Assignment 4 - ECEN 689

Name: Prabhasa Kalkur

UIN: 127003549

General Organization of the report: Problem Description, Setup, Results & Analysis, Tables

PROBLEM 1

Problem Description - To design non-linear classifiers for **large Dataset** using NNs and SVMs

a) Different number of hidden layers (1, 2, 3) were used for the non-linear classifiers. The input and output layers had **64 ReLu and 4 Softmax** neurons, respectively. Each hidden layer has **128 ReLu** neurons (This choice of activation layers was motivated by the Stanford CS231n assignments). The number of parameters can be calculated accordingly.
b) Keras was used for the **NNs**. **Binary cross-entropy** loss was used for three (SGD, RMSprop, Adam) different optimizers
c) The SVC function was used for **SVMs**. Sigmoid, RBF and Polynomial **Kernels** were tried and tested.
d) Accuracy of the classifier is the ratio of correctly estimated labels to true labels (the test label depends on the 'threshold' of the classifier)

Setup – **33%** of the data is taken as test data. Batch size for the NNs is 10. Number of Epochs is 10 as well. Tolerance for SVM was set to 10^{-5}

Results – Classifier accuracy is the performance metric. A lot of parameters had to be varied, so I have attached a table of the results. The analysis below is common to all subproblems:

a) **SGD is a poor choice** for both noiseless and noisy datasets, compared to RMSprop and Adam (and possibly AdaGrad).
b) **RBF performs really well**, whereas Sigmoid is a poor choice. This is most likely owing to the size of the dataset.

Neural Networks

		1a	1a	1b	1b	1c	1c
Hidden layers	**Optimizer**	**Train_acc**	**Test_acc**	**Train_acc**	**Test_acc**	**Train_acc**	**Test_acc**
1	SGD	77.31	76.83	71.98	70.28	52.35	52.88
	RMSprop	86.98	92.83	78.58	79.73	61.03	65.93
	Adam	88.03	90.83	80	79.18	64.53	61.53
2	SGD	96.93	96.88	86.13	86.33	68.55	67.83
	RMSprop	97.36	97.33	87.88	86.63	72.41	71.11
	Adam	98.13	98.33	86.86	84.28	72.06	71.73
3	SGD	96.36	96.53	86.56	87.68	71.36	70.13
	RMSprop	96.96	96.33	86.85	85.43	72.16	74.08
	Adam	97.66	97.43	86.7	86.13	73.11	70.33

SVMs

Kernel	Scores - 1a	Scores - 1b	Scores - 1c
Sigmoid	14.46	6.73	12.94
RBF	93.09	83.13	65.55
Poly-3	53.89	47.15	42.94
Poly-4	62.11	55.35	44.72
Poly-5	51.92	47.15	39.31
Poly-6	47.81	43.21	37.45

PROBLEM 2

Problem Description - Design non-linear classifiers for **MNIST Dataset** using NNs and SVMs

a) Different number of hidden layers (1, 2, 3) were used for the non-linear classifiers. The input and output layers had **64 ReLu and 10 Softmax** neurons, respectively. Each hidden layer has **128 ReLu** neurons

b) Keras was used for the **NNs**. **Categorical cross-entropy** loss was used for three (SGD, RMSprop, Adam) different optimizers

c) Sigmoid, RBF and Polynomial **Kernels** were tried and tested for the **SVMs**.

d) Accuracy of the classifier is the ratio of correctly estimated labels to true labels (the test label depends on the 'threshold' of the classifier)

Setup – Batch size for the NNs is 64. Number of Epochs is 10. SVM tolerance is set to 10^{-5}

Results – a) All optimizers are **equally good choices** for the NN!

b) **RBF performs really well**, whereas **Sigmoid is a great choice for larger training to test ratio datasets**, but a poor choice for smaller training to test ratio datasets.

Subproblem 1 - MNIST: 50000 training + 10000 test samples

Neural Networks

Hidden layers	Optimizer		Train_acc	Test_acc
1	SGD		93.51	83.66
	RMSprop		98.74	97.26
	Adam		98.97	97.29
2	SGD		95.19	95.16
	RMSprop		99.16	97.16
	Adam		99.24	97.29
3	SGD		96.03	95.93
	RMSprop		99.14	97.7
	Adam		99.22	97.09

SVMs

Kernel	Scores
Sigmoid	91.04
RBF	92.11
Poly-3	15.91
Poly-5	11.21

Subproblem 2 - MNIST-rot: 12000 training + 10000 test samples

Neural Networks

Hidden layers	Optimizer		Train_acc	Test_acc
1	SGD		79.63	54.35
	RMSprop		93.29	79.64
	Adam		93.43	76.03
2	SGD		80.92	56.79
	RMSprop		95.25	86.04
	Adam		96.26	85.75
3	SGD		82.06	57.35
	RMSprop		92.05	82.65
	Adam		96.66	85.32

SVMs

Kernel	Scores
Sigmoid	10.12
RBF	96.23
Poly-3	95.55
Poly-5	95.55

Improving the classifier by using the Rotated MNIST datasets: Concatenate and mix the two datasets (rotated and normal). Train the new training set on a NN with different parameters. The results are as shown above. **This is a much better classifier, as it can classify both normal and rotated images.**

PROBLEM 3

Problem Description - Design non-linear classifiers for **MRI Dataset** using NNs and SVMs

a) Different number of hidden layers (1, 2, 3) were used for the non-linear classifiers. The input and output layers had **16 ReLu and 2 Softmax** activations, respectively. Each hidden layer has **16 ReLu** activations
b) Keras was used for the **NNs**. **Cross-entropy** loss was used for three (SGD, RMSprop, Adam) different optimizers
c) Sigmoid, RBF and Polynomial **Kernels** were tried and tested for the **SVMs**.
d) Accuracy of the classifier is the ratio of correctly estimated labels to true labels (the test label depends on the 'threshold' of the classifier)

Setup – **15%** of the data is taken as test data. Batch size for the NNs is 18. Number of Epochs is 20. SVM tolerance is set to 10^{-5}

Results – a) All optimizers **perform equally poor** on the NN. This is owing to the smaller dataset, where every feature becomes important.
b) **Sigmoid performs really well**, owing to the high training to test ratio of datasets. All other SVM Kernels perform **equally poorly.**

Neural Networks

			Alzheimer	Alzheimer		Normal	Normal
Hidden layers		Optimizer		Train_acc	Test_acc		Train_acc	Test_acc
1		SGD		58.63	38.27		64.52	53.66
		RMSprop		74.81	38.27		39.52	45.96
		Adam		55.69	92.12		65.99	61.35
2		SGD		64.52	69.04		65.99	45.96
		RMSprop		51.28	61.35		67.46	45.96
		Adam		52.75	76.73		39.52	76.73
3		SGD		61.57	53.66		64.52	22.89
		RMSprop		60.1	45.96		35.1	61.35
		Adam		54.53	76.73		36.57	76.73

SVMs

Kernel	Alzheimer's Score	Normal's score
Sigmoid	61.45	30.68
RBF	23.07	23.07
Poly-3	23.07	23.07
Poly-5	23.07	23.07

PROBLEM 4

Problem Description - Design non-linear classifiers for **EEG Dataset** using NNs and SVMs

a) Different number of hidden layers (1, 2, 3) were used for the non-linear classifiers. The input and output layers had **64 ReLu and 6 Softmax** activations, respectively. Each hidden layer has **128 ReLu** activations

b) Keras was used for the **NNs**. **Binary cross-entropy** loss was used for three (SGD, RMSprop, Adam) different optimizers

c) Sigmoid, RBF and Polynomial **Kernels** were tried and tested for the **SVMs**.

d) Accuracy of the classifier is the ratio of correctly estimated labels to true labels (the test label depends on the 'threshold' of the classifier)

Setup – **33%** of the data is taken as test data. Batch size for the NNs is 10. Number of Epochs is 10. SVM tolerance is set to 10^{-5}

Results – a) All optimizers are **equally good choices** for the NN.

b) All other SVM Kernels perform **equally poorly.**

Neural Networks

Hidden layers	Optimizer		Train_acc	Test_acc
1	SGD		76.65	76.39
	RMSprop		80.47	79.51
	Adam		80.65	80.24
2	SGD		84.9	84.54
	RMSprop		84.81	81.7
	Adam		85.81	82.91
3	SGD		86.31	84.54
	RMSprop		88.69	85.64
	Adam		88.4	87.06

SVMs

Kernel	Scores
Sigmoid	18
RBF	22
Poly-3	41
Poly-5	38
Poly-7	35

Summary: There is no such thing as a “best” classifier for all datasets. It depends heavily on the choice of **optimizers**, **hyperparameters**, and varies with **type** (noisy, noiseless) and **size** (MNIST vs MRI) of dataset.