
    - Softmax
- Mini Batch Gradient Descent (size = 64)
- 10 Epochs
- Cross Entropy Loss Function
- Xavier Initialization



Performance measures: (Training)

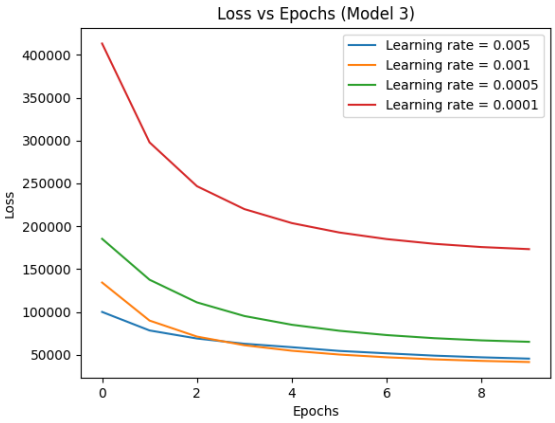
Performance measure	**0.005**	**0.001**	**0.0005**	**0.0001**
Accuracy (%)	91.8175	95.4930	95.5222	92.8525
F1-macro	0.91821	0.9551	0.9554	0.9287
Loss	37347.05	20561.62	20683.13	35048.62

Performance measures: (Validation)

Performance measure	**0.005**	**0.001**	**0.0005**	**0.0001**
Accuracy (%)	90.5021	90.8066	90.8653	89.5886
F1-macro	0.9047	0.9079	0.9086	0.8958
Loss	7791.31	8976.74	8623.74	8958.32

Model 3

- Layers:
    - Dense Layer with 512 Node (optimizer: None)
    - ReLU Layer
    - DropOut Layer with drop rate 0.5
    - Dense Layer with 256 Node (optimizer: None)
    - ReLU Layer
    - DropOut Layer with drop rate 0.2
    - Dense Layer with 128 Node (optimizer: None)
    - Sigmoid Layer
    - Dense Layer with 26 Node (optimizer: None)
    - Softmax



- Mini Batch Gradient Descent (size = 64)
- 10 Epochs
- Cross Entropy Loss Function
- Xavier Initialization

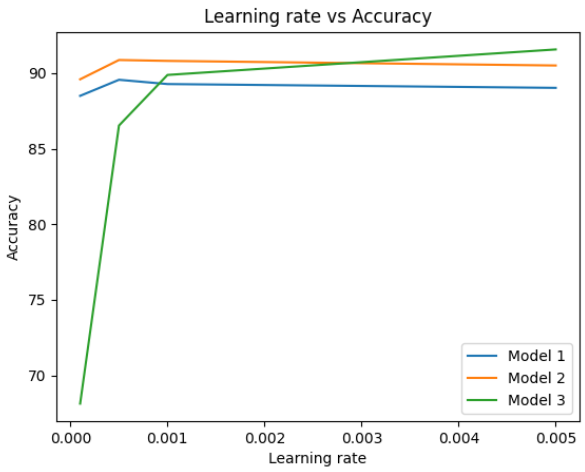
Performance measures: (Training)

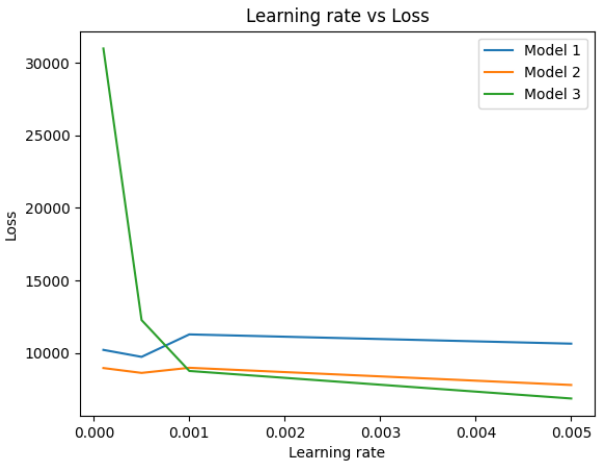
Performance measure	**0.005**	**0.001**	**0.0005**	**0.0001**
Accuracy (%)	93.4662	91.6619	87.4490	68.5275
F1-macro	0.93469	0.9166	0.8744	0.6818
Loss	29137.76	41503.76	65086.656	173258.39

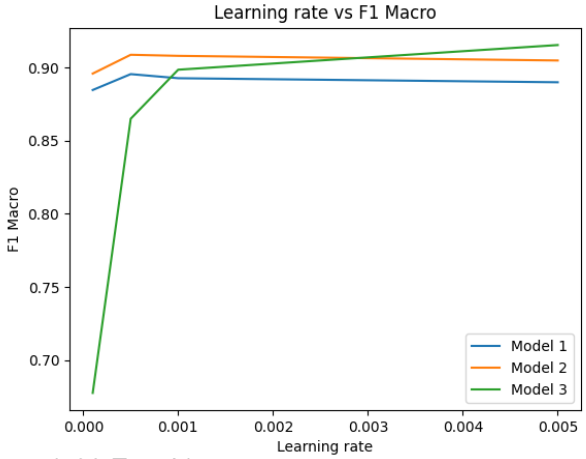
Performance measures: (Validation)

Performance measure	**0.005**	**0.001**	**0.0005**	**0.0001**
Accuracy (%)	91.5651	89.8771	86.5277	68.1250
F1-macro	0.9153	0.8984	0.8649	0.6777
Loss	6858.61	8758.06	12268.97	30996.81

Model Comparison







Top 2 Models:

1. Model 3 with 0.005 Learning Rate
2. Model 2 with 0.0005 Learning Rate

Model 3 with 0.005 Learning Rate Performance measures (500 Epoch):

Performance measure	Train	Validation
Accuracy (%)	98.3861	92.8205
F1-macro	0.9839	0.9278
Loss	5840.14	6724.59

Model 3 with 0.005 Learning Rate Performance measures (500 Epoch):

Performance measure	Train	Validation
Accuracy (%)	96.9165	92.3985
F1-macro	0.9692	0.9235
Loss	14369.79	7532.96

Best Model

Model 3 with learning rate 0.005 (Epoch 500)

Performance measures:

Performance measure	Train	Validation	Test
Accuracy (%)	98.3861	92.8205	92.8846
F1-macro	0.9839	0.9278	0.9287
Loss	5840.14	6724.59	8063.42

Confusion Matrix:

Letters↓	TP	TN	FP	FN
A	608	19822	178	192
B	715	19890	110	85
C	711	19918	82	89
D	641	19905	95	159
E	691	19886	114	109
F	709	19887	113	91
G	433	19923	77	367
H	663	19847	153	137
I	524	19777	223	276
J	701	19868	132	99
K	666	19885	115	134
L	604	19707	293	196
M	704	19965	35	96
N	646	19841	159	154
O	753	19897	103	47
P	734	19917	83	66
Q	666	19638	362	134
R	656	19885	115	144
S	721	19947	53	79
T	712	19888	112	88
U	698	19908	92	102
V	718	19859	141	82
W	752	19899	101	48
X	697	19927	73	103
Y	684	19886	114	116
Z	721	19956	44	79