ED5340 - Lab 13

ABIL MOHAMMED K
ED19B041

1. A large number of machine learning datasets involve thousands and sometimes millions of features. These features make training very slow. Also, there is plenty of space in high dimensional datasets making the high-dimensional datasets very sparse.

This increases the risk of overfitting since the predictions will be based on much larger extrapolations as compared to those on dimensional data.

Manifolds can act as stepping stone from complex space to a simplex smoother subset.

. Classification problems are prime examples for manifold learning.

Or 2D shape made ~~into~~ to fit to higher-dimensional space by twisting or bending it.

2. Locally Linear Embedding (LLE) is a Manifold learning technique that is used for non-linear dimensionality reduction. It is an unsupervised learning algorithm that produces low-dimensional embeddings of high-dimensional inputs relating each training instance to its closest neighbour.

For each training instance $x^{(i)}$, algo finds its k nearest neighbours and express $x^{(i)}$ as linear function of them

eg: $\sum_{i=1}^{m} \left(x^{(i)} - \sum_{j=1}^{m} w_{i,j} x^{(j)} \right)^2$

And $\sum_{j=1}^{m} w_{i,j} = 1$

Finally each high-dimensional training instance $x^{(i)}$ is mapped to low-dimensional vector $y^{(i)}$ while preserving neighbourhood relationships. This is done by choosing d-dimensional coordinates

$$\sum_{i=1}^{m} \left(y^{(i)} - \sum_{j=1}^{m} w_{i,j} y^{(j)} \right)^2$$

Advantages:

i) Ability to deal with large amounts of high dimensional data
ii) Non-iterative way of finding the embeddings

Disadvantages

i) Sensitivity to noise
ii) Inevitable ill-conditioned eigen problems.
iii) Inability to deal with novel data.