Manifold Learning for Image Processing

Pankaj Kumar

Moscow Institute of Physics and Technology

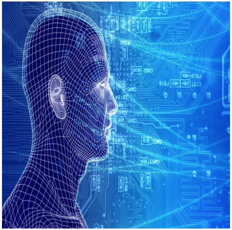
kumar.x.pankaj@gmail.com

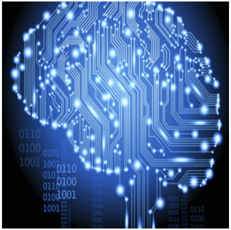
June 14, 2017

Overview

Introduction

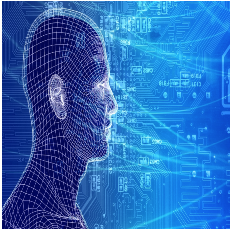
- Data is everywhere
- Data is being recorded at an unprecedented rate.
- How do we find out the useful information?

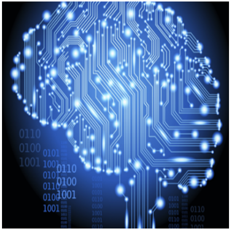




Introduction

- Data is everywhere
- Data is being recorded at an unprecedented rate.
- How do we find out the useful information?





- By effective modeling of high-dimensional observed data, such that it capture the structure of the information content in a concise manner.

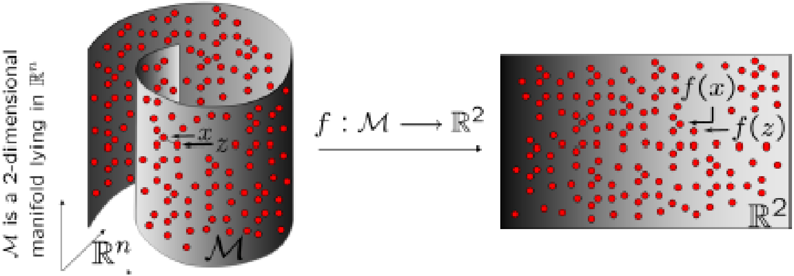
Image Processing

- Digital revolution have infused enormous amount of images and video, which is growing constantly.
- Each image sequence is a point in a space of dimension equal to the number of image pixels.
- So the observation space is of possibly thousands/millions of dimensions.
- The data is embedded as sub-manifold in high dimensional space.
- Learning this manifold is a natural approach to the problem of manifold learning algorithms.

Manifold Learning

Manifold learning is form of unsupervised machine learning algorithms, which extract low-dimensional structure from high dimensional data.

It can be further illustrated in below figure, where manifold learning algorithms for non-linear data builds an embedding function f mapping $\mathcal{M}$ to $\mathbb{R}^2$.



Problem

Given a set of of high-dimensional training instances $\mathbb{X} = \{x_1, x_2, ..., x_N\}$, where $x_i \in \mathbb{R}^D$. We assume that $\mathbb{X}$ approximately lie on a smooth manifold $\mathcal{X}$. The idea of manifold learning algorithms is to find an embedding set $\mathbb{Y} = \{y_1, y_2, ..., y_N\}$ of $\mathbb{X}$ in low dimension space $\mathbb{R}^d$, where $d < D$. The local manifold structure formed by $\mathbb{X}$ in the original space $\mathbb{R}^D$ is preserved in the embedded space $\mathbb{R}^d$.

Linear Methods

- Principal Components Analysis (PCA)
- Metric Multidimensional Scaling (MDS)

Principal Components Analysis (PCA)

Given $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^D$
Find $y_1, \ldots, y_n \in \mathbb{R}$ such that

$$y_i = \mathbf{w} \cdot \mathbf{x}_i$$

and

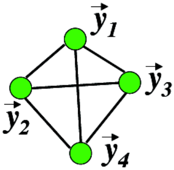
$$\max_{\mathbf{w}} \text{Variance}(\{y_i\}) = \sum_i y_i^2 = \mathbf{w}^T \left(\sum_i \mathbf{x}_i \mathbf{x}_i^T \right) \mathbf{w}$$

$$\mathbf{w}_* = \text{leading eigenvector of } \sum_i x_i x_i^T$$

Metric Multidimensional Scaling (MDS)

- An alternative approach to PCA based on preserving pairwise distances

$$\begin{bmatrix} 0 & \Delta_{12} & \Delta_{13} & \Delta_{14} \\ \Delta_{12} & 0 & \Delta_{23} & \Delta_{24} \\ \Delta_{13} & \Delta_{23} & 0 & \Delta_{34} \\ \Delta_{14} & \Delta_{24} & \Delta_{34} & 0 \end{bmatrix}$$



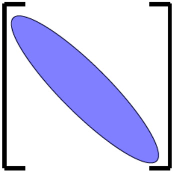
Given $n(n-1)/2$ pairwise distances $d_{ij} = \|X_i - X_j\|$, find a low-dimensional embedding $X \to y$ such that $\|y_i - y_j\| \approx d_{ij}$.

Non-Linear Methods

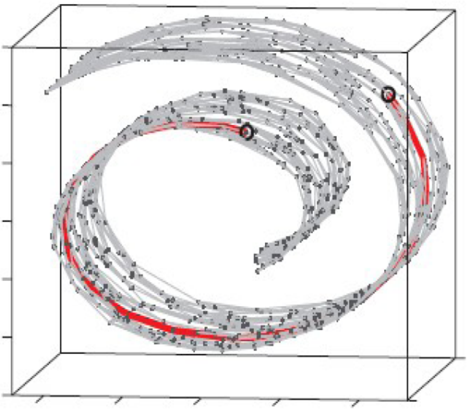
- Isomap
- Locally Linear Embedding
- Laplacian Eigenmaps
- Diffusion Maps
- Hessian eigenmaps
- Local Tangent Space Alignment

Isomap

1. Build a sparse graph with K-nearest neighbors

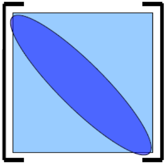
$D_g =$ 

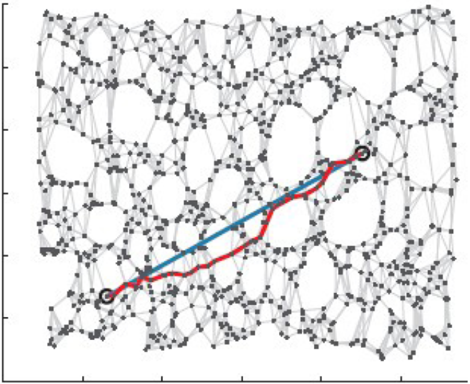
(distance matrix is sparse)



Isomap

2. Infer other interpoint distances by finding shortest paths on the graph (Dijkstra's algorithm).

$D_g =$ 



Isomap

3. Build a low-D embedded space to best preserve the complete distance matrix.

Error function:

inner product distances in graph

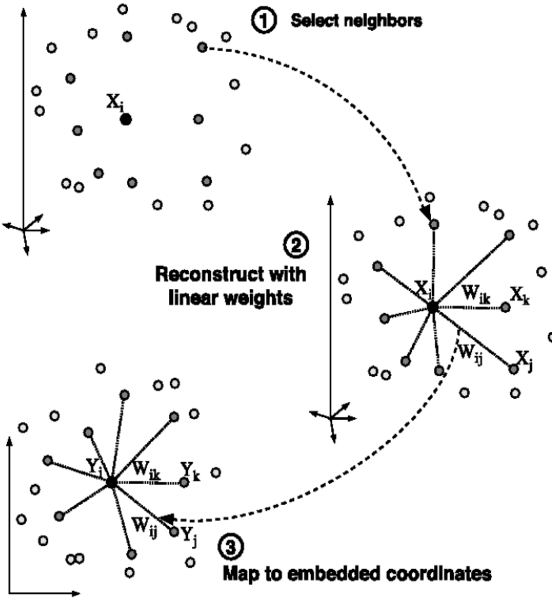
inner product distances in new coordinate system

L2 norm

$$E = \|\tau(D_G) - \tau(D_Y)\|_{L^2}$$

Solution – set points Y to top eigenvectors of D_g

Locally Linear Embedding



Contribution

1. The research exploring the comparison of manifold learning algorithms is quite shallow. There are few theoretical comparisons results for these algorithms, but the primary evaluation methodology has been to run the algorithm on artificial data sets and do comparison. We use MNIST datasets with optical character recognition in mind to evaluate the performance of the manifold learning algorithms.
2. Motivated by recent study, we presents a novel manifold learning approach for high dimensional data, with emphasis on the problem of anomaly detection in image.

Introduction

Introduction

- optical character recognition have found applications in bank check processing, assistive technology for visually impaired users, automatic number plate recognition and many more
- A basic requirement for credibility of classification algorithms requires a high accuracy on MNIST datasets, a hello world datasets.
- The earlier experimental results doing comparison between different manifold learning algorithms show that these methods can work pretty well at least in our toy examples.

Data

- The standardised MNIST database has a training set of 60,000 examples, and a test set of 10,000 examples.
- Each example is of size 28×28, a total of 784 greyscale pixels.
- Training set have 60,000 rows (images) and 785 (784 pixels and a one label) columns.
- Each row component contains a value between one and zero and this describes the intensity of each pixel.
- Test set is same as training set, except no label and 10000 images.

First look

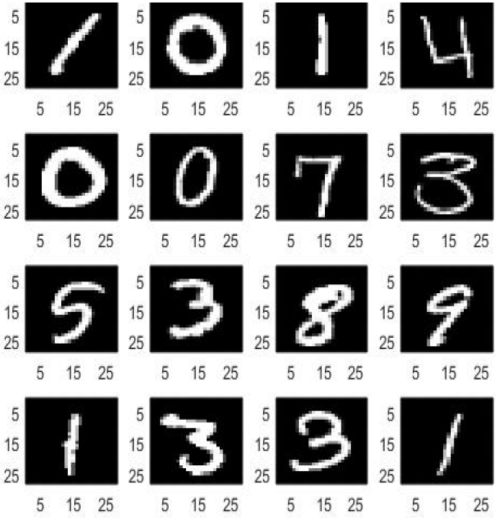


Figure: First look at MNIST data

Preprocessing: Hough transform

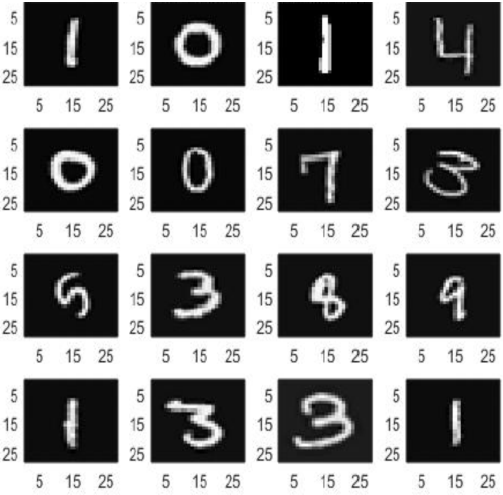


Figure: Hough transform

Pearson Correlation plot

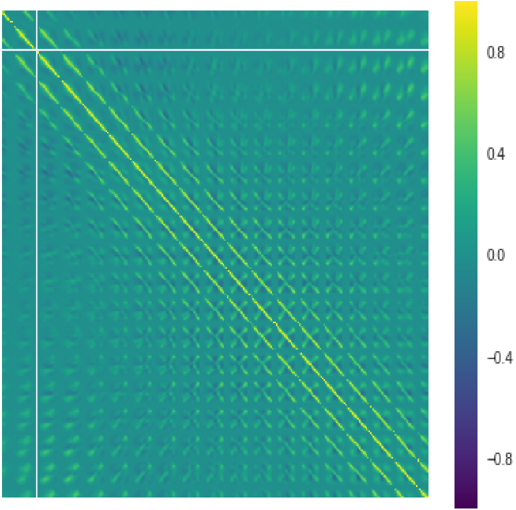


Figure: Pearson Correlation plot 500 observations

Explained Variance

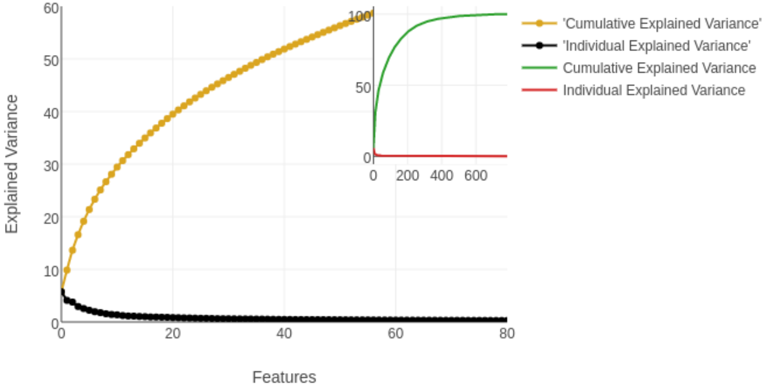


Figure: Explained Variance

Eigenvalues

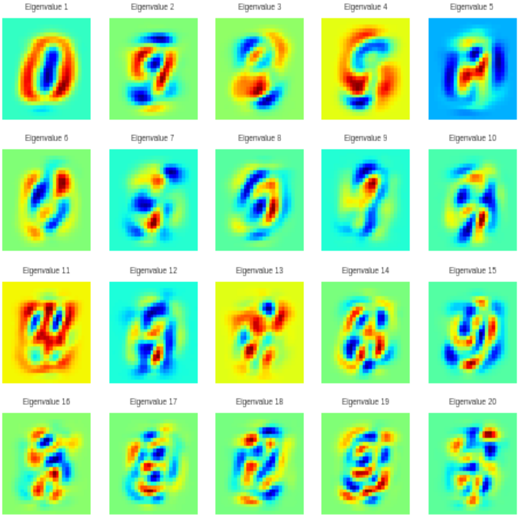


Figure: Top 200 eigenvalues

Digit Classification

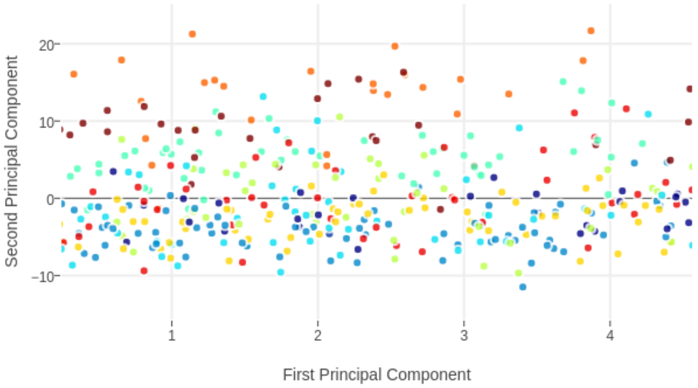


Figure: PCA

TSNE

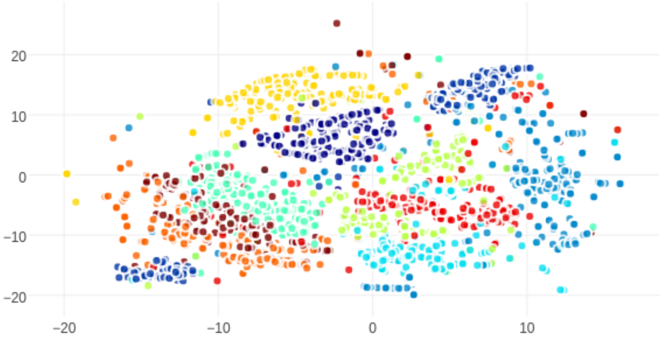


Figure: TSNE (t-Distributed Stochastic Neighbour Embedding)

Comparison: Manifold learning Algorithms

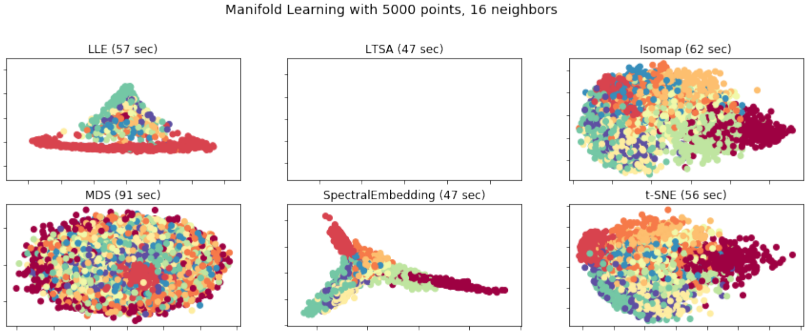
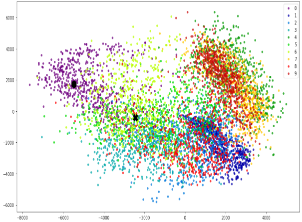
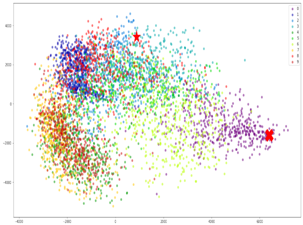


Figure: Manifold Learning Algorithms Output

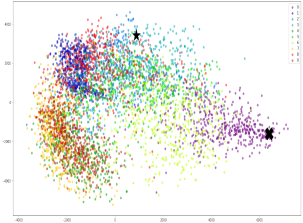
Out of sample



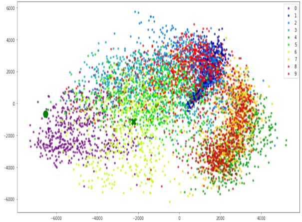
(a) LLE



(b) ISOMAP



(c) SE



(d) TSNE

Figure: Out of sample predication for manifold learning algorithms

Models

- To perform digit recognition task we used three supervised algorithms: K-Nearest Neighbour (KNN) , Support Vector Machine (SVM) and Convolutional Neural Networks (CNN).
- By extracting n-components from the low dimension, we build train and text datasets corresponding to manifold learning algorithms. For example, $\mathbf{TEST}_{Isomap}$ and $\mathbf{Train}_{Isomap}$.
- For dimension reduction, we take three manifold algorithms namely Isomap, Locally Linear Embedding and t-distributed Stochastic Neighbor Embedding.

Training

- The 5-fold cross validation is performed on the training set comprising of original data and data created out of dimension reduction.
- Then for every model, we choose the parameters yielding highest cross-validation classification accuracy.
- Finally, to obtain a real out of sample performance evaluation, we use the best model to predict on the test set.

Accuracy: Original

Table: Model Summary for **Data**(original)

Model	Data	Accuracy
KNN	original	91.00
SVM	orignal	95.60
CNN	orignal	93.00

Accuracy: Isomap

Table: Model Summary for **Data**(isomap)

Model	Data	Accuracy (%)
KNN	isomap	92.00
SVM	isomap	95.50
CNN	isomap	93.00

Accuracy: TSNE

Table: Model Summary for **Data**(tsne)

Model	Data	Accuracy
KNN	tsne	92.00
SVM	tsne	96.00
CNN	tsne	93.00

Accuracy: LLE

Table: Model Summary for **Data**(lle)

Model	Data	Accuracy
KNN	lle	90.00
SVM	lle	91.20
CNN	lle	91.20

Takeaway

- The out of sample predication for manifold algorithms indicates that test points are projected on the corrects clusters of training datasets. But the projection of the test points is not so clear.
- The prepossessing using Hough transform didn't improve the accuracy.
- There was no substantial improvement in the accuracy of hand written digit recognition with preprocessed data using manifold learning algorithms.

Anomaly Detection

Anomaly detection refers to the problem of finding patterns in data that do not conform to expected behavior. There are numerous approach:

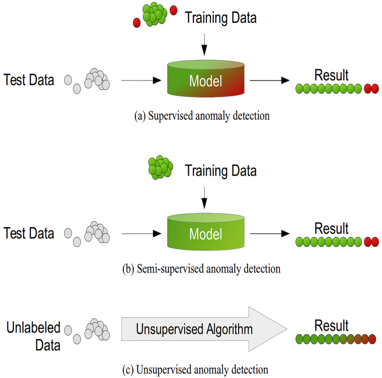


Figure: Anomaly detection approach based on data labels.

Data

(a)

(b)

(c)

(d)

(e)

(f)

Figure: Input Image

Multiscale Algorithms

1. Input Image: $\mathbf{I}$. Size: 400×400.
2. For $l = 0 : L$ compute Gaussian pyramid $\{\mathbf{G}_i\}_{l=0}^{L}$, where $\mathbf{G}_0$ is the original image and $\mathbf{G}_L$ is the coarsest resolution as discussed in section **??**.
3. Starting with $\mathbf{G}_L$, sample a subset $\mathbf{I}_s$ from $\mathbf{I}$.
4. Compute Diffusion map using $\mathbf{I}_s$.
5. Extend above step to remaining pixels.
6. Calculate anomaly score $\mathbf{C}_l$ for 400 pixels.
7. Set threshold τ_l to 95th percentile of the anomaly score.
8. If $\mathbf{C}_l > \tau_l$, the anomaly will be sampled more densely. Goto step 2, else
9. output.

Flowchart

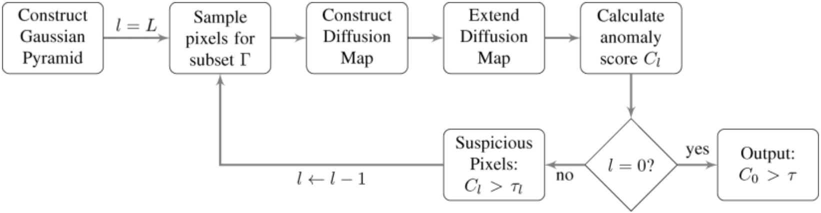


Figure: Multiscale Algorithm.

Results: True Positive (Basic)

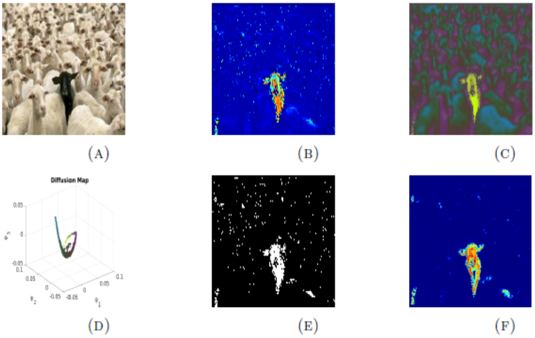


FIGURE 5.4: Anomaly Detection results for multiscale detector

TABLE 5.1: Parameters used in Multiscle algorithms for figure 5.3d (Black Sheep)

L	Image	Patch	Window	Mask	Pixels (%)	d
0	200×200	32×32	16×16	5×5	0.10	6
1	100×100	16×16	8×8	3×3	0.20	3
2	50×50	8×8	4×4	3×3	0.30	3

Results: True Positive (Complex)

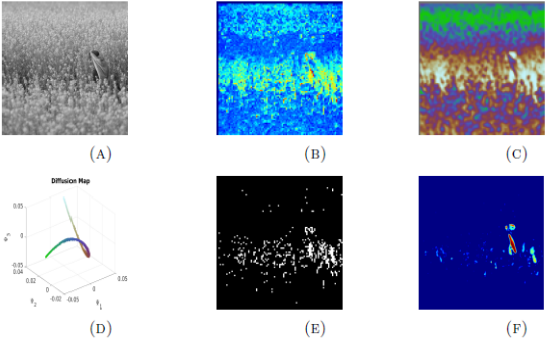


FIGURE 5.5: Anomaly Detection results for multiscale detector

TABLE 5.2: Parameters used in Multiscle algorithms for figure 5.3c (Village Girl)

L	Image	Patch	Window	Mask	Pixels (%)	d
0	200×200	8×8	40×40	6×6	0.10	6
1	100×100	4×4	20×20	3×3	0.30	3
2	50×50	2×2	10×10	2×2	0.50	3

Results: False Alarm (Complex)

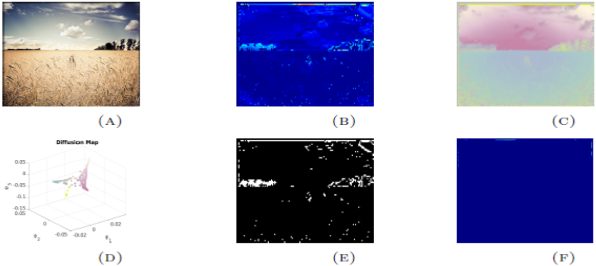


FIGURE 5.6: Anomaly Detection results for multiscale detector

TABLE 5.3: Parameters used in Multiscle algorithms for figure 5.3e (girl in the field)

L	Image	Patch	Window	Mask	Pixels (%)	d
0	400×400	64×64	32×32	8×8	0.10	12
1	200×200	32×32	16×16	4×4	0.20	6
2	100×100	16×16	8×8	2×2	0.40	3
3	50×50	8×8	4×4	2×2	0.50	3

Conclusions

1. Deviating from primary evaluation methodology of manifold leaning algorithms based on artificial data sets, we use MNIST datasets with optical character recognition in mind.
2. The out of sample prediction for manifold algorithms indicates that test points are projected on the corrects clusters of training datasets, but separation of some numbers were blurred.
3. The preprocessing of MNIST data by manifold learning algorithms didn't improve accuracy in optical character recognition.
4. The proposed algorithms performs well the images which is not too contrast, but gives false positive on image which is of high resolutions.

Future Work

1. When answering the questions on evaluating the results of manifold learning, we need to take in account of manifold structure and noises.
2. When there is multiple sub-manifolds of possibly different dimensionalities embedded in high dimensional complex data, it is most unlikely that the existing manifold learning approaches can discover all the interesting lower-dimensional structures. There is need to develop a hierarchical manifolds learning framework to discover a variety of the underlying low dimensional structures.
3. In optical character recognition, there is need to optimize the model when calculating accuracy.
4. A future extension of this work would be combine anomaly scores from the different multiscale levels into a single anomaly score. This will enable detecting anomalies of different size in an image, which this algorithms fails to address.

Questions?