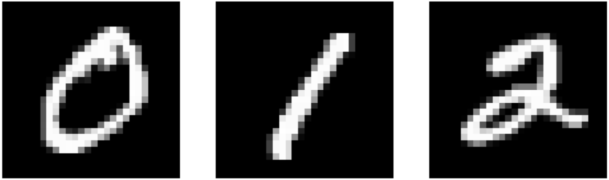
Principal component analysis

229351

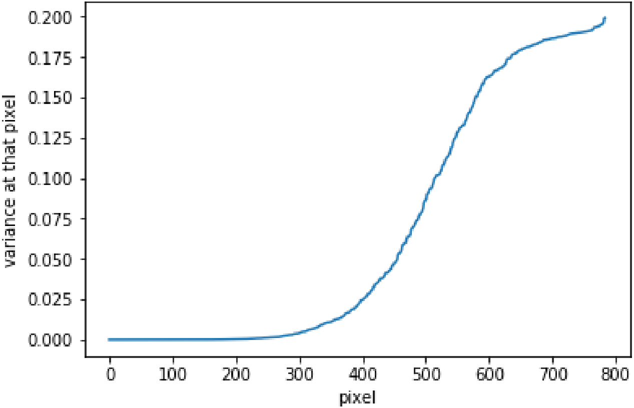
Dimensionality reduction

Why remove some of the features?

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



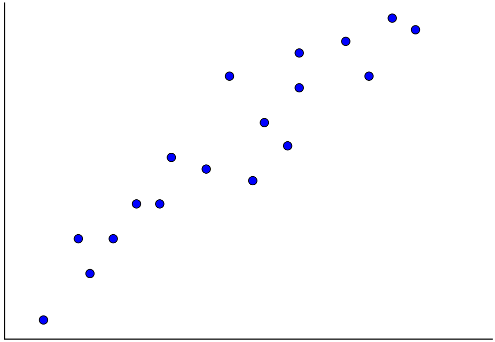
MNIST example



First 300 pixels with the lowest variance are undesirable features.

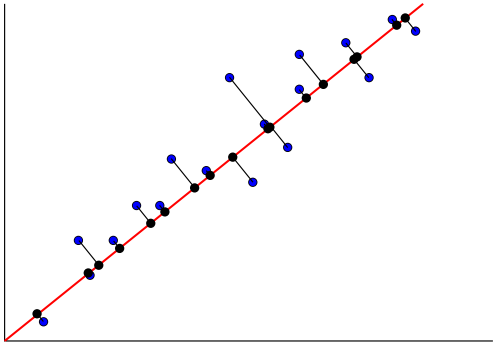
A simple case

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



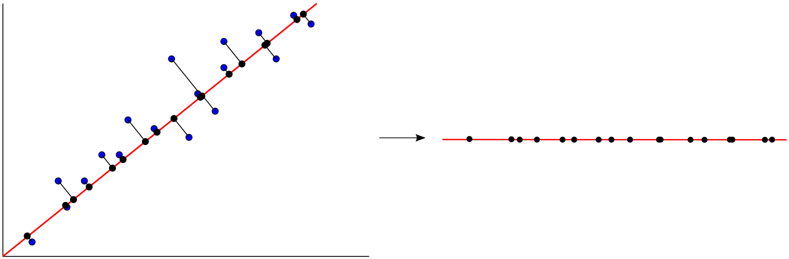
A simple case

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



Make projections on this line.

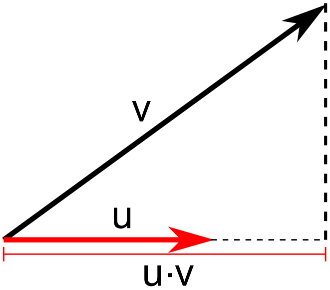
From 2D to 1D



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 $u \cdot v$.

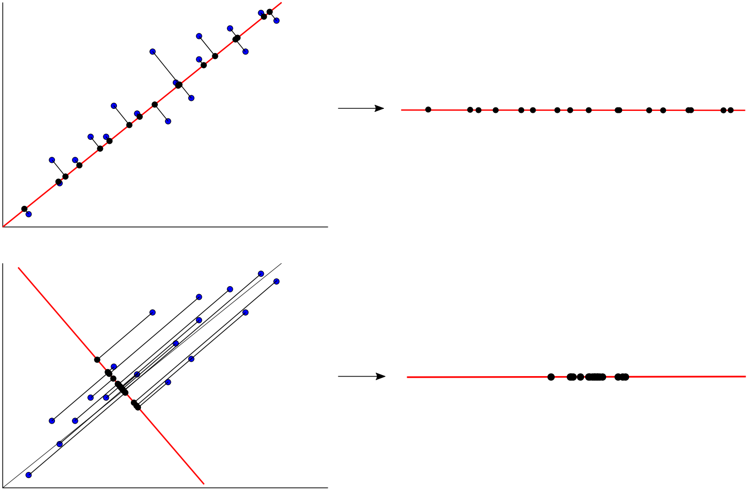
Examples

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?

Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is the technique of finding directions (principal components) that capture the variance of the data

Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is the technique of finding directions (principal components) that capture the variance of the data

The **first principal component** is the direction of the maximum variance

Sample Covariance

Suppose we have a data table:

	X_1	X_2	X_3	$\dots$	X_d
1	a_1	b_1			
2	a_2	b_2			
3	a_3	b_3			
$\vdots$	$\vdots$		$\vdots$		
n	a_n	b_n			

The **Sample Covariance** between X_1 and X_2 is

$$\mathrm{Cov}(X_1, X_2) = \frac{1}{n-1}\left[X_1 \cdot X_2 - \overline{X}_1 \overline{X}_2\right].$$

Sample Covariance

Positive correlation

Negative correlation

No correlation

Covariance matrix

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

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

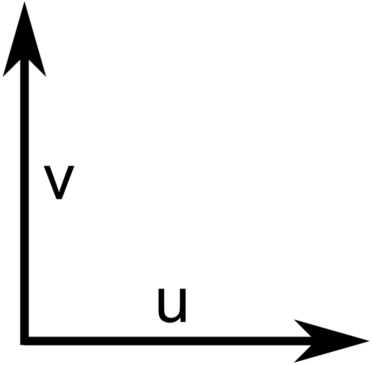
$$\Sigma = \begin{bmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \dots & \text{Cov}(X_1, X_d) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \dots & \text{Cov}(X_2, X_d) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_d, X_1) & \text{Cov}(X_d, X_2) & \dots & \text{Var}(X_d) \end{bmatrix}$$

Example

Data with two variables:
$D = \{(0,1),(2,3),(5,0),(1,8)\}$.

$\Sigma =$

Orthogonal vectors



A basic of orthogonal unit vectors

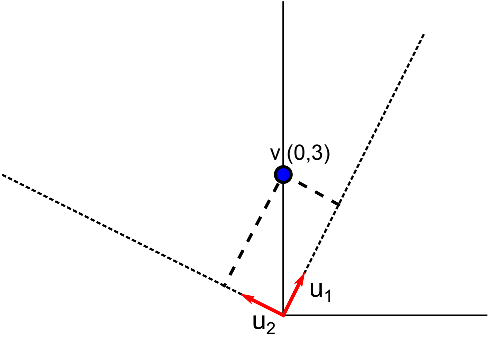
Suppose we have d orthogonal unit vectors in $\mathbb{R}^d$:

$$u_1, u_2, \ldots, u_d$$

We can write any vector $u \in \mathbb{R}^d$ as a linear combination of $u_1 \ldots, u_d$:

$$u = a_1 u_1 + a_2 u_2 + \ldots + a_d u_d = \sum_{i=1}^{d} a_i u_i$$

Example



$$u_1 = [0.6, 0.8] \qquad u_2 = [-0.8, 0.6]$$

Next slide is the main result from linear algebra that we will use for PCA...

Spectral decomposition

Fact: The covariance 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: The covariance 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 is the direction with maximum variance
- The maximum variance is λ_1

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 the second largest variance λ_2.

- The third best direction is u_3 with the third largest variance λ_3.

- and so on...

Principal component analysis

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

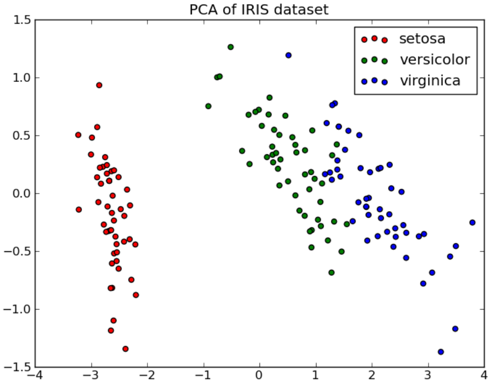
Principal axes ($k < d$):

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

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

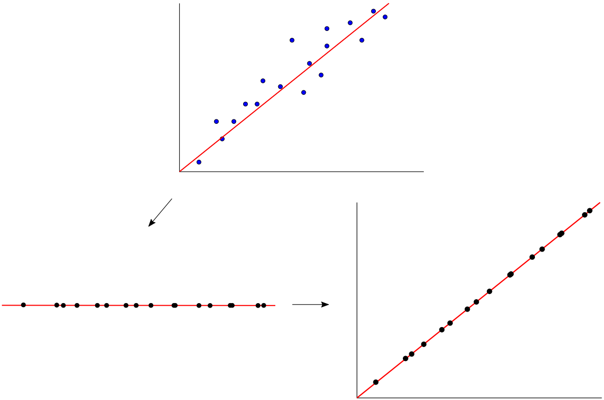
- In these axes, the coordinate of the PCA of a point u is

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

  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.

$k = 200$

$k = 150$

$k = 100$

$k = 50$