Principal component analysis

229351

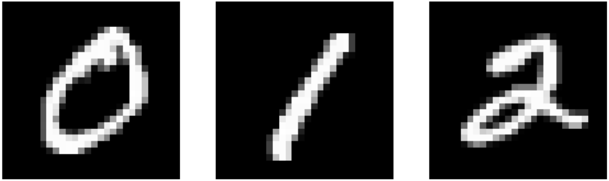
Dimensionality reduction

Why remove some of the features?

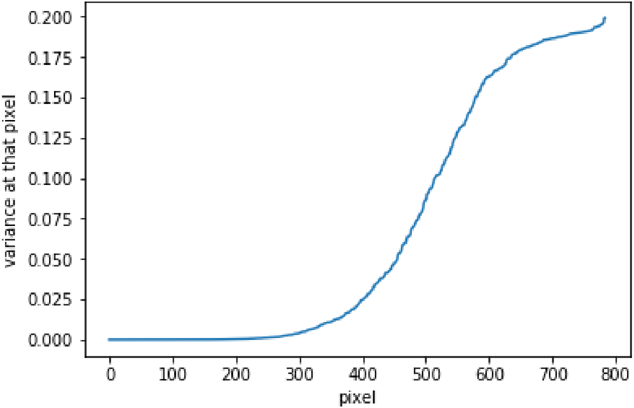
- Save storage and computation time.

- Reduce some redundancy in the data.

- Remove noises in the data.



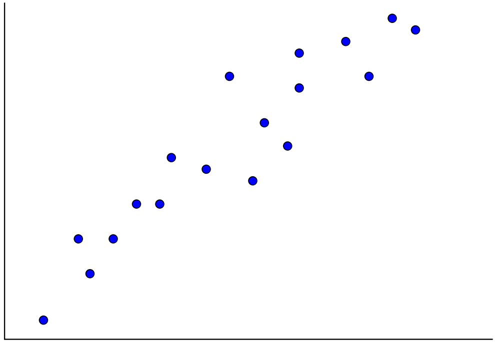
MNIST example



First 300 pixels with the lowest variance are undesirable features.

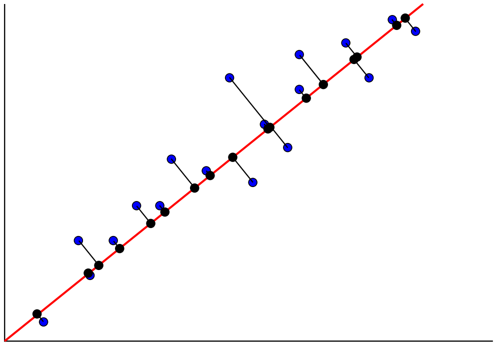
A simple case

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



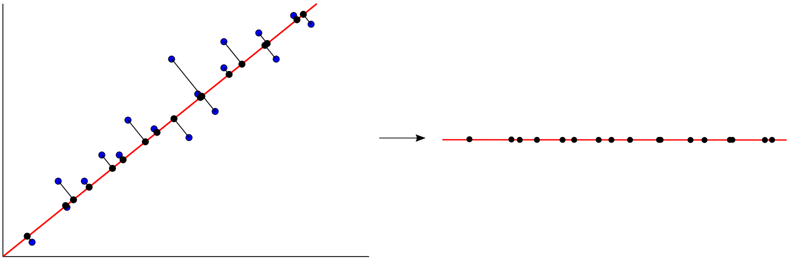
A simple case

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



Make projections on this line.

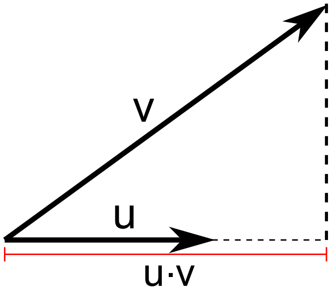
From 2D to 1D



The red line becomes the 1D axis.

Vector Projection

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



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

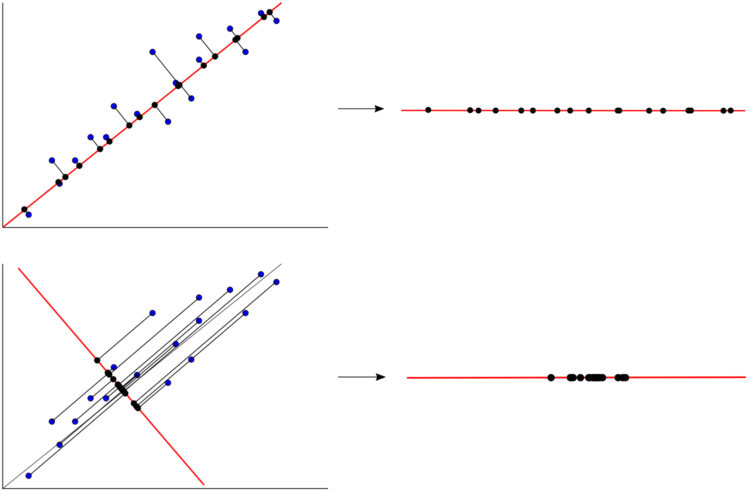
Examples

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

- The x axis.

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

Comparison between two directions



Which red line is better?

The best direction

Normalized data with d variables.

$$v_1, v_2, \ldots, v_n \in \mathbb{R}^d$$

The best direction

Normalized data with d variables.

$$v_1, v_2, \ldots, v_n \in \mathbb{R}^d$$

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

$$v_1 \cdot u, v_2 \cdot u, \ldots, v_n \cdot u$$

How can we find such u?

Answer: Look at the **covariance matrix** Σ.

Covariance matrix

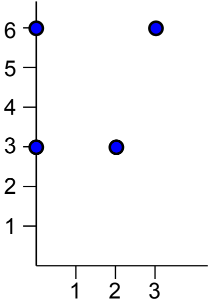
Let $X_1, X_2, \ldots, X_d$ be the variable vectors.

The covariance matrix is a $d \times d$ matrix defined by

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

Example

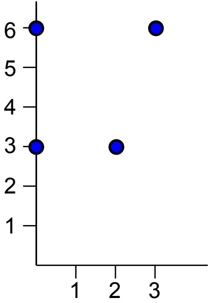
$D = \{(0,3), (2,3), (3,6), (0,6)\}.$



Answer: $X_1 =$, $X_2 =$

Example

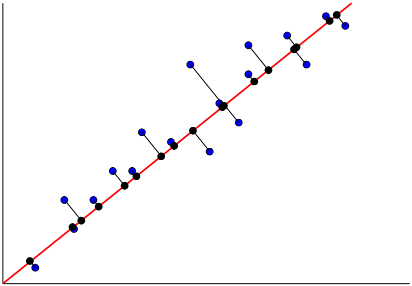
$D = \{(0,3),(2,3),(3,6),(0,6)\}.$



Answer: $X_1 =$, $X_2 =$

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

Finding the best direction u



Data: $v_1, v_2, \ldots, v_n$

Projections on u: $(v_1 \cdot u, v_2 \cdot u, \ldots, v_n \cdot u)$

$$\text{Var}(v_1 \cdot u, \ldots, v_n \cdot u) = u^T \Sigma u.$$

Example

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

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

The variance of the projections on u is

Spectral decomposition

Fact: Any symmetric matrix Σ can be decomposed as

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

where

- $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d$ are the **eigenvalues**.
- $u_1, u_2, \ldots, u_d$ are the **eigenvectors** of length d.
- $u_1, u_2, \ldots, u_d$ are orthogonal unit vectors .

Eigenvectors

Fact: Any symmetric matrix Σ can be decomposed as

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

The eigenvectors $u_1, u_2, \dots, u_d$ are:

- unit vectors
- perpendicular to each other.

Example

Suppose that data $v_1, v_2, \ldots, v_n$ has convariance matrix Σ.
Compute Var$(v_1 \cdot u_1, v_2 \cdot u_1, \ldots, v_n \cdot u_1)$.

Example

Example

$$\Sigma = \begin{pmatrix} 2.25 & 0.5 \\ 0.5 & 3 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} 3.25 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}$$

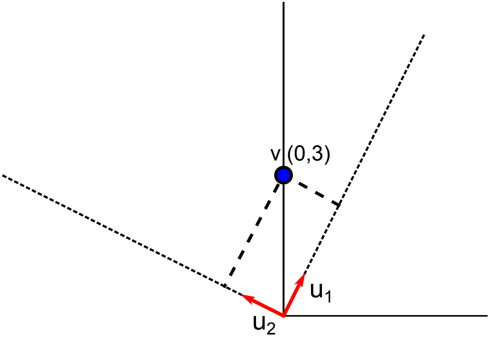
- Eigenvalues:

- Eigenvectors:

Matrix as a transformation

$$\Sigma = \begin{pmatrix} 2.25 & 0.5 \\ 0.5 & 3 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} 3.25 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}$$

$$= \begin{pmatrix} \cos(63) & -\sin(63) \\ \sin(63) & \cos(63) \end{pmatrix} \begin{pmatrix} 3.25 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} \cos(63) & \sin(63) \\ -\sin(63) & \cos(63) \end{pmatrix}$$



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

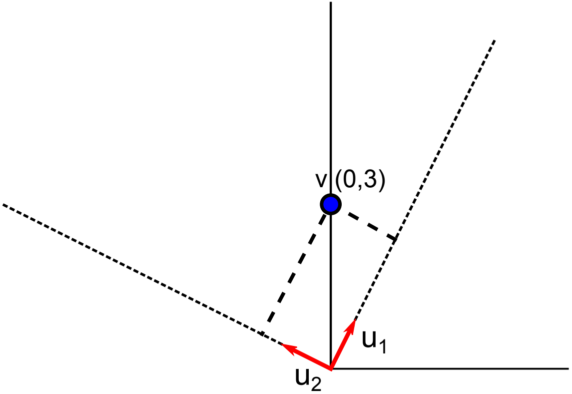
$$(v \cdot u_1, v \cdot u_2)$$

where

$$v \cdot u_1 =$$

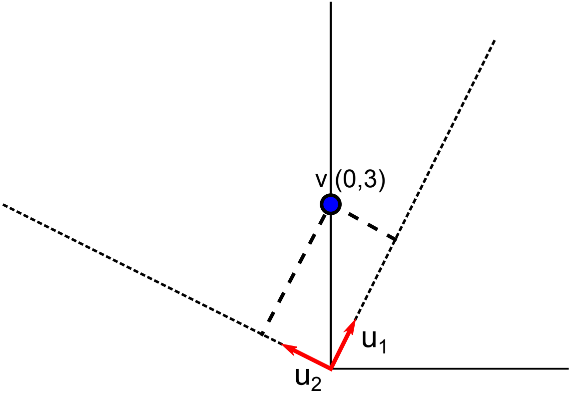
$$v \cdot u_2 =$$

Rotations preserve the length



Observation: $\|v\| = \|(v \cdot u_1, v \cdot u_2)\|$

Rotations preserve the length



Observation: $\|v\| = \|(v \cdot u_1, v \cdot u_2)\|$

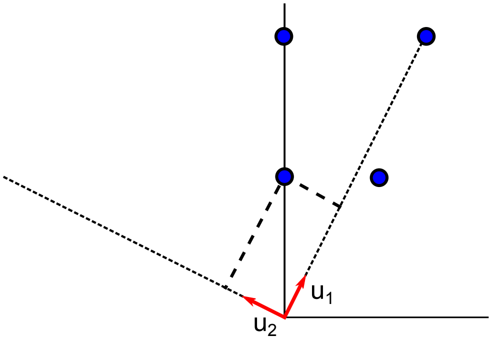
In higher dimension, $\|Uv\| = \|v\|$.

The highest variance is λ_1

$$\text{Var}(v_1 \cdot u, v_2 \cdot u, \ldots, v_n \cdot u)$$

Finding the best direction u

Example



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

Spectral decomposition

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

- The second best direction is u_2 with associated variance λ_2.
- The third best direction is u_3 with associated variance λ_3.
- and so on...

Principal component analysis

Let $u \in \mathbb{R}^d$ be a data point.

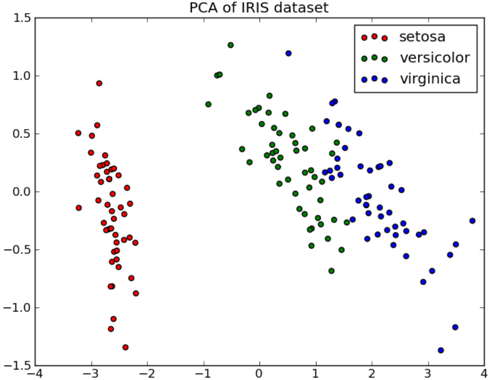
Principal axes ($k < d$):

$$u_1, u_2, \ldots, u_k$$

The PCA of u is

$$(u \cdot u_1, u \cdot u_2, \ldots, u \cdot u_k) \in \mathbb{R}^k.$$

PCA of iris flowers



$$\lambda_1 = 4.23, \quad \lambda_2 = 0.24$$
$$u_1 = (0.36, -0.08, 0.86, 0.36)$$
$$u_2 = (0.66, 0.73, -0.17, -0.07)$$

Three species of iris

- Setosa
- Versicolor
- Virginica

Four variables

- x_1: sepal length
- x_2: sepal width
- x_3: petal length
- x_4: petal width

Reconstruction

Eigenvectors: $u_1, u_2, \ldots, u_d$.

- k principal axes: $u_1, u_2, \ldots, u_k \in \mathbb{R}^d$.

- In these axes, the coordinate of the PCA of a point u is

$$(u \cdot u_1, u \cdot u_2, \ldots, u \cdot u_k) \in \mathbb{R}^k.$$

Reconstruction

Eigenvectors: $u_1, u_2, \ldots, u_d$.

- k principal axes: $u_1, u_2, \ldots, u_k \in \mathbb{R}^d$.

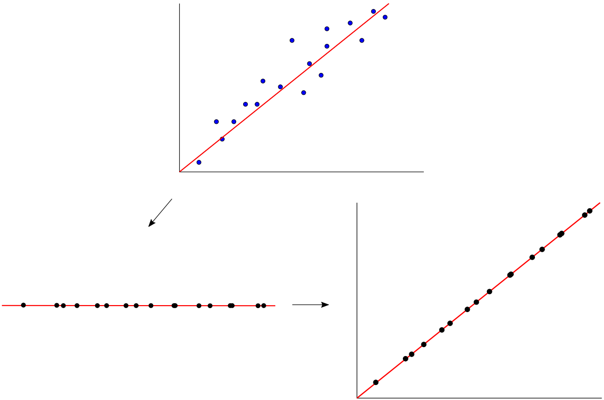
- In these axes, the coordinate of the PCA of a point u is

$$(u \cdot u_1, u \cdot u_2, \ldots, u \cdot u_k) \in \mathbb{R}^k.$$

Reverse this point back to the original coordinate using

$$(u \cdot u_1)u_1 + (u \cdot u_2)u_2 + \ldots + (u \cdot u_k)u_k \in \mathbb{R}^d.$$

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.

$k = 200$ $\quad$ $k = 150$ $\quad$ $k = 100$ $\quad$ $k = 50$