Pattern Recognition - Tutorial 3 Report

Manousaridis Ioannis

Table 1: Accuracy Scores for MNIST Dataset before data augmentation

	SVM-linear	SVM-polynomial	SVM-RBF	SVM-sigmoid	KNN
Train Accuracy	99.34%	100%	100%	31%	95.65%
Test Accuracy	92.13%	97.4%	18.38%	31.7%	93.80%

Table 2: Accuracy Scores for Fashion-MNIST Dataset before data augmentation

	SVM-linear	SVM-polynomial	SVM-RBF	SVM-sigmoid	KNN
Train Accuracy	85.88%	20.64%	82.02%	80.60%	88.00%
Test Accuracy	83.86%	20.44%	81.24%	79.68%	83.60%

Table 3: Accuracy Scores for Wdbc Dataset without data augmentation

	SVM-linear	SVM-polynomial	SVM-RBF	SVM-sigmoid	KNN
Train Accuracy	96.85%	90.02%	90.02%	43.56%	93.43%
Test Accuracy	95.74%	92.55%	92.55%	45.212%	94.68%
5 Folds Cross Validation (mean value)	94.55%	90.86%	91.21%	46.04%	92.79%

Table 4: Accuracy Scores for MNIST Dataset after using data augmentation

	SVM-linear	SVM-polynomial	SVM-RBF	SVM-sigmoid	KNN
Train Accuracy	94.51%	52.50%	93.85%	91.61%	97.45%
Test Accuracy	93.24%	53.07%	92.89%	92.20%	95.73%

Table 5: Accuracy Scores for Fashion-MNIST Dataset after using data augmentation

	SVM-linear	SVM-polynomial	SVM-RBF	SVM-sigmoid	KNN
Train Accuracy	86.90%	64.51%	84.03%	82.34%	90.00%
Test Accuracy	85.60%	64.25%	83.56%	82.22	85.44%

Conclusions:

- For the **MNIST dataset with no augmentation** excellent accuracy scores were achieved by the SVM model with the polynomial and the linear kernel and by the KNN model. The best scores were given by the SVM model with the polynomial (100% training accuracy and 97.4% testing accuracy). The SVM model with rbf kernel provide excellent training accuracy and very low testing accuracy (overfitting). Also, the SVM model with the sigmoid kernel gave very poor results and it's not good for this dataset. When the dataset was tested with **augmented data** the results were a bit different. The SVM with the polynomial kernel had accuracy around 50% and all the other models become more stable and provide accuracy values around 95% (SVM with linear, rbf and sigmoid kernels, knn). In general the augmentation improved the SVM-sigmoid kernel model and deteriorated the performance of the SVM-polynomial kernel.

- For the **F-MNIST dataset** all the models gave results with accuracy around to 80-85% apart from the SVM model with the polynomial kernel. Its performance was poor. The highests score were achieved by the KNN model (88.00% training accuracy and 83.60% testing accuracy). When the dataset was tested with **augmented data** the results were improved in all models. The average accuracy is around 85%. Again the performance of the SVM with the polynomial kernel provided the lowest results with mean accuracy around 60%.

- For the **WDBC dataset** excellent results were achieved by all models apart from the SVM with the sigmoid kernel. Furthermore, this dataset was tested also with 5-Folds Cross Validation. However, this technique didn't improve the already high results. The highest scores were given by the SVM with the linear kernel.

- In general the svm and the knn models had very good results in the three datasets proving. It seems that both non-linear and linear methods can provide good results in those datasets. The data augmentation is used for improvement of the accuracy. This technique was really good for the F-MNIST dataset. This is not always the case as shown with the mnist dataset were the results varied a lot (some models improved, some were stable and one got worse). The SVM models with the linear and the sigmoid kernels seems to be unstable. The SVM model with the linear kernel and the K-NN model were stable at all experiments and gave the highest results. The SVM model with the rbf kernel was also very good but in one case provide bad results (MNIST dataset with no augmented data - overfitting).

- To conclude the knn and svm-linear kernel were the best models.