Deniz Aytemiz

Machine Learning Algorithms for Image Classification of hand digits and face recognition dataset

In this paper 5 different algorithms for image classification of hand digits and face recognition dataset are applied and each algorithm is discussed as well as their respective performance in this task. I find the algorithms and their modeling for specific purposes interesting that is why I choose this paper to read.

Image classification can be realized either with supervised or unsupervised methods. Supervised methods train learning models on datasets of images or some kind of spectral significant data to classify. Whereas unsupervised methods leverage image classification algorithms which provide clusters and metrics of performances such that no analyst is needed in the process.

Five machine learning algorithms are discussed in this paper are as follows; Nearest Class Centroid (NCC) classifier, Nearest Sub Class (NSC) Centroid Classifier, Nearest Neighbor Classifier Perceptron trained using Backpropagation and Perceptron trained using MSE.

In Nearest Class Centroid Classifier, each class has clusters of samples and the centroid is taken into consideration. When an observation is made it is mapped to whichever classes centroid it is most proximate to. Data conceptually is modeled as vector and centroid of data is also vector. Therefore in order to determine the proximity cosine between two vectors is used as a metric. The resemblance determines the class of the new observation.

Nearest Sub Class Centroid classifier is based on Maximum Variance Cluster Algorithm. The algorithm aims to firstly determine the clusters and means such that the mean squared error of each object with respect to the mean is minimized. However this may cause overfitting of the data, hence to prevent this it also imposes joint variance constraint where the variance of every instance have ro be lower than the constrained variance value.

K-Nearest Neighbor Classifier, use Euclidean distance as a metric. It is calculated between test samples to be classified and neighboring training sample group/cluster specified by the number k. Whichever has the least difference the test sample is classified as them.

Perceptron trained using Backpropagation is a supervised method common in Artificial Neural Networks. In the multilayer structured network, backpropagation is for gradient based optimization of weights of each node. By the result of this optimization least error in output should be achieved. This error function here is an equation describing the difference between expected output and the output and the optimization schmee is based on its gradient with respect to each node's weight.

In Percepton trained using MSE classifier is a single layer binary classifier with linear decision boundaries. The mean square error based on the maximum likelihood principle maps training data to target values randomly and creates a corresponding weight matrix.

The dataset used in the experiment are firstly a training set of 60000 and a test set of 10000 examples of handwritten digits. They are all the same size and centered. Secondly, a set of face images from an official website. All images are in the frontal position of the face but there is some margin for movement.

The experiment is conducted as follows. For the algorithms sklearn libraries are leveraged. For the dataset of handwritten digits (MNIST) raw grayscale pixel intensities are used as features. Each classification is then modeled, trained and evaluated. For the MNIST dataset there is one train and test set (60k/10k) and another one train/test set after applying PCA to the dataset. PCA is principal component analysis for relevant feature selection for the task and helps remove redundancy and complexity in the process. For the ORL dataset which contains human face images, the first train and test set split is performed randomly with 70% and 30% ratios. Second train/test set is again obtained after applying PCA. In Figure 1 the dataset with PCA applied is represented visually also with the dataset that no PCA is applied..

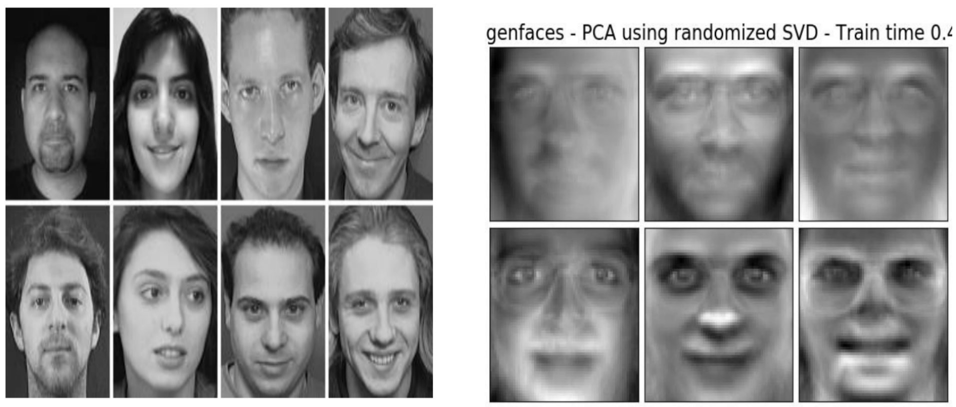

Figure 1: Dataset with no PCA applied and PCA applied

The first method Nearest Class Centroid (NCC) is applied to ORL dataset (face image dataset) with nearest neighbor number 10 and Shrinkage is none or 0.2. The shrinked algorithm(Shrinkage with 0.2) has benefits over other one which are; reducing noise and making classification more accurate and also automatic gene selection. For Nearest Sub-class Centroid Classifier data seems overfit as shown in Figure 2.

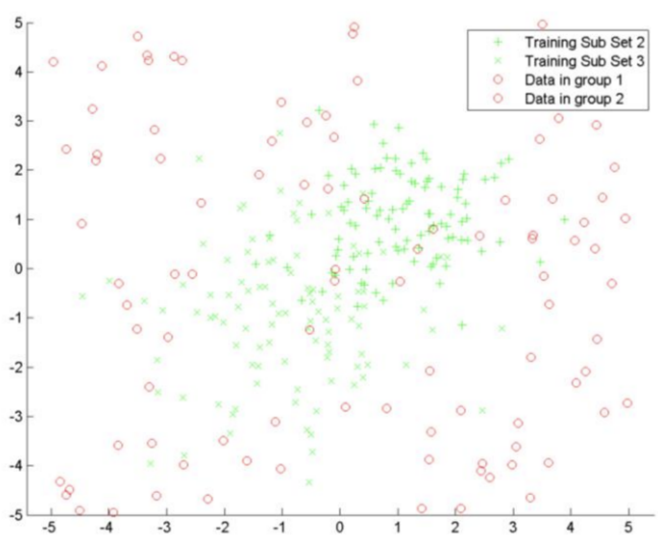

Figure 2: Overfit data with NSC applied

For KNN algorithms many k values and their accuracy in training and other metrics such as precision, recall, f1 score and support are printed which is shown in Figure 3. For each k KNeighborsClassifier is trained over a loop, since knn is a fast algorithm and a large number of trials are conducted.

k=29, accuracy=97.04%

k=1 achieved highest accuracy of 99.26% on validation data

EVALUATION ON TESTING DATA

	precision	recall	f1-score	support
0	1.00	1.00	1.00	43
1	0.95	1.00	0.97	37
2	1.00	1.00	1.00	38
3	0.98	0.98	0.98	46
4	0.98	0.98	0.98	55
5	0.98	1.00	0.99	59
6	1.00	1.00	1.00	45
7	1.00	0.98	0.99	41
8	0.97	0.95	0.96	38
9	0.96	0.94	0.95	48
avg / total	0.98	0.98	0.98	450

Figure 3: k values and respective accuracies

The result also indicates that higher k values does not mean better performance.

For evaluation 5 digits are given to the best KNN model in which the algorithm outputs each correctly.

While testing Perceptron trained using Backpropagation, the learning rate is either increased or decreased based on the validation error. This hints convergence to where early stopping is achieved. In the next phase weight optimizations are made within the network according to local minima from earlier step. The performance of this algorithm is summarized in Figure 4.

```
Train twolayer perceptron with 700 hidden units.
Learning rate: 1.000000e-01.
Validation:
Classification errors: 788
Correctly classified: 9212
applyStochasticSquaredErrorTwoLayerPerceptronMNIST
Train twolayer perceptron with 700 hidden units.
Learning rate: 1.000000e-01.
Validation:
Classification errors: 757
Correctly classified: 9243
```

Figure 4: Performance of Perceptron with Backpropogation

The model has a success rate of 92% approximately for both train-test pairs.

The last model is Perceptron trained using MSE which is a feed forward algorithm. Since it is a binary classifier, each perceptron has binary values according to an activation function. During the iterative training perceptron calculates an output for each input value in the training set . If the output 1 where it should have been 0 the threshold value is increased and all perceptrons associated with the outcome are decreased by 1. If the opposite case happens, the threshold is decreased and weights increase. If the outcome is correct no changes applied. The dataset is divided randomly and Levenberg-Marquardt as training type is applied. The MSE applied as a metric which indicates an error of 0.14% which seems pretty accurate.

Comparing these classifiers one can conclude that KNN works better than NCC and NSC. The result is intuitive in the sense that KNN algorithms basis lies in approximating instances to small neighbors of similar objects.However Back Propagation outperforms KNN especially in the cases with more complex datasets. However the model training significantly takes more time.

From this paper I have learned about NSC, NCC and Peceptron with MSE algorithms along with PCA. In addition to how these methods are implemented within such experiments and the evaluation metric for each. I personally liked the algorithms being compared involved simple classification techniques and more complex one with Artificial Neural Network appliances in the task of object recognition. However, I believe there are some not negligible shortcomings. For example there is no explicit conclusion on which model works better since each evaluation metric for each model is different; it is a bit ambiguous to infer the best model among the presented. In addition there are no separate results given for datasets with PCA applied and not applied hence the reader does not know which outputed better results.

Link:
https://www.researchgate.net/profile/Tanmoy-Das-12/publication/330798350_Machine_Learning_algorithms_for_Image_Classification_of_hand_digits_and_face_recognition_dataset/links/5c54619f92851c22a3a12b5f/Machine-Learning-algorithms-for-Image-Classification-of-hand-digits-and-face-recognition-dataset.pdf