Principal Component Analysis (PCA)

DS351

Last time

- 2D Matrices as transformations (rotations etc.)

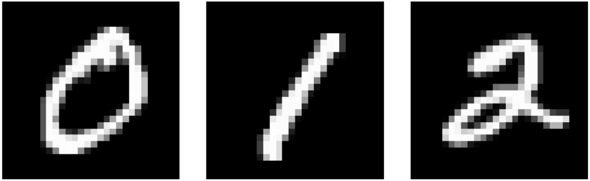
Today

- Basic ideas of Principal Component Analysis
- Transformations in higher dimension
- Eigenvalues and eigenvectors

Dimensionality reduction

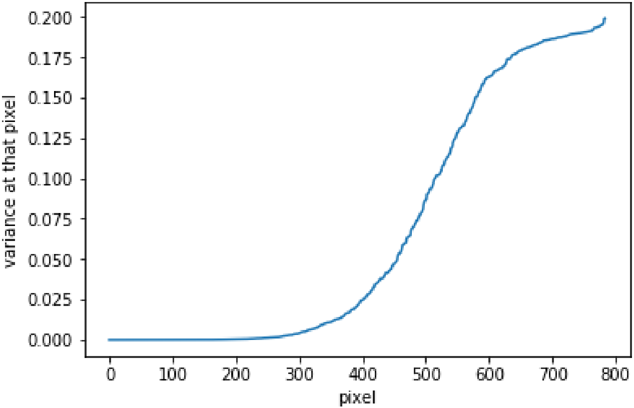
Why remove some of the features?

- Save storage and computation time.
- Reduce some redundancy in the data.
- Remove noises in the data.



... but sometimes it's not easy to find which feature should be removed.

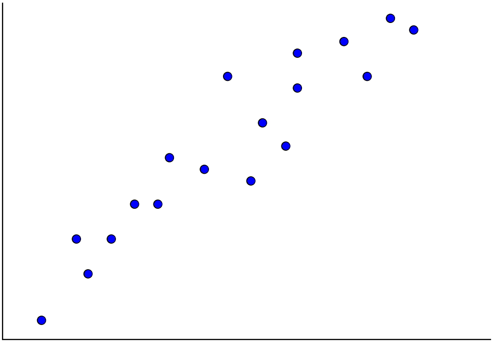
MNIST example



First 300 pixels with the lowest variance are undesirable features.

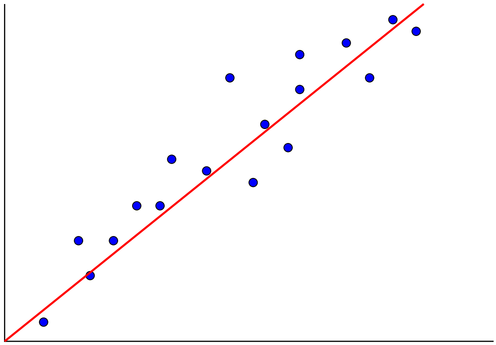
A simple case

Suppose we want to reduce from 2D data to 1D.



A simple case

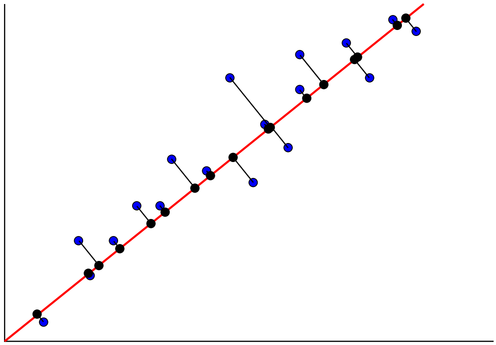
Suppose we want to reduce from 2D data to 1D.



The line is in the direction of maximum variance

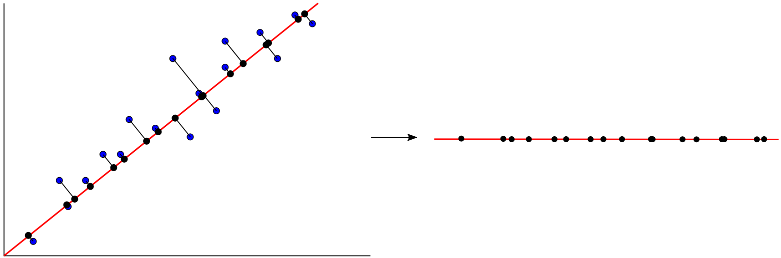
A simple case

Suppose we want to reduce from 2D data to 1D.



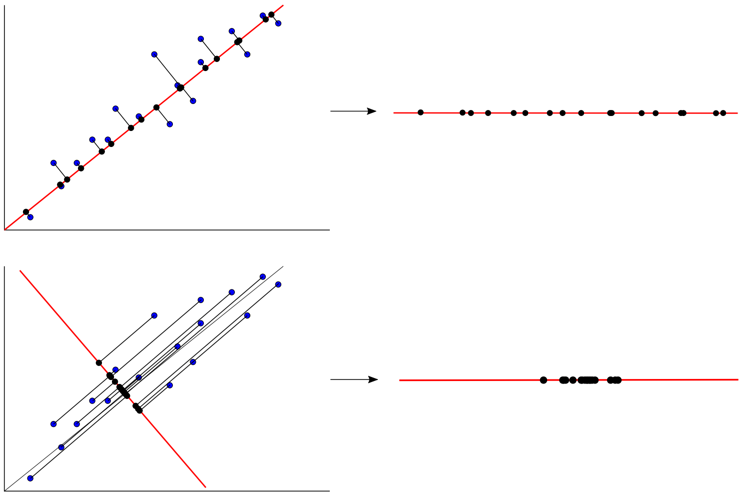
Make projections on this line.

From 2D to 1D



The line becomes the 1D axis.

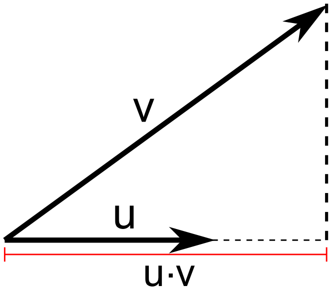
Comparison between two directions



Which red line is better?

Vector Projection

If we want to project a vector v in a direction of a **unit vector** u,



then the length of projection is $u \cdot v$.

Examples

What is the projection of $v = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ in the following directions?

- The x axis.
- The direction of $u = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$.

The best direction

Suppose we have d-dimensional data

$x_1, x_2, x_3, \ldots, x_n$ (These are d-dimensional vectors.)

The best direction

Suppose we have d-dimensional data

$$x_1, x_2, x_3, \ldots, x_n \qquad \text{(These are } d\text{-dimensional vectors.)}$$

The goal is to find the unit vector u that maximizes the variance in the direction of u i.e. the variance of

$$x_1 \cdot u, x_2 \cdot u, \ldots, x_n \cdot u$$

How can we find such u?

To answer this question, we look at the **covariance matrix** of X.

Covariance matrix

For $i = 1, 2, \ldots, d$, let X_i be the list of observed values of the i-th variable.

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,i} & \cdots & x_{1,d} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,i} & \cdots & x_{2,d} \\ x_{3,1} & x_{3,2} & \cdots & x_{3,i} & \cdots & x_{3,d} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,i} & \cdots & x_{n,d} \end{bmatrix}$$

Covariance matrix

For $i = 1, 2, \ldots, d$, let X_i be the list of observed values of the i-th variable.

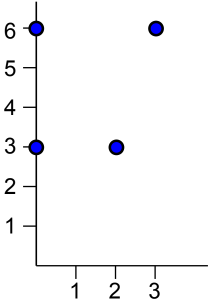
$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & \ldots & x_{1,i} & \ldots & x_{1,d} \\ x_{2,1} & x_{2,2} & \ldots & x_{2,i} & \ldots & x_{2,d} \\ x_{3,1} & x_{3,2} & \ldots & x_{3,i} & \ldots & x_{3,d} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \ldots & x_{n,i} & \ldots & x_{n,d} \end{bmatrix}$$

Then the covariance matrix is a $d \times d$ matrix defined by

$$\Sigma = \begin{bmatrix} \mathrm{Cov}(X_1, X_1) & \mathrm{Cov}(X_1, X_2) & \ldots & \mathrm{Cov}(X_1, X_d) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Cov}(X_2, X_2) & \ldots & \mathrm{Cov}(X_2, X_d) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_d, X_1) & \mathrm{Cov}(X_d, X_2) & \ldots & \mathrm{Cov}(X_d, X_d) \end{bmatrix}$$

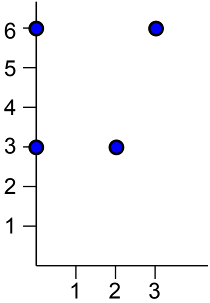
Example

Suppose we have four data points $X = \{(0,3),(2,3),(3,6),(0,6)\}$.
Compute the covariance matrix



Example

Suppose we have four data points $X = \{(0,3),(2,3),(3,6),(0,6)\}$. Compute the covariance matrix



Answer: $X_1 = (0,2,3,0)$, $X_2 = (3,3,6,6)$

$$\Sigma = \begin{bmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}(X_2) \end{bmatrix} = \begin{bmatrix} 2.25 & 0.5 \\ 0.5 & 3 \end{bmatrix}.$$

Find the best direction

Find the unit vector u that maximizes the variance in the direction of u i.e. the variance of

$$x_1 \cdot u, x_2 \cdot u, \ldots, x_n \cdot u$$

How can we find such u?

Find the best direction

Find the unit vector u that maximizes the variance in the direction of u i.e. the variance of

$$x_1 \cdot u, x_2 \cdot u, \ldots, x_n \cdot u$$

How can we find such u?

Fact:

- Let Σ be the covariance matrix of X.
- The variance of X in direction u is given by $u^T \Sigma u$.

Example

The data $X = \{(0,3),(2,3),(3,6),(0,6)\}$ has the covariance matrix

$$\Sigma = \begin{pmatrix} 2.25 & 0.5 \\ 0.5 & 3 \end{pmatrix}, \qquad u = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

After projecting X on u, the variance of the projections are

$$\begin{aligned} u^T \Sigma u &= \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 2 \end{pmatrix} \begin{pmatrix} 2.25 & 0.5 \\ 0.5 & 3 \end{pmatrix} \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix} \\ &= \frac{1}{5} \begin{pmatrix} 1 & 2 \end{pmatrix} \begin{pmatrix} 3.25 \\ 6.5 \end{pmatrix} \\ &= \end{aligned}$$

Spectral decomposition

Fact: Any real symmetric matrix Σ can be decomposed as

$$\Sigma = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ u_1 & u_2 & \ldots & u_d \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_d \end{pmatrix} \begin{pmatrix} \longleftarrow & u_1 & \longrightarrow \\ \longleftarrow & u_2 & \longrightarrow \\ & \vdots & \\ \longleftarrow & u_d & \longrightarrow \end{pmatrix}$$

where

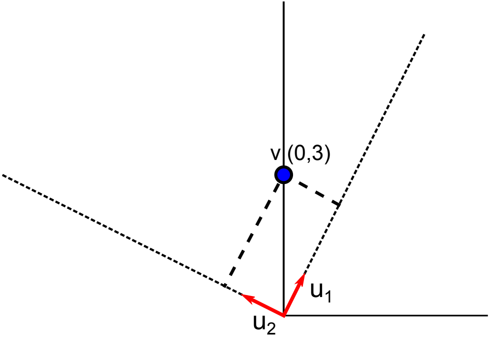
- $\lambda_1 > \lambda_2 > \ldots > \lambda_d$ are the **eigenvalues**.
- $u_1, u_2, \ldots, u_d$ are the **eigenvectors** of length d.
- The maximum variance that we can get is λ_1.

Eigenvectors

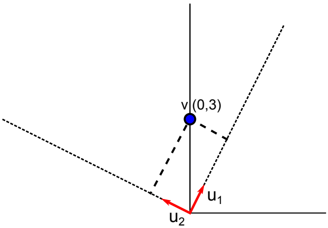
Fact: The eigenvectors $u_1, u_2, \ldots, u_d$ are **orthonormal**, meaning that

- They have length one.
- They are perpendicular to each other.

Therefore, $u_1, u_2, \ldots, u_d$ form another coordinate for the data points.

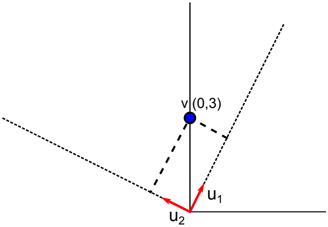


Example



$$\Sigma = \begin{pmatrix} 2.25 & 0.5 \\ 0.5 & 3 \end{pmatrix}$$

- Eigenvalues: $\lambda_1 = 3.25, \lambda_2 = 2$
- Eigenvalues: $u_1 = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix}, u_2 = \frac{1}{\sqrt{5}} \begin{pmatrix} -2 \\ 1 \end{pmatrix}$.



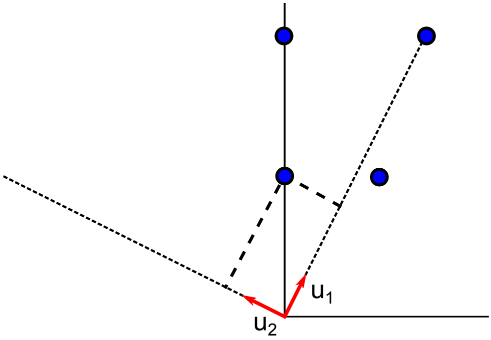
The point $v = (0, 3)$ in the new axis is

$$(v \cdot u_1, v \cdot u_2)$$

where

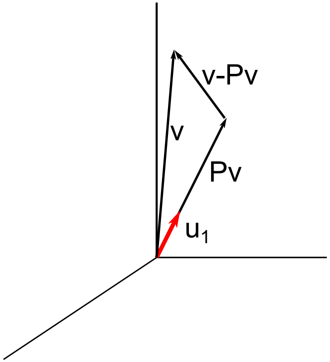
$v \cdot u_1 =$

$v \cdot u_2 =$



- Highest variance $= 3.25$ in the direction of u_1.
- Variance $= 2$ in the direction of u_2.

The second best direction



- Suppose we have 3D data.
- Explain the data using the eigenvector u_1 which has the highest eigenvalue.
- What is the best among the remaining directions?

Spectral decomposition (revisited)

$$\Sigma = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ u_1 & u_2 & \dots & u_d \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_d \end{pmatrix} \begin{pmatrix} \longleftarrow & u_1 & \longrightarrow \\ \longleftarrow & u_2 & \longrightarrow \\ & \vdots & \\ \longleftarrow & u_d & \longrightarrow \end{pmatrix}$$

- The second best direction is u_2 with associated variance $lambda_2$.
- The third best direction is u_2.
- and so on...

Spectral decomposition (revisited)

$$\Sigma = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ u_1 & u_2 & \cdots & u_d \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_d \end{pmatrix} \begin{pmatrix} \longleftarrow & u_1 & \longrightarrow \\ \longleftarrow & u_2 & \longrightarrow \\ & \vdots & \\ \longleftarrow & u_d & \longrightarrow \end{pmatrix}$$

- The second best direction is u_2 with associated variance $lambda_2$.
- The third best direction is u_2.
- and so on...

To find the best k directions ($k < d$), pick $u_1, u_2, \dots, u_k$.

Reconstruction

How can we know if PCA does not destroy the structure of the data? Reconstruction.

- k principal axes: $u_1, u_2, \ldots, u_k$.
- In these axes, the coordinate of the PCA of a point u is

$$(u \cdot u_1, u \cdot u_2, \ldots, u \cdot u_k).$$

Reconstruction

How can we know if PCA does not destroy the structure of the data? Reconstruction.

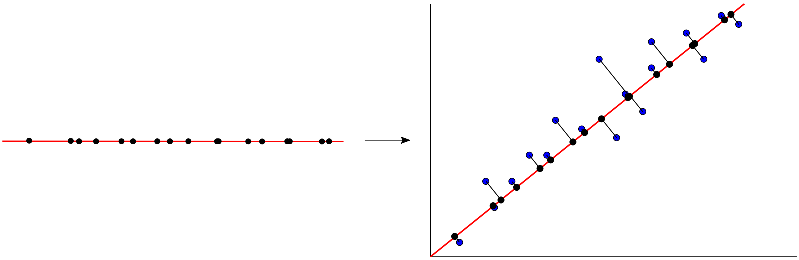
- k principal axes: $u_1, u_2, \ldots, u_k$.
- In these axes, the coordinate of the PCA of a point u is

$$(u \cdot u_1, u \cdot u_2, \ldots, u \cdot u_k).$$

Reverse this point back to the original coordinate using

$$(u \cdot u_1)u_1 + (u \cdot u_2)u_2 + \ldots + (u \cdot u_k)u_k.$$

Reconstruction



The reconstructions are the black points on the red line. We see that there is some information loss in the process.

Reconstruction of MNIST

Reconstruct this original image x from its PCA projection to k dimensions.

$k = 200$

$k = 150$

$k = 100$

$k = 50$