CS-5341 Pattern Recognition

Homework-2

Argument: Performance on small datasets

Traditional classifiers when given an input dataset of small size gives the best results but when working on a large dataset, the performance will be decreased. In the case of deep learning algorithms, when working on small datasets, the performance will be low. But when working on large datasets, the performance is getting decreased. The performance in the following experiment is shown in terms of Test Accuracy.

Experiment conducted:

Dataset used: IRIS (one with 150(small) data points and the other with 10877(considering it as large) points)

Algorithms: SVM and Deep Learning

i) **Training with small dataset:**

In this case, SVM is given IRIS dataset which has four features and 3 classes. Initially a small dataset was given for training which contains 150 datapoints. In this case, the test accuracy achieved is 97.36%.

Next the same IRIS dataset is given as input for training the deep learning model. In this case, the test accuracy achieved is 86.67%

ii) **Training with large dataset:**

In the above case, since only 150 datapoints were used to predict. But now using a dataset with 10877 records is considered a large dataset when compared to 150 records. Now, SVM is trained with this dataset containing 10877 datapoints and obtained the test accuracy of 70%.

Next the same IRIS dataset with 10877 data points is given as input to train the deep learning model. In this case, the test accuracy obtained is 87.27%.

By analyzing the results from cases (i) and (ii) we can conclude that traditional classifier performance decreases as the dataset size increases. This is clear because we have seen a drop in test accuracy from 97.36% to 70% upon increasing data points. We can also come to another conclusion that deep learning model performance increases as the size of the dataset increases. This is evident because we have seen a rise in test accuracy from 86.67 to 87.27% (on an average it is 88%).

Graphs:
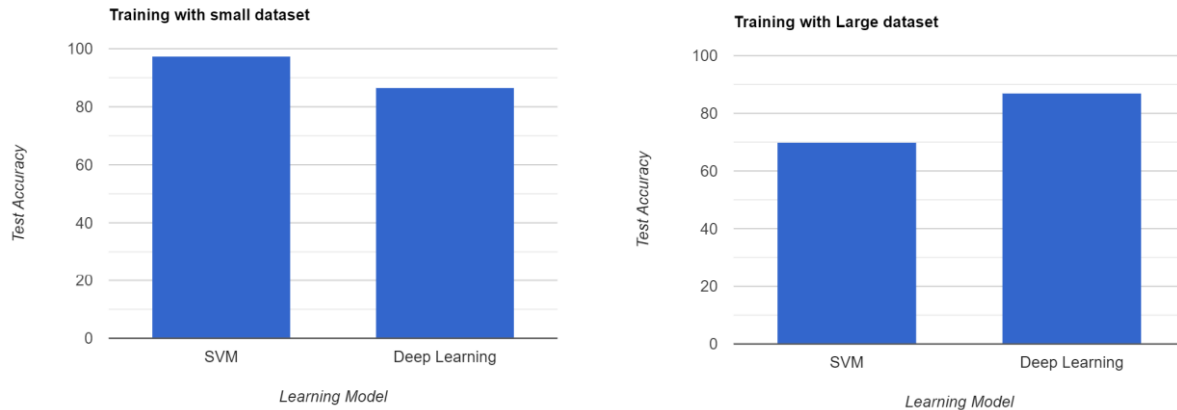
Argument: Robustness to Overfitting

Traditional machine learning models are less prone to overfitting as observed in the case of SVM model it always gave the expected results when the model is trained properly.

But in the case of deep learning models, when the input dataset is very small, and there are more epochs, then there is a chance that the model may learn again and again from the small training model itself making it perfect over the training set. But it may not give correct results for test set. As observed in my case, when the epochs are increased from 100 to 1000 on the 150 datapoints dataset, test accuracy is becoming less than train accuracy where train accuracy is nearly 99% and test accuracy is around 96% resulting in overfitting.

References:

SVM Code: https://github.com/amansk2050/svm-impementation-in-pyhton/

Deep Learning code: https://google.com, https://stackoverflow.com