Lecture 11: Effective Dimension of (Visual) Data

Nikola Zlatanov

Innopolis University

Advanced Statistics

17-th of April to 24-th of April, 2023

Motivation

- The visual data received by the human eye is high dimensional
- However, we are low-dimensional creatures
- Then how does our mind comprehend the high-dimensional visual data around us?
- Is it possible that the high-dimensional visual data around us is simply a disguised low-dimensional data, and our mind can somehow remove this disguise and 'see' the low dimensional data
- For example, imagine a picture of someone that you know.
- Imagine you are presented a very low-dimensional picture of that person such that we are able to recognize the person.
- Then this means that if we are presented with the same picture but in very high-dimensions, you won't gain any new information in terms of recognizing who this person is.
- It seems that, at least for face recognition, our mind sees some low-dimensional structure in high-dimensional visual data.

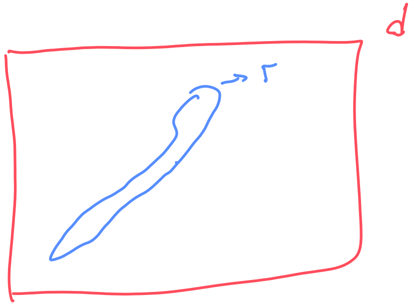
Low-Dimensional Paradigm

- For the phenomenon described in the previous slide the following explanation is suggested.
- Low-Dimensional Paradigm: High-dimensional (visual) data has some low-dimensional structure and our mind is able to see the low-dimensional structure.
- For example, it is known that our mind recognizes human faces using 5-10 features, independent of how many pixels (i.e., dimensions) the picture has.
- Now how do we translate this Low-Dimensional Paradigm to computers?
- How can we make algorithms that make computers see the low-dimensional structure in high-dimensional (visual) data?

Low-Dimensional Manifold in High-Dimensional Data

- The Low-Dimensional Paradigm is possible if the high dimensional data 'lives' in a low-dimensional manifold
- Example:

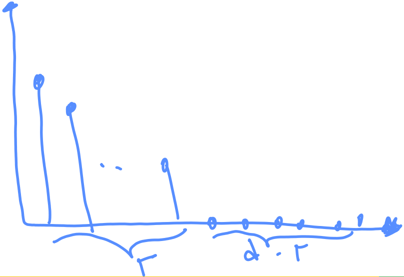
$r \ll d$



- Data $\in \mathbb{R}^d$, but manifold $\in \mathbb{R}^r$, where $r \ll d$.

Theoretical Test For Low-Dimensional Manifold

- How can we test if our high-dimensional data has low-dimensional manifold?
- Theoretically, if we know the exact covariance matrix, $\mathbf{\Sigma}$, of our d-dimensional data $\boldsymbol{X}_1, \boldsymbol{X}_2, ...\boldsymbol{X}_N \in \mathbb{R}^d$, then our data has a low-dimensional manifold if the covariance matrix, $\mathbf{\Sigma}$ has only r non-zero eigenvalues, where $r \ll d$.
- If that is the case, then the (sorted) eigenvalues of $\mathbf{\Sigma}$ would be like this:



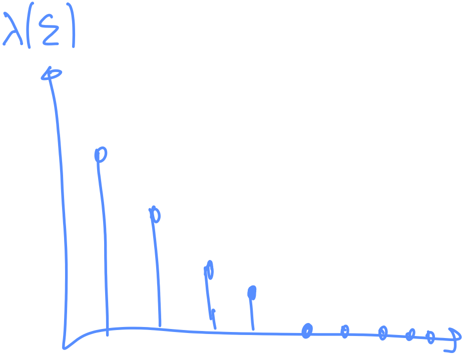
Practical Test For Low-Dimensional Manifold

- The problem is that in practice we do NOT know the exact covariance matrix, $\mathbf{\Sigma}$, of our d-dimensional data $\boldsymbol{X}_1, \boldsymbol{X}_2, ...\boldsymbol{X}_N \in \mathbb{R}^d$.
- What we know (or can find out), by exploiting the d-dimensional data $\boldsymbol{X}_1, \boldsymbol{X}_2, ...\boldsymbol{X}_N \in \mathbb{R}^d$, is only the estimated covariance matrix obtained as

$$\hat{\mathbf{\Sigma}}_N = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{X}\boldsymbol{X}^T \tag{1}$$

- Can we see the low-dimensional manifold, present in $\mathbf{\Sigma}$, from $\hat{\mathbf{\Sigma}}_N$?

- If we now plot the eigenvalues of $\hat{\mathbf{\Sigma}}_N$, we would guess that some low-dimensional manifold exists, if we have the following plot:



An Algorithm For Obtaining A Low-Dimensional Manifold

The algorithm:

- Let us have d-dimensional data $\boldsymbol{X}_1, \boldsymbol{X}_2, ...\boldsymbol{X}_N \in \mathbb{R}^d$
- Let us find the estimated covariance matrix, $\hat{\boldsymbol{\Sigma}}_N$, of this data using (1).
- Let us find the eigenvalues of $\hat{\boldsymbol{\Sigma}}_N$, and let they be given by
$$\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \hat{\lambda}_3 \geq ... \geq \hat{\lambda}_d$$
- Then, let us set the dimension of the manifold, r, as the solution to the following optimization problem
$$\begin{aligned} &\min \quad r \\ &\text{s.t.} : \sum_{i=1}^{r} \hat{\lambda}_i \geq \beta \sum_{i=r+1}^{d} \hat{\lambda}_i, \end{aligned} \tag{2}$$
where $\beta > 1$ (continues on the following slide...)

An Algorithm For Obtaining A Low-Dimensional Manifold

- If $r \ll d$, i.e., r is order(s) of magnitudes smaller than d, then we can claim that the d-dimensional data has a low-dimensional approximate representation
- Otherwise, if $r \approx d$, i.e., r is of the same order as d, then we can claim that the d-dimensional data does NOT have a low-dimensional approximate representation.
- If $r \ll d$ holds, then the r eigenvectors $\hat{\boldsymbol{v}}_1$, $\hat{\boldsymbol{v}}_2$, ..., $\hat{\boldsymbol{v}}_r$ of $\hat{\boldsymbol{\Sigma}}_N$ associated with its eigenvalues $\hat{\lambda}_1$, $\hat{\lambda}_2$, ... ,$\hat{\lambda}_r$, respectively, form the r-dimensional (i.e., lower-dimensional) projection basis of the d-dimensional data.
- Thereby, to obtain a r-dimensional representation of $\boldsymbol{X}_1, \boldsymbol{X}_2, ...\boldsymbol{X}_N \in \mathbb{R}^d$, we need to project $\boldsymbol{X}_1, \boldsymbol{X}_2, ...\boldsymbol{X}_N \in \mathbb{R}^d$ onto the basis formed by $\hat{\boldsymbol{v}}_1$, $\hat{\boldsymbol{v}}_2$, ..., $\hat{\boldsymbol{v}}_r$.
- This ends the algorithm.

Accuracy of the Proposed Algorithm

- We now want to test how accurate our proposed algorithm is.
- Let the true $d \times d$ covariance matrix $\mathbf{\Sigma}$ have a r^*-dimensional manifold.
- Then, $\mathbf{\Sigma}$ has r^* non-zero eigenvalues given by λ_1, λ_2, ..., λ_{r^*}.
- Let the $d \times d$ estimated covariance matrix $\hat{\mathbf{\Sigma}}_N$ have a d non-zero eigenvalues given by $\hat{\lambda}_1$, $\hat{\lambda}_2$, ..., $\hat{\lambda}_d$ with corresponding eigenvectors, denoted by $\hat{\boldsymbol{v}}_1$, $\hat{\boldsymbol{v}}_2$, ..., $\hat{\boldsymbol{v}}_d$.
- Let us take only the r largest eigenvalues of $\hat{\mathbf{\Sigma}}_N$, where r is found from (2), and their corresponding eigenvectors, and use them to construct a new estimated covariance matrix, denoted by $\hat{\mathbf{\Lambda}}_r$, obtained as

$$\hat{\mathbf{\Lambda}}_r = \sum_{i=1}^{r} \hat{\lambda}_i \hat{\boldsymbol{v}}_i \hat{\boldsymbol{v}}_i^T$$

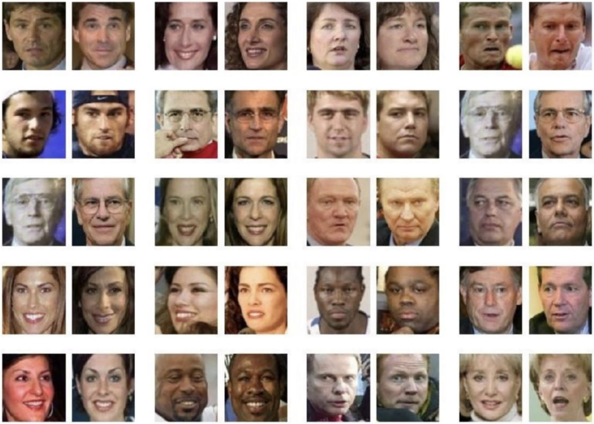
- Then, what is the distance between $\mathbf{\Sigma}$ and $\hat{\mathbf{\Lambda}}_r$, i.e., $||\mathbf{\Sigma} - \hat{\mathbf{\Lambda}}_r||_{\text{op}}$?

Proposal

- If you can upper bound the distance $||\mathbf{\Sigma} - \hat{\mathbf{\Lambda}}_r||_{\text{op}}$, as a function of r, probabilistically or deterministically, **I will give 30 pts towards your final mark!**

Eigenfaces

- How do we know that visual data has a low-dimensional manifold, and thereby for visual data $||\mathbf{\Sigma} - \hat{\mathbf{\Lambda}}_r||_{\text{op}}$ is indeed well bounded?
- Algorithm: Face recognition using Eigenfaces
  1. Let us have training data of face images, such as



Test ?

Eigenfaces

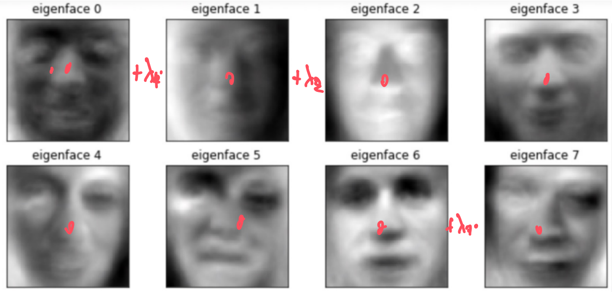
Algorithm: Face recognition using Eigenfaces

- Make color images into black and white images
- Transform the images into vectors
- Find the estimated covariance matrix, $\hat{\boldsymbol{\Sigma}}_N$, from the vectors
- Keep only the first r eigenvalues of $\hat{\boldsymbol{\Sigma}}_N$ and extract their corresponding eigenvectors $\hat{\boldsymbol{v}}_1$, $\hat{\boldsymbol{v}}_2$, ..., $\hat{\boldsymbol{v}}_r$
- Convert these r eigenvalues into images (you should obtain something like the following)

Nitro

$[\lambda_0 \dots \lambda_7]$

$[2\ 1 \dots 2]$

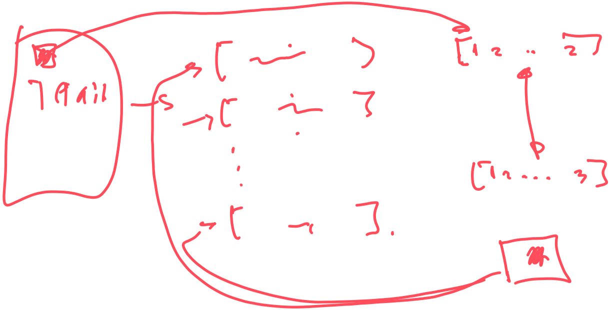


$\lambda_0 \cdot$ $+\lambda_1\cdot$ $+\lambda_2$ $+\lambda_7\cdot$

Eigenfaces

Algorithm: Face recognition using Eigenfaces

1. Obtain the test data of face images
2. Use the r eigenvectors, extracted from the training data, to recognize the images in the test data. How?



It turns out that this algorithms works well for face recognition, which leads us to the conclusion that, at least, face images have low-complexity manifold.