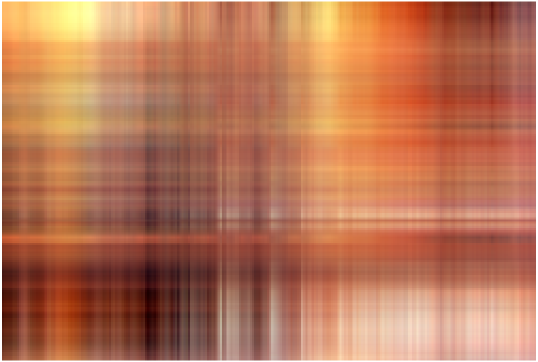
Principal component analysis

Ingo Blechschmidt

December 17th, 2014

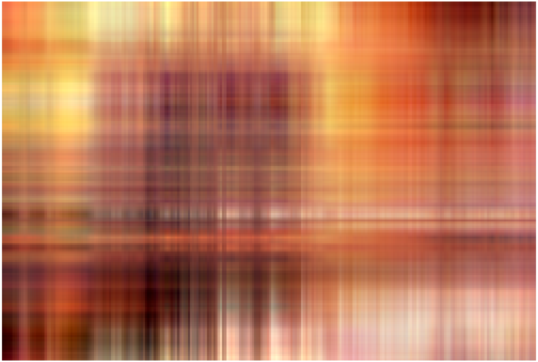
Principal component analysis



Ingo Blechschmidt

December 17th, 2014

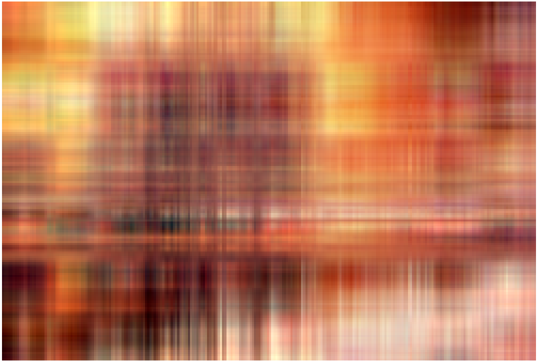
Principal component analysis



Ingo Blechschmidt

December 17th, 2014

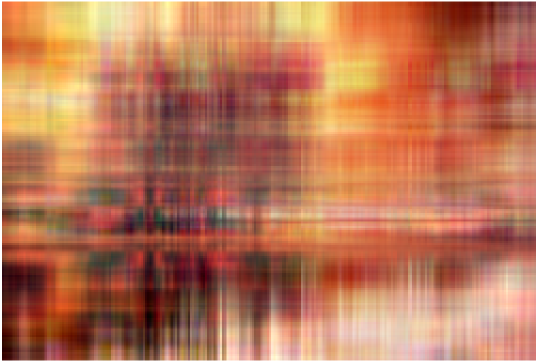
Principal component analysis



Ingo Blechschmidt

December 17th, 2014

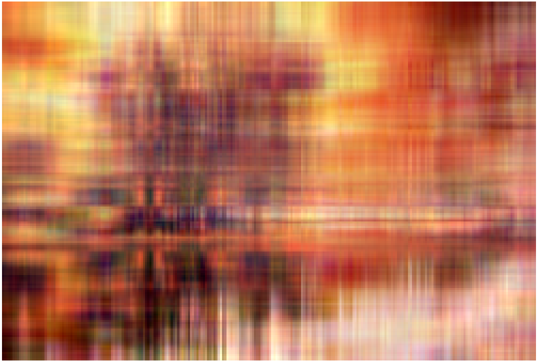
Principal component analysis



Ingo Blechschmidt

December 17th, 2014

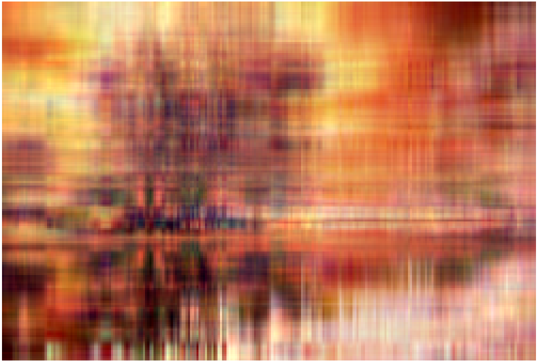
Principal component analysis



Ingo Blechschmidt

December 17th, 2014

Principal component analysis



Ingo Blechschmidt

December 17th, 2014

Principal component analysis

Ingo Blechschmidt

December 17th, 2014

Principal component analysis

Ingo Blechschmidt

December 17th, 2014

Principal component analysis

Ingo Blechschmidt

December 17th, 2014

Principal component analysis

Ingo Blechschmidt

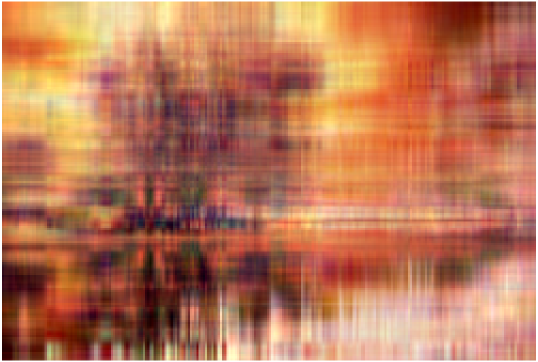
December 17th, 2014

Principal component analysis

Ingo Blechschmidt

December 17th, 2014

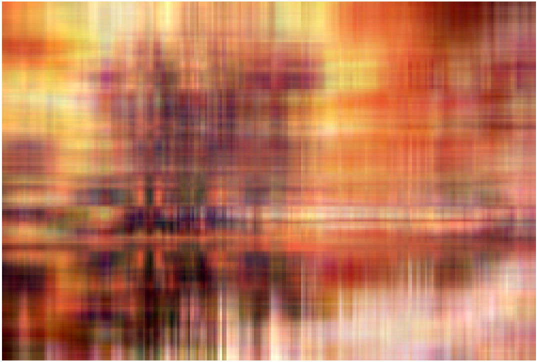
Principal component analysis



Ingo Blechschmidt

December 17th, 2014

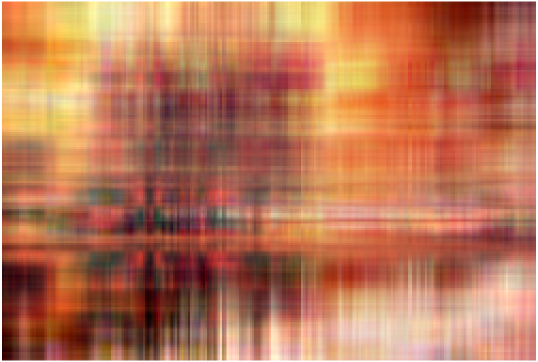
Principal component analysis



Ingo Blechschmidt

December 17th, 2014

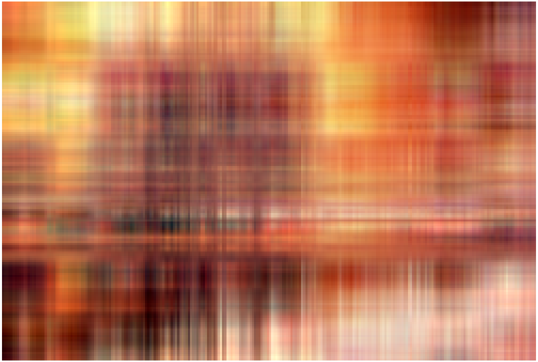
Principal component analysis



Ingo Blechschmidt

December 17th, 2014

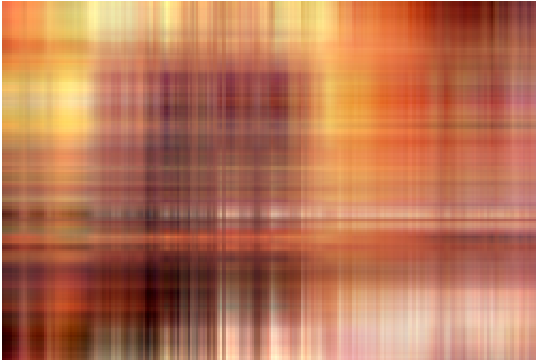
Principal component analysis



Ingo Blechschmidt

December 17th, 2014

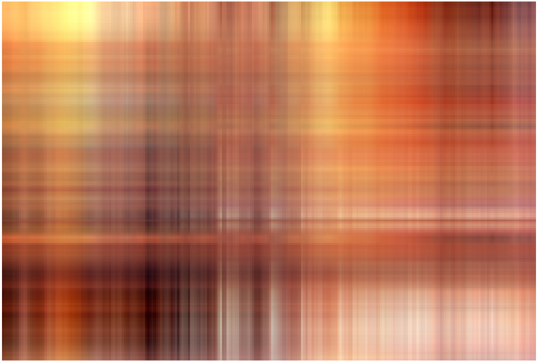
Principal component analysis



Ingo Blechschmidt

December 17th, 2014

Principal component analysis



Ingo Blechschmidt

December 17th, 2014

Principal component analysis

Ingo Blechschmidt

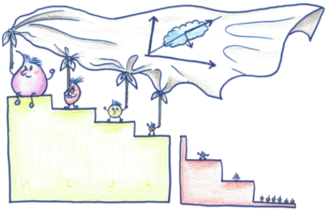
December 17th, 2014

Principal component analysis

Ingo Blechschmidt

December 17th, 2014

Outline



1. Theory
   - Singular value decomposition
   - Pseudoinverses
   - Low-rank approximation
2. Applications
   - Image compression
   - Proper orthogonal decomposition
   - Principal component analysis
   - Eigenfaces
   - Digit recognition

Singular value decomposition

Let $A \in \mathbb{R}^{n \times m}$. Then there exist

- numbers $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_m \geq 0$,
- an orthonormal basis $\mathbf{v}_1, \ldots, \mathbf{v}_m$ of $\mathbb{R}^m$, and
- an orthonormal basis $\mathbf{w}_1, \ldots, \mathbf{w}_n$ of $\mathbb{R}^n$,

such that

$$A\mathbf{v}_i = \sigma_i \mathbf{w}_i, \quad i = 1, \ldots, m.$$

In matrix language:

$$\begin{aligned} A &= W \Sigma V^t, \\ \text{where} \quad V &= (\mathbf{v}_1 | \ldots | \mathbf{v}_m) \in \mathbb{R}^{m \times m} \text{ orthogonal}, \\ W &= (\mathbf{w}_1 | \ldots | \mathbf{w}_n) \in \mathbb{R}^{n \times n} \text{ orthogonal}, \\ \Sigma &= \operatorname{diag}(\sigma_1, \ldots, \sigma_m) \in \mathbb{R}^{n \times m}. \end{aligned}$$

The pseudoinverse of a matrix

Let $A \in \mathbb{R}^{n\times m}$ and $\mathbf{b} \in \mathbb{R}^n$. Then the solutions to the optimization problem

$$\|A\mathbf{x} - \mathbf{b}\|_2 \longrightarrow \min$$

under $\mathbf{x} \in \mathbb{R}^m$ are given by

$$\mathbf{x} = A^+\mathbf{b} + V\begin{pmatrix}0\\ \star\end{pmatrix},$$

where $A = W\Sigma V^t$ is the SVD and

$$A^+ = W\Sigma^+ V^t,$$
$$\Sigma^+ = \mathrm{diag}(\sigma_1^{-1}, \dots, \sigma_m^{-1}).$$

Low-rank approximation

Let $A = W\Sigma V^t \in \mathbb{R}^{n\times m}$ and $1 \leq r \leq n, m$. Then a solution to the optimization problem

$$\|A - M\|_{\text{Frobenius}} \longrightarrow \min$$

under all matrices M with $\operatorname{rank} M \leq r$ is given by

$$\begin{aligned} M &= W\Sigma_r V^t, \\ \text{where} \quad \Sigma_r &= \operatorname{diag}(\sigma_1, \dots, \sigma_r, 0, \dots, 0). \end{aligned}$$

The approximation error is

$$\|A - W\Sigma_r V^t\|_{\text{F}} = \sqrt{\sigma_{r+1}^2 + \cdots + \sigma_m^2}.$$

Image compression

- Think of images as matrices.
- Substitute a matrix $W\Sigma V^t$ by $W\Sigma_r V^t$ with r small.
- To reconstruct $W\Sigma_r V^t$, only need to know
  - the r singular values $\sigma_1, \dots, \sigma_r$, — r
  - the first r columns of W, and — $\text{height} \cdot r$
  - the top r rows of V^t. — $\text{width} \cdot r$
- Total amount:
  $r \cdot (1 + \text{height} + \text{weight}) \ll \text{height} \cdot \text{width}$

Proper orthogonal decomposition

Given data points $\mathbf{x}_i \in \mathbb{R}^N$, want to find a low-dimensional linear subspace which **approximately contains** the $\mathbf{x}_i$.

Minimize

$$J(U) := \sum_i \|\mathbf{x}_i - P_U(\mathbf{x}_i)\|^2$$

under all r-dimensional subspaces $U \subseteq \mathbb{R}^N$, $r \ll N$, where $P_U : \mathbb{R}^N \to \mathbb{R}^N$ is the orthogonal projection onto U.

Proper orthogonal decomposition

Given data points $\mathbf{x}_i \in \mathbb{R}^N$, want to find a low-dimensional linear subspace which **approximately contains** the $\mathbf{x}_i$.

Minimize

$$J(U) := \sum_i \|\mathbf{x}_i - P_U(\mathbf{x}_i)\|^2$$

under all r-dimensional subspaces $U \subseteq \mathbb{R}^N$, $r \ll N$, where $P_U : \mathbb{R}^N \to \mathbb{R}^N$ is the orthogonal projection onto U.

More concrete formulation: Minimize

$$J(\mathbf{u}_1, \ldots, \mathbf{u}_r) := \sum_i \left\| \mathbf{x}_i - \sum_{j=1}^{r} \langle \mathbf{x}_i, \mathbf{u}_j \rangle \mathbf{u}_j \right\|^2,$$

where $\mathbf{u}_1, \ldots, \mathbf{u}_r \in \mathbb{R}^N$, $\langle \mathbf{u}_j, \mathbf{u}_k \rangle = \delta_{jk}$.

Principal component analysis

Given observations $x_i^{(k)}$ of random variables $X^{(k)}$, want to find **linearly uncorrelated** principal components.

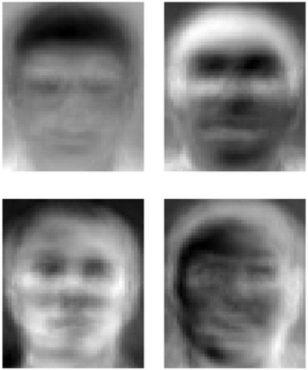
Write $X = (\mathbf{x}_1 | \cdots | \mathbf{x}_\ell) \in \mathbb{R}^{N \times \ell}$. Calculate $X = W \Sigma V^t$.
Then the principal components are the variables

$$Y^{(j)} = \sum_k W_{kj} X^{(k)}.$$

Most of the variance is captured by $Y^{(1)}$; second to most is captured by $Y^{(2)}$; and so on.

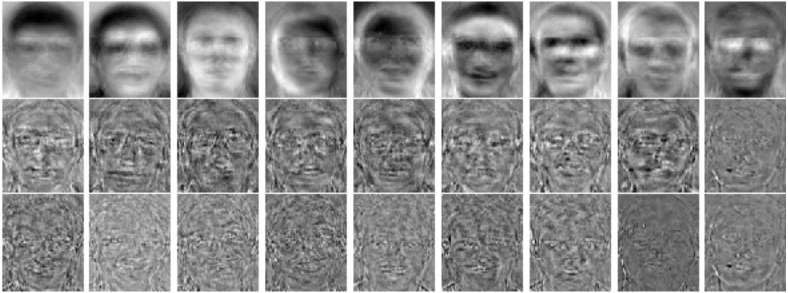
Eigenfaces

- Record sample faces $\mathbf{x}_1, \ldots, \mathbf{x}_N \in \mathbb{R}^{\text{width}\cdot\text{height}}$.
- Calculate a POD basis of **eigenfaces**.
- Recognize faces by looking at the coefficients of the most important eigenfaces.



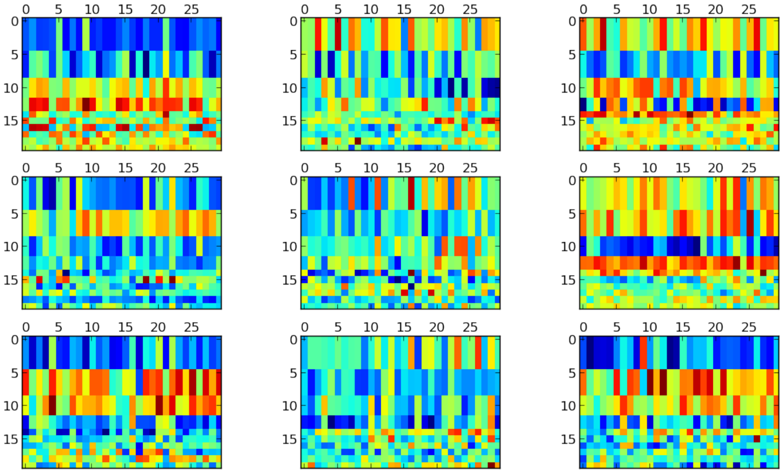
Eigenfaces resemble faces.

More eigenfaces

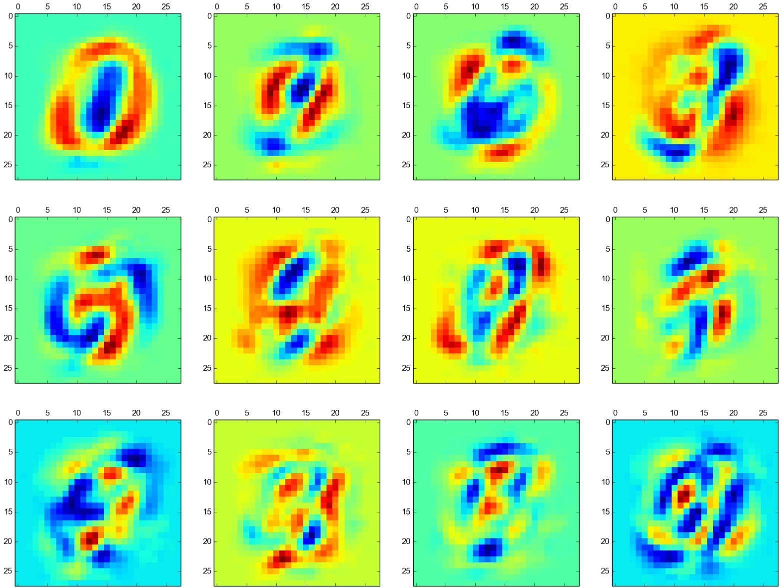


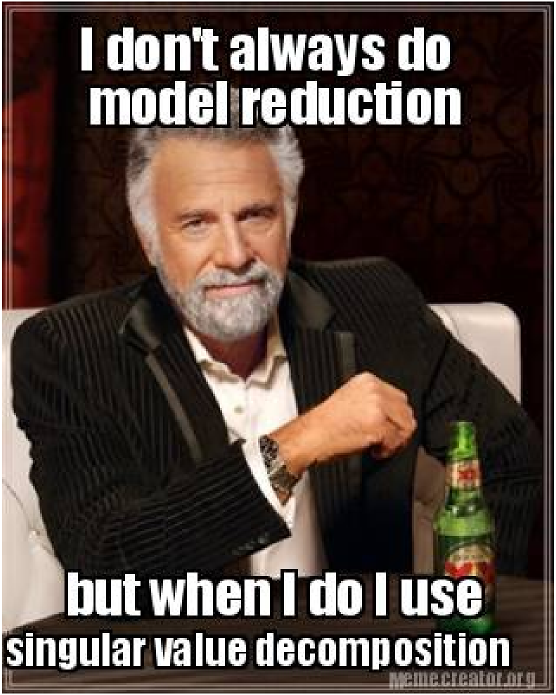
Digit recognition

Apply POD for dimension reduction, then use some similarity measure or clustering technique. Results:



Eigendigits




I don't always do model reduction
but when I do I use
singular value decomposition
memecreator.org