REPORT

CSE 512: Assignment 1

Chumki Acharya (112683478)

Performance:

Algorithm	Kernel	Dataset	Learning Rate	Training Error	Testing Error	Accuracy
SVM	linear	Breast Cancer	1.0	0.372	0.00	0.627
SVM	RBF	Breast Cancer	1.0	0.372	0.056	0.627
SVM	linear	MNIST	1.0	0.26	0.26	0.73
SVM	RBF	MNIST	1.0	0.092	0.092	0.98

SVM:

Linear kernel:

The soft-SVM with linear kernel (that is SVM with no kernel function) has been implemented using the stochastic gradient descent. It has been observed that linear kernel runs faster than RBF kernel. The weights calculated are dependent on different hyperparameter values. The change in the learning rate did not effect the accuracy, however changing the regularization strength(C) has also very less effect on the accuracy. For breast cancer data the linear kernel was much faster than the MNIST dataset but the accuracy has been found more better in MNIST dataset. In breast cancer data the labelling has been done as per the labels given in dataset and for MNIST the labelling is done such that for each digit whether it is a digit D or not.

RBF kernel:

The RBF kernel has been implemented with implementation of the RBF kernel function and calculation of gram matrix. It has been observed that the RBF kernel takes much longer time to execute than linear kernel. The accuracy of the model was not affected by RBF kernel for Breast cancer data, however for MNIST it has been seen slight variation. The gamma value taken for the computation kernel function was 0.7. Thus we can say RBF kernel works better in multidimension where the data is not linearly separable.

ReadMe:

SVM Algorithm

Breast cancer data:

- File Name: svm.py
- Kernel: linear/rbf
- Datasets: bcd
- input: file path to input dataset folder
- output: file path to the folder of output file
- Run following on command line mentioning the appropriate dataset file and mode:

python svm.py --kernel linear|rbf
--dataset bcd
--input /path/to/training/data/
--output /path/to/weightvector/

2. MNIST data

- Kernel: linear/rbf
- Datasets: mnist
- input: file path to input dataset folder
- output: file path to the folder of output file
- Run following on command line mentioning the appropriate dataset file and mode:

python svm.py --kernel linear|rbf
--dataset mnist
--input /path/to/training/data/
--output /path/to/weightvector/

Note: Keep the extracted .gz files in a "samples" folder in the input file path. It is necessary to read file using inbuilt functions of python-mnist

References:

- https://towardsdatascience.com/svm-implementation-from-scratch-python-2db2fc52e5c2#d7d8
- https://github.com/metpallyv/SVM-Kernels/blob/master/rbf_kernel.py
- https://stats.stackexchange.com/questions/215524/is-gradient-descent-possible-for-kernelized-svms-if-so-why-do-people-use-quadr
- https://www.youtube.com/watch?v=efR1C6CvhmE
- https://www.youtube.com/watch?v=05VABNfa1ds
- https://www.youtube.com/watch?v=wBVSbVktLIY
- https://www.youtube.com/watch?v=Qc5IyLW_hns