

III Report

1.

Performance metrics of the learning models –

A) Perceptron -

Dataset	Mode	Accuracy	Epochs	Loss/Error
Linearly-separable-dataset	ERM	1	18	0
Linearly-separable-dataset	Cross-validation	0.999	19	0.0008
Breast_cancer_data	ERM	0.85	50	0.15
Breast_cancer_data	Cross-validation	0.82	50	0.18

B) AdaBoost –

Dataset	Mode	Accuracy	No Of Weak learners	Loss/Error
Breast_cancer_dataset	ERM	0.94	6	0.056
Breast_cancer_dataset	Cross-validation	0.91	6	0.0875

2.

For the perceptron algorithm, when executing using the linearly-separable-dataset, the learning algorithm converges after certain epochs and 100% accuracy is observed. The same algorithm over the breast_cancer_dataset which is not linearly separable, it failed to converge and error rate fluctuated between 12%-25%. Increasing the number of iterations did not lead to better accuracy.

Tuning:

Initially, I tried training the perceptron using batches of data but the model didn't converge because on each modification of weights it was adding to many examples and it lead to the weights moving up and down with large magnitude. Then I tried point wise learning for perceptron.

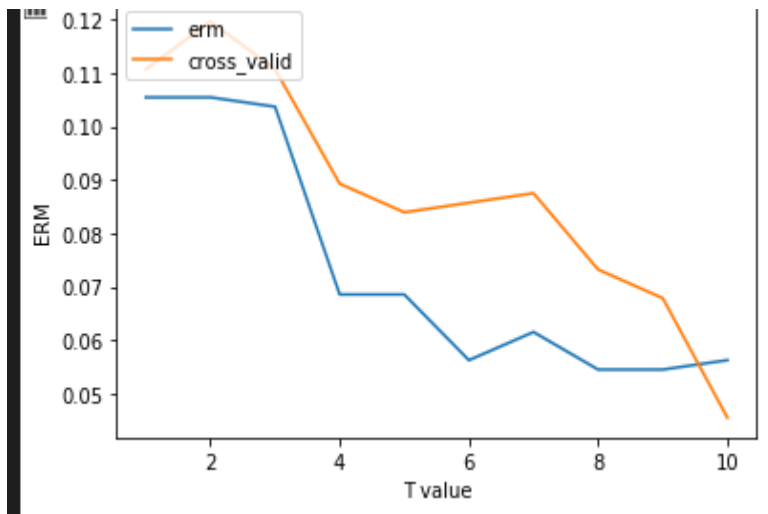
Perceptron is usually used when learning halfspaces. Perceptron is a powerful algorithm, stacking perceptrons in layers will give the most powerful Model => **Neural Network**

3.

I have tried different no of weak learners from 1-10 and found that the performance was good when $T=6$. Upon increasing the T to large number the erm error was very small (almost zero). But cross_valid error was very high. i.e, overfitting was observed.

Adaboost is a good algorithm as it gives control over bias-variance tradeoff in terms of Hyper parameter T .

Adaboost is computationally inexpensive as it learns weak classifiers which are fairly simple to learn.



4.

Unzip and extract the homework submission zip file. Open the code directory and in a new command prompt window (Windows)/ new terminal window(Linux/Mac) execute the code by providing dataset and mode as parameters-

1. `python perceptron.py --dataset linearly-separable-dataset.csv --mode erm`
2. `python perceptron.py --dataset linearly-separable-dataset.csv --mode cross_valid`
3. `python perceptron.py --dataset breast_cancer_data.csv --mode erm`
4. `python perceptron.py --dataset breast_cancer_data.csv --mode cross_valid`
5. `python adaboost.py --dataset breast_cancer_data.csv --mode erm`
6. `python adaboost.py --dataset breast_cancer_data.csv --mode cross_valid`