Principal Components Analysis

Dimensionality reduction

Why reduce the number of features in a data set?

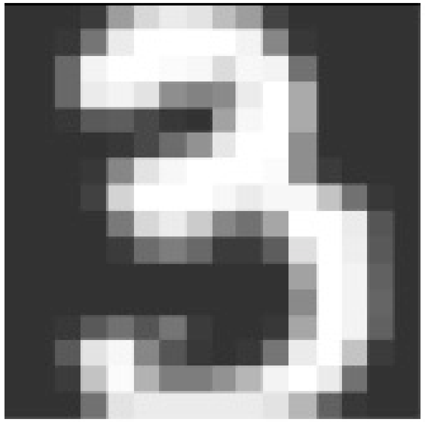
1. It reduces storage and computation time.
2. High-dimensional data often has a lot of redundancy.
3. Remove noisy or irrelevant features.

Dimensionality reduction

Why reduce the number of features in a data set?

1. It reduces storage and computation time.
2. High-dimensional data often has a lot of redundancy.
3. Remove noisy or irrelevant features.

Example: are all the pixels in an image equally informative?

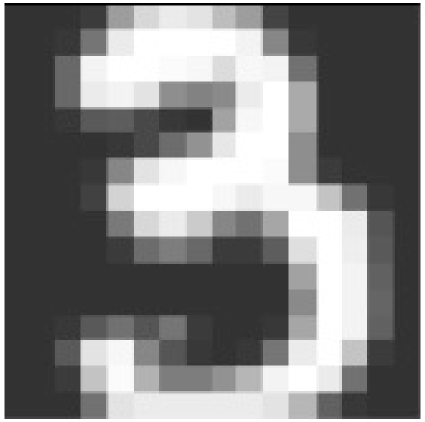


Dimensionality reduction

Why reduce the number of features in a data set?

1. It reduces storage and computation time.
2. High-dimensional data often has a lot of redundancy.
3. Remove noisy or irrelevant features.

Example: are all the pixels in an image equally informative?



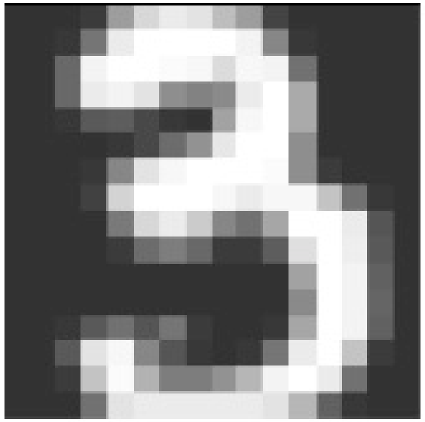
If we were to choose a few pixels to discard, which would be the prime candidates?

Dimensionality reduction

Why reduce the number of features in a data set?

1. It reduces storage and computation time.
2. High-dimensional data often has a lot of redundancy.
3. Remove noisy or irrelevant features.

Example: are all the pixels in an image equally informative?



If we were to choose a few pixels to discard, which would be the prime candidates?

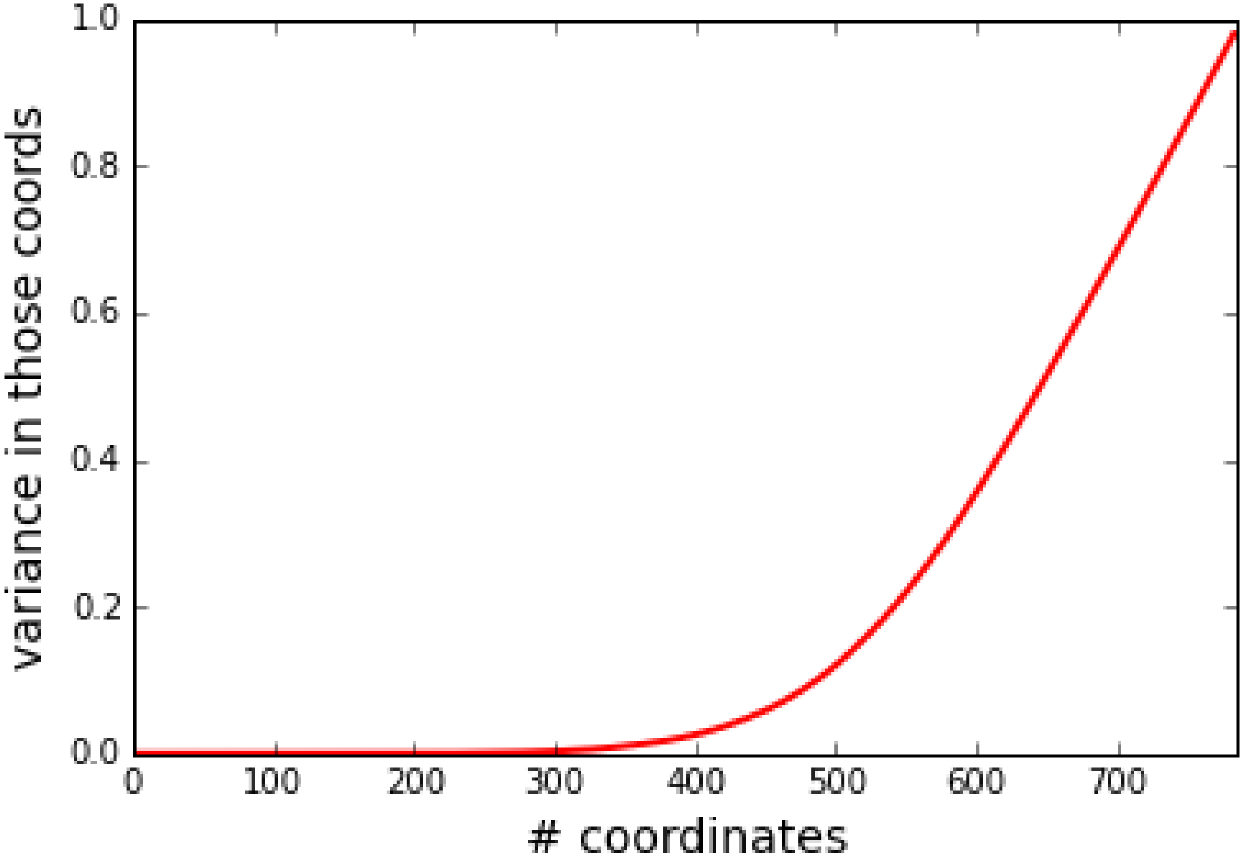
Those with lowest variance...

Eliminating low variance coordinates

Example: MNIST. What fraction of the total variance is contained in the 100 (or 200, or 300) coordinates with lowest variance?

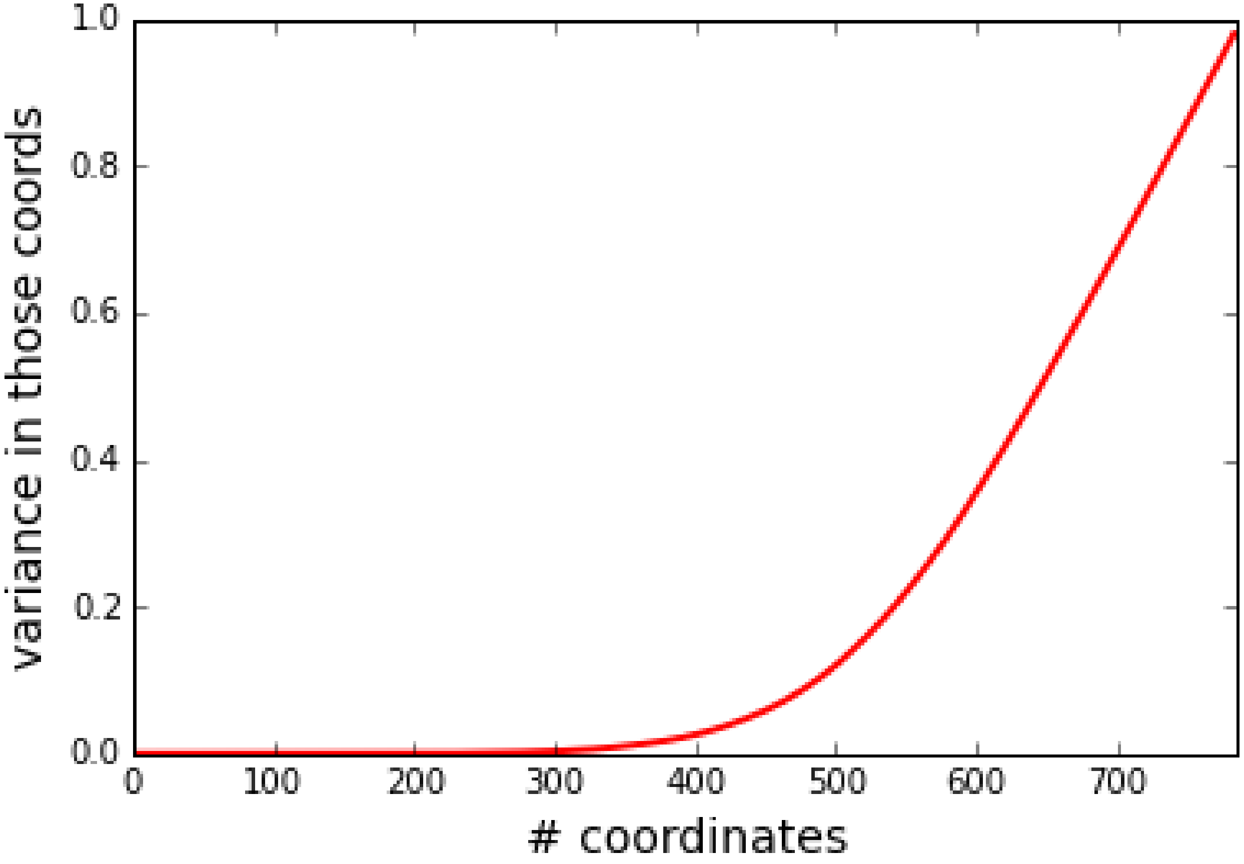
Eliminating low variance coordinates

Example: MNIST. What fraction of the total variance is contained in the 100 (or 200, or 300) coordinates with lowest variance?



Eliminating low variance coordinates

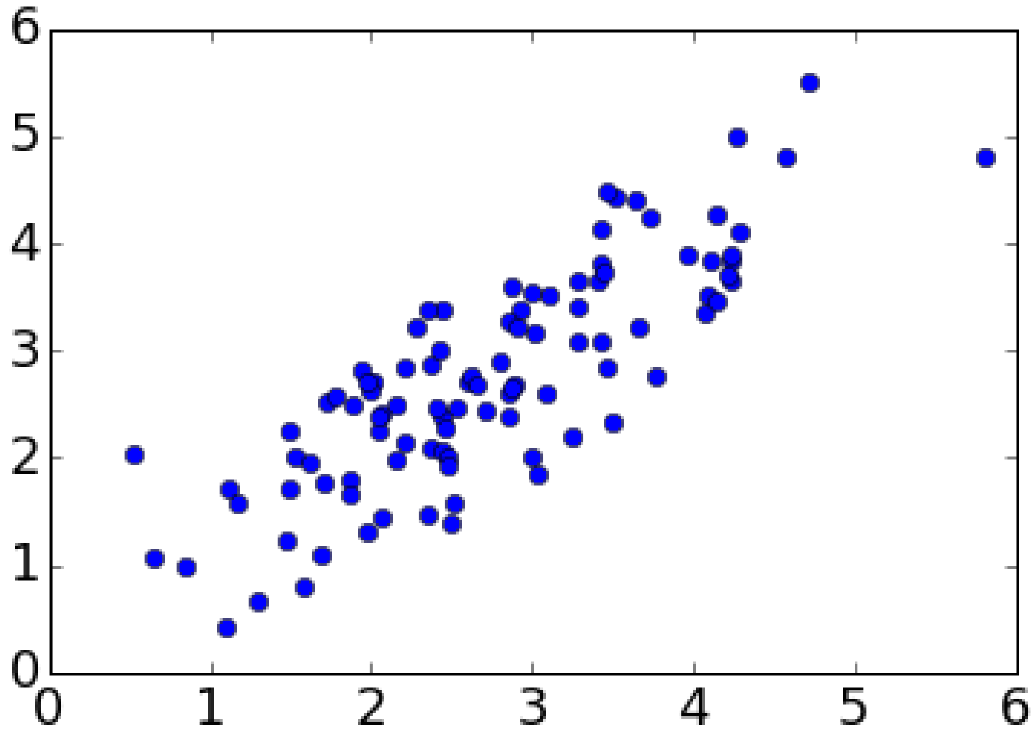
Example: MNIST. What fraction of the total variance is contained in the 100 (or 200, or 300) coordinates with lowest variance?



Could easily drop 300-400 pixels...

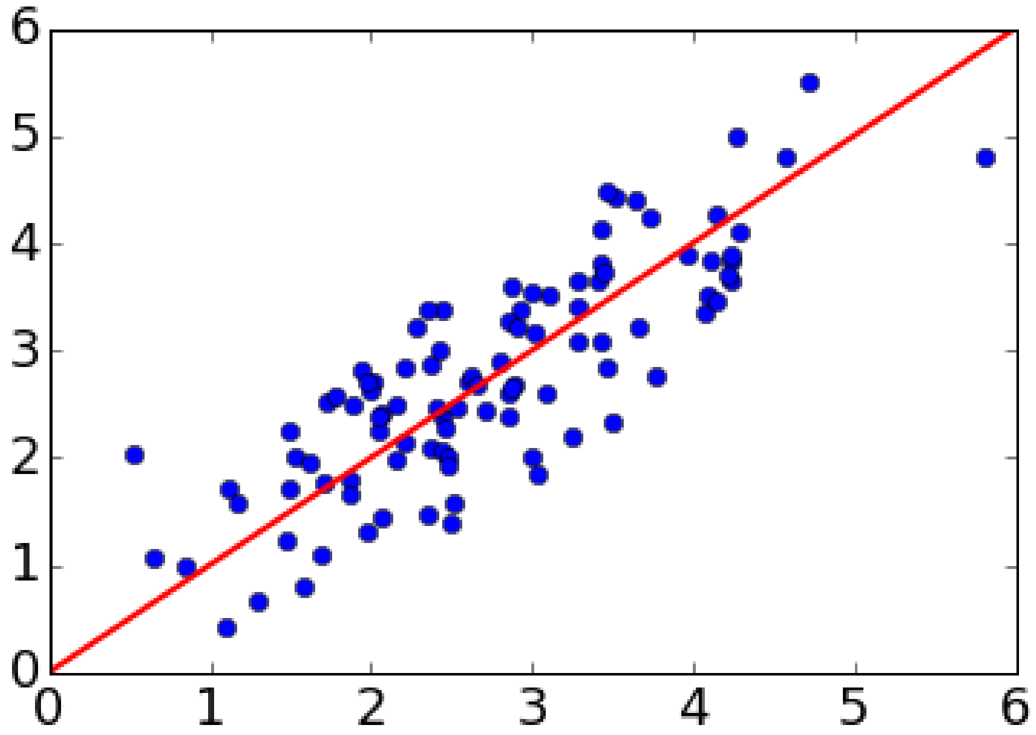
The effect of correlation

Suppose we wanted just one feature for the following data.



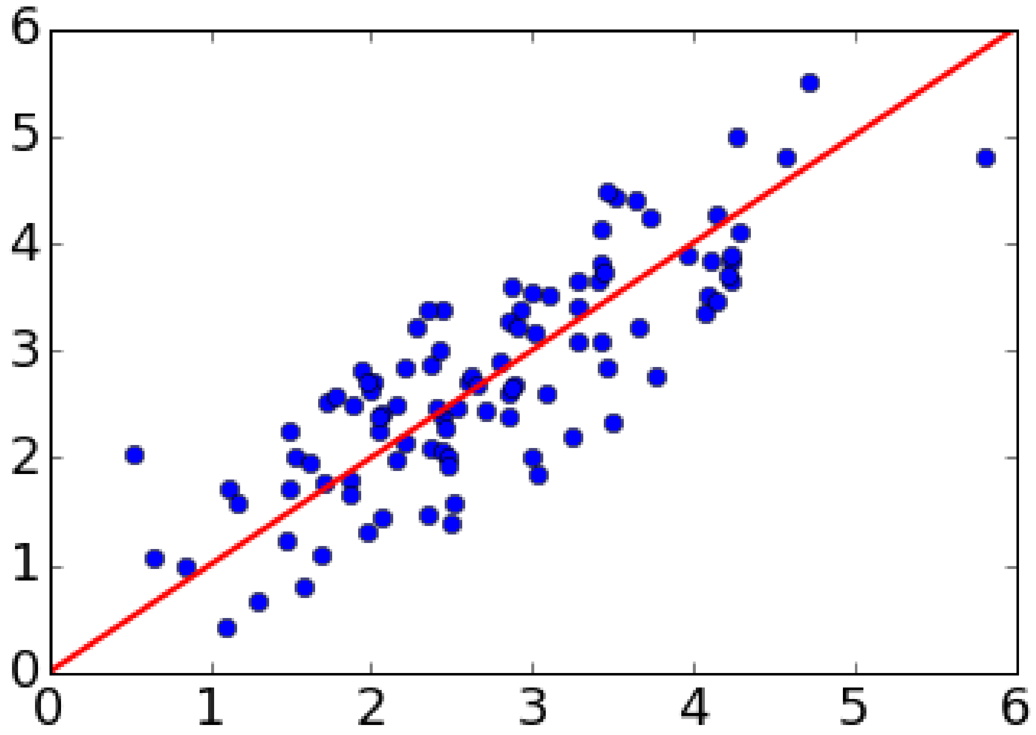
The effect of correlation

Suppose we wanted just one feature for the following data.



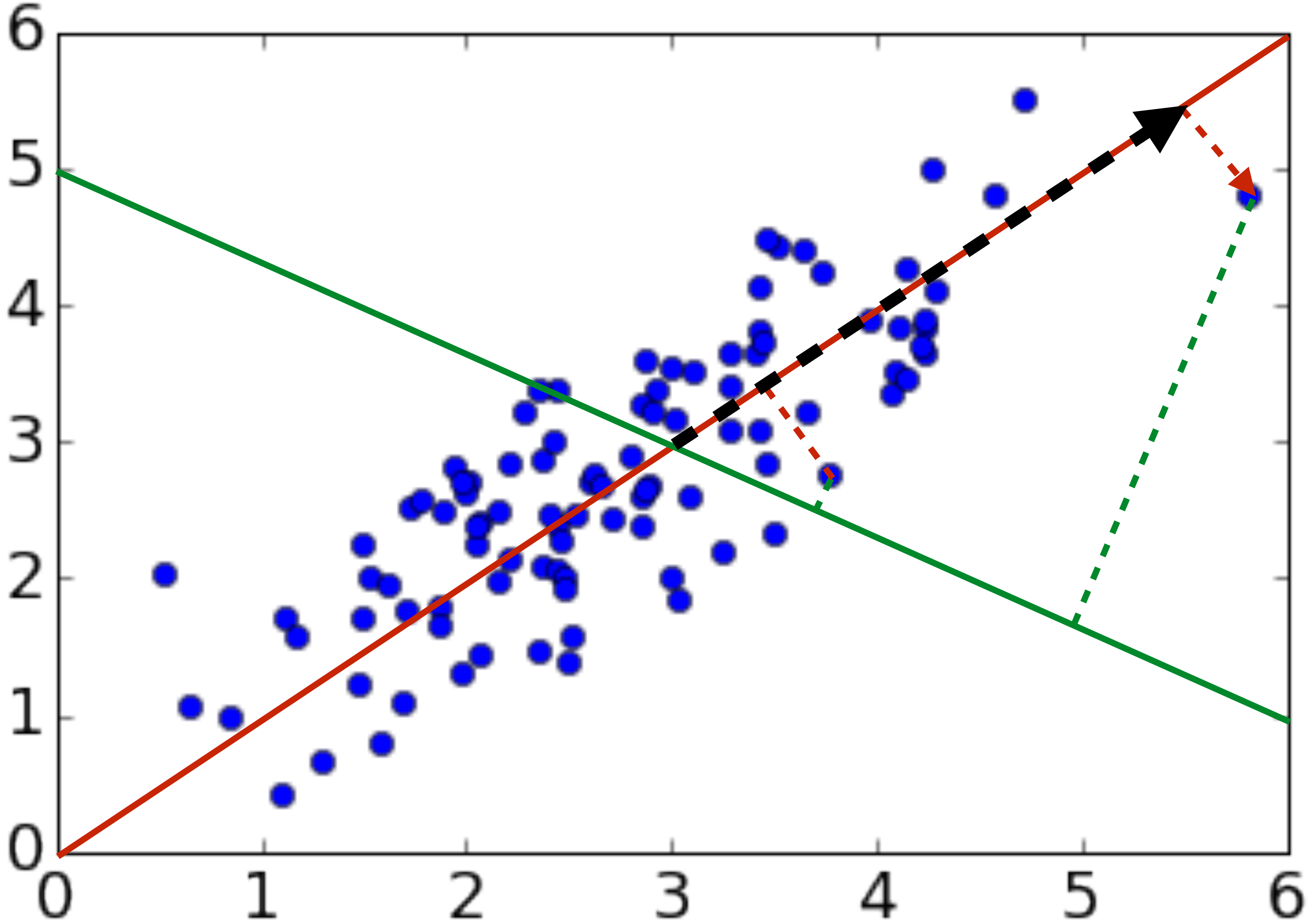
The effect of correlation

Suppose we wanted just one feature for the following data.

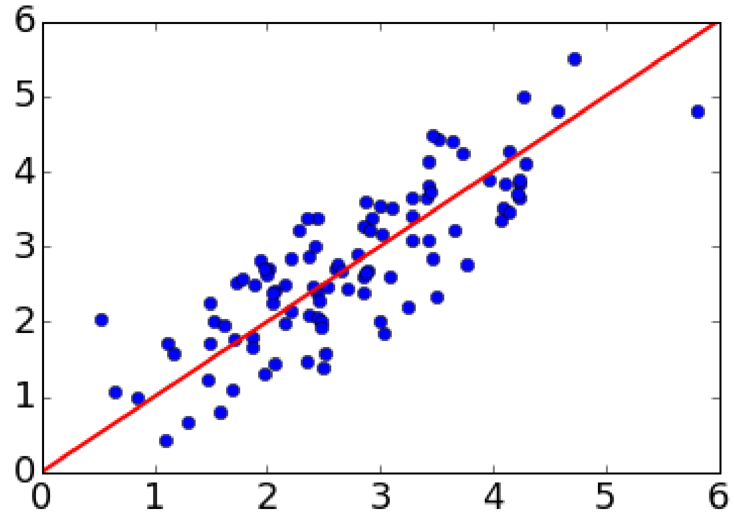


This is the **direction of maximum variance**.

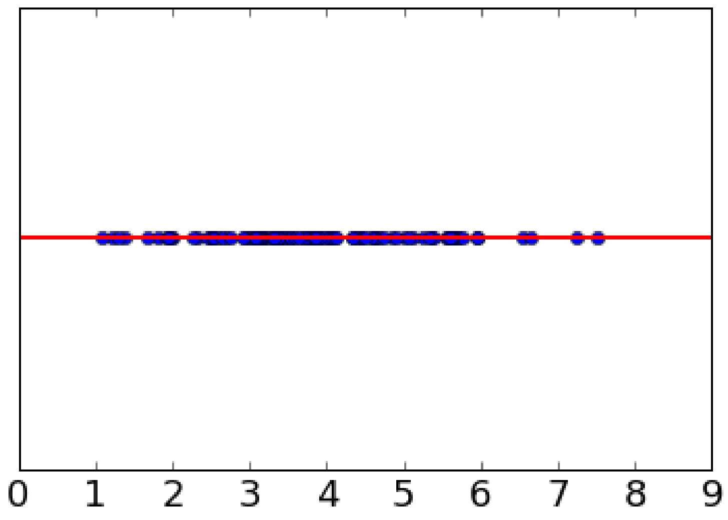
Reconstruction from 1D projection



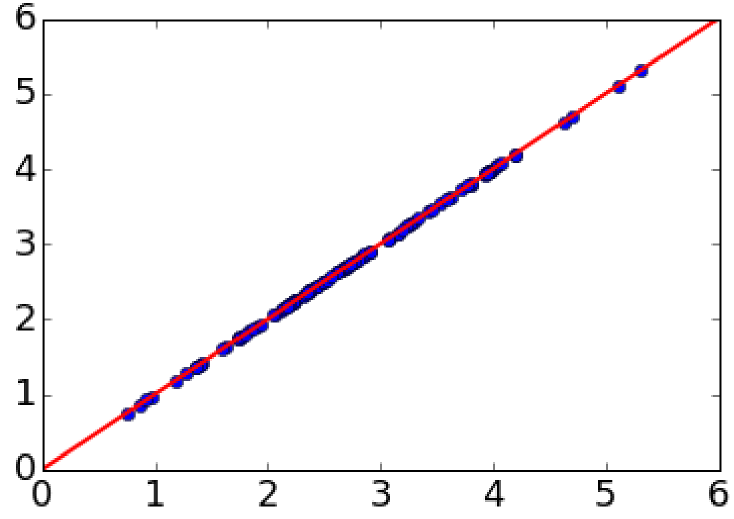
Two types of projection



Projection onto $\mathbb{R}$:

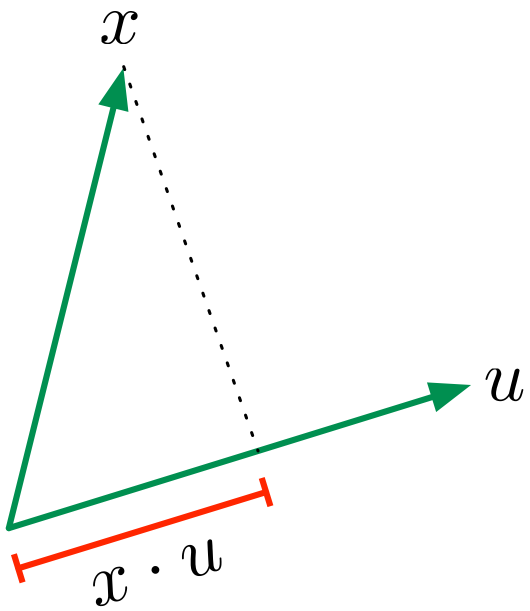


Projection onto a 1-d line in $\mathbb{R}^2$:



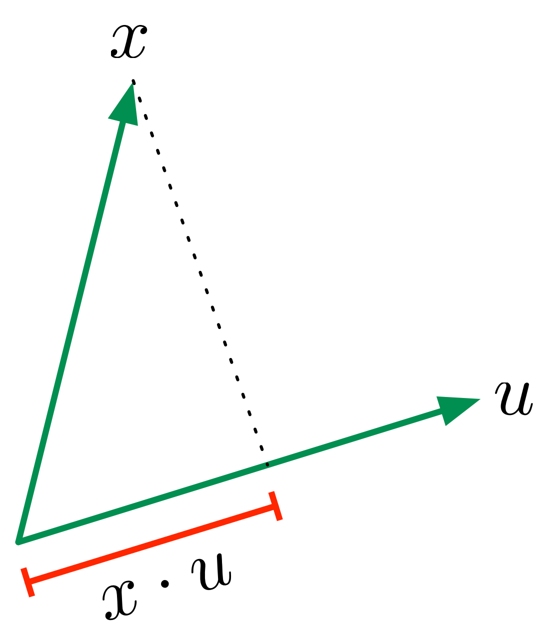
Projection: formally

What is the projection of $x \in \mathbb{R}^p$ onto direction $u \in \mathbb{R}^p$ (where $\|u\| = 1$)?



Projection: formally

What is the projection of $x \in \mathbb{R}^p$ onto direction $u \in \mathbb{R}^p$ (where $\|u\| = 1$)?

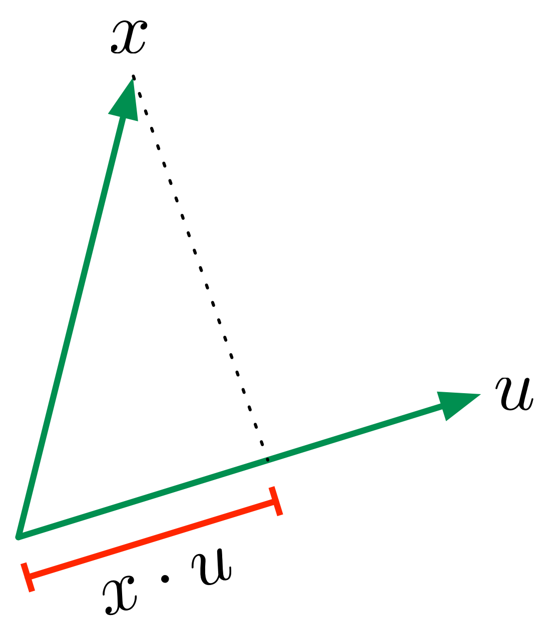


As a one-dimensional value:

$$x \cdot u = u \cdot x = u^T x = \sum_{i=1}^{p} u_i x_i.$$

Projection: formally

What is the projection of $x \in \mathbb{R}^p$ onto direction $u \in \mathbb{R}^p$ (where $\|u\| = 1$)?



As a one-dimensional value:

$$x \cdot u = u \cdot x = u^T x = \sum_{i=1}^{p} u_i x_i .$$

As a p-dimensional vector:

$$(x \cdot u)u = uu^T x$$

"Move $x \cdot u$ units in direction u"

Matrix notation

The compact way for representing projections and rotations

Inner Products

Row vector $\mathbf{a} = (a_1, a_2 \ldots, a_n)$

Column vector $\mathbf{b}^T = \begin{pmatrix} b_1 \\ b_2 \\ . \\ . \\ . \\ b_n \end{pmatrix}$

$$\mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^{n} a_i b_i = \mathbf{a}\mathbf{b}^T = (a_1, a_2 \ldots, a_n) \begin{pmatrix} b_1 \\ b_2 \\ . \\ . \\ . \\ b_n \end{pmatrix}$$

Orthogonality and norm

$\mathbf{a} \cdot \mathbf{b} = 0$ a and b are orthogonal vectors

Norm $\|\mathbf{a}\|_2^2 \doteq \mathbf{a} \cdot \mathbf{a} = \sum_{i=1} a_i^2$

Unit vector $\|\mathbf{a}\|_2 = 1$

matrix-vector product

$$A = \begin{pmatrix} a_{1,1}, a_{1,2}, \ldots, a_{1,m} \\ a_{2,1}, a_{2,2}, \ldots, a_{2,m} \\ \ldots \\ \ldots \\ a_{n,1}, a_{n,2}, \ldots, a_{n,m} \end{pmatrix} = \begin{pmatrix} \mathbf{a_1} \\ \mathbf{a_2} \\ . \\ . \\ . \\ \mathbf{a_n} \end{pmatrix}$$

$$A\mathbf{b}^T = \begin{pmatrix} \mathbf{a_1}\mathbf{b}^T \\ \mathbf{a_2}\mathbf{b}^T \\ \ldots \\ \ldots \\ \mathbf{a_n}\mathbf{b}^T \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^m a_{1,i}b_i \\ \sum_{i=1}^m a_{2,i}b_i \\ \ldots \\ \ldots \\ \sum_{i=1}^m a_{n,i}b_i \end{pmatrix}$$

matrix-matrix product

$$A = \begin{pmatrix} a_{1,1}, a_{1,2}, \ldots, a_{1,m} \\ a_{2,1}, a_{2,2}, \ldots, a_{2,m} \\ \ldots \\ \ldots \\ a_{n,1}, a_{n,2}, \ldots, a_{n,m} \end{pmatrix} = \begin{pmatrix} \mathbf{a_1} \\ \mathbf{a_2} \\ . \\ . \\ . \\ \mathbf{a_n} \end{pmatrix}$$

$$B = \begin{pmatrix} b_{1,1}, b_{1,2}, \ldots, b_{1,l} \\ b_{2,1}, b_{2,2}, \ldots, b_{2,l} \\ \ldots \\ \ldots \\ b_{m,1}, b_{m,2}, \ldots, b_{m,l} \end{pmatrix} = \left(\mathbf{b_1}^T, \mathbf{b_2}^T, \ldots, \mathbf{b_l}^T\right)$$

$$AB = \begin{pmatrix} (\mathbf{a_1} \cdot \mathbf{b_1}), (\mathbf{a_1} \cdot \mathbf{b_2}), \ldots, (\mathbf{a_1} \cdot \mathbf{b_l}) \\ (\mathbf{a_2} \cdot \mathbf{b_1}), (\mathbf{a_2} \cdot \mathbf{b_2}), \ldots, (\mathbf{a_2} \cdot \mathbf{b_l}) \\ \ldots \\ \ldots \\ (\mathbf{a_n} \cdot \mathbf{b_1}), (\mathbf{a_n} \cdot \mathbf{b_2}), \ldots, (\mathbf{a_n} \cdot \mathbf{b_l}) \end{pmatrix}$$

diagonal matrices

$$D\mathbf{b}^T = \begin{pmatrix} \lambda_1, 0, 0, \ldots, 0 \\ 0, \lambda_2, 0, \ldots, 0 \\ 0, 0, \lambda_3, \ldots, 0 \\ \ldots \\ \ldots \\ 0, 0, 0, \ldots, \lambda_n \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ . \\ . \\ . \\ b_n \end{pmatrix} = \begin{pmatrix} \lambda_1 b_1 \\ \lambda_2 b_2 \\ . \\ . \\ . \\ \lambda_n b_n \end{pmatrix}$$

$$I\mathbf{b}^T = \begin{pmatrix} 1, 0, 0, \ldots, 0 \\ 0, 1, 0, \ldots, 0 \\ 0, 0, 1, \ldots, 0 \\ \ldots \\ \ldots \\ 0, 0, 0, \ldots, 1 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ . \\ . \\ . \\ b_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ . \\ . \\ . \\ b_n \end{pmatrix}$$

Orthonormal Matrices

A is a square matrix (nxn)

$$A = \begin{pmatrix} \mathbf{a_1} \\ \mathbf{a_2} \\ . \\ . \\ . \\ \mathbf{a_n} \end{pmatrix} \qquad A^T = \left(\mathbf{a_1}^T, \mathbf{a_2}^T, \ldots, \mathbf{a_n}^T\right)$$

$$AA^T = I$$

The rows of ***A*** define an orthonormal basis

$$\forall i \neq j,\ \mathbf{a_i} \cdot \mathbf{a_i} = \|\mathbf{a}\|_2^2 = 1,\ \mathbf{a_i} \cdot \mathbf{a_j} = 0$$

Orthonormal matrices and vector bases

Orthonormal matrices and vector bases

- A set of n orthogonal unit vectors in R^n defines an orthonormal basis.

Orthonormal matrices and vector bases

- A set of n orthogonal unit vectors in R^n defines an orthonormal basis.
- Multiplying a vector by an orthonormal matrix corresponds to expressing it in terms of the orthonormal basis.

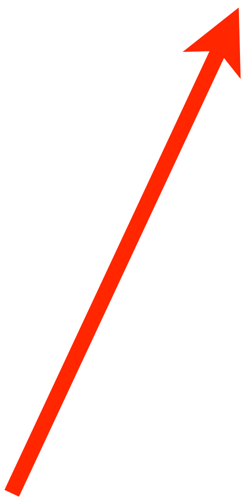
Changing basis in R^2

Changing basis in R^2

Orthonormal basis = coordinate system

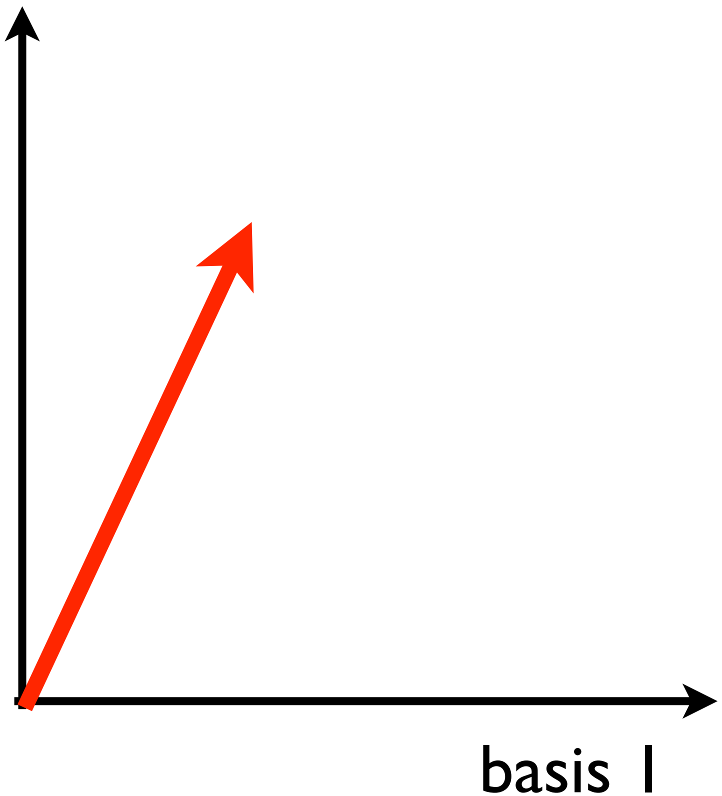
Changing basis in R^2

Orthonormal basis = coordinate system



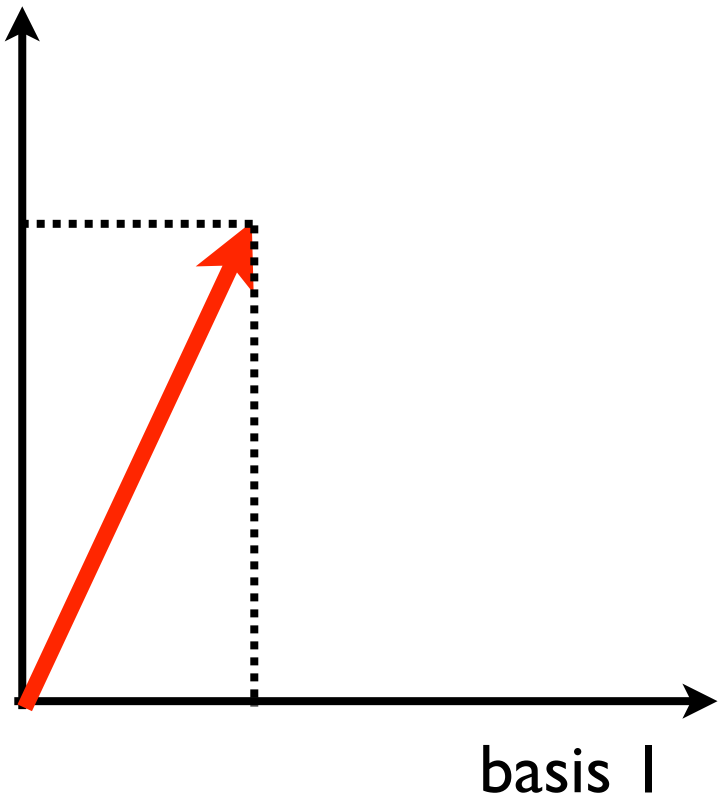
Changing basis in R^2

Orthonormal basis = coordinate system



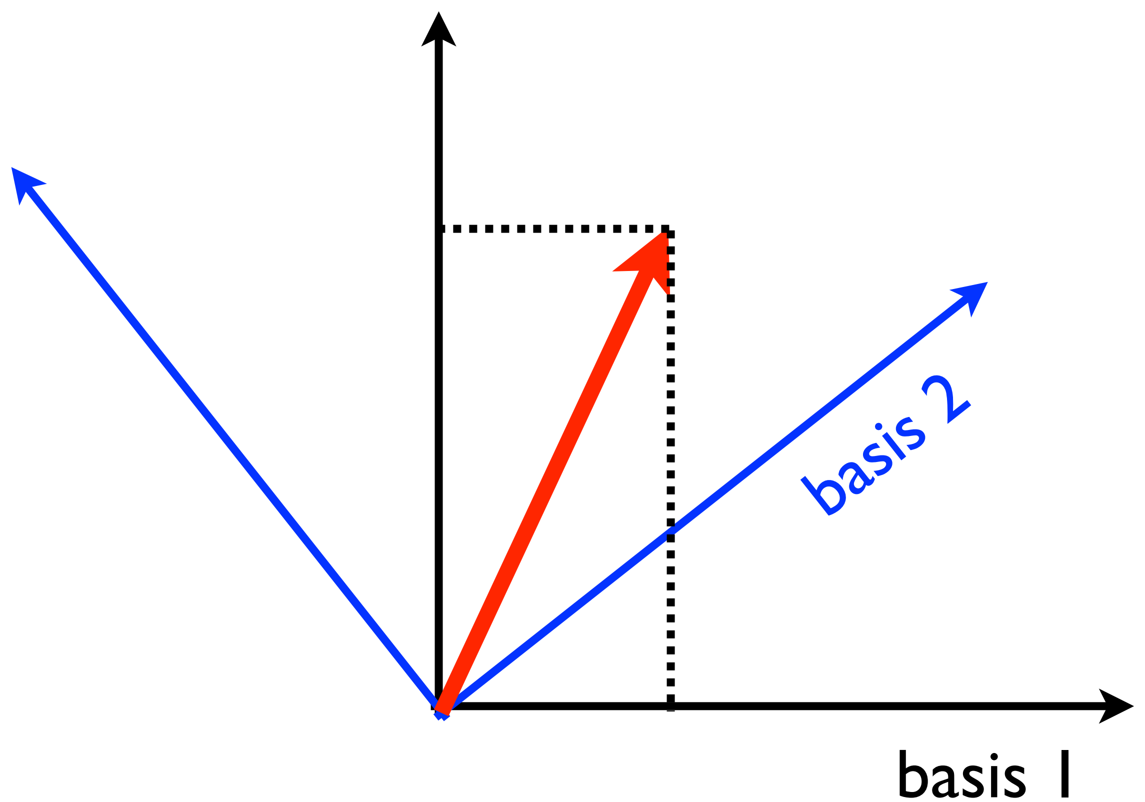
Changing basis in R^2

Orthonormal basis = coordinate system



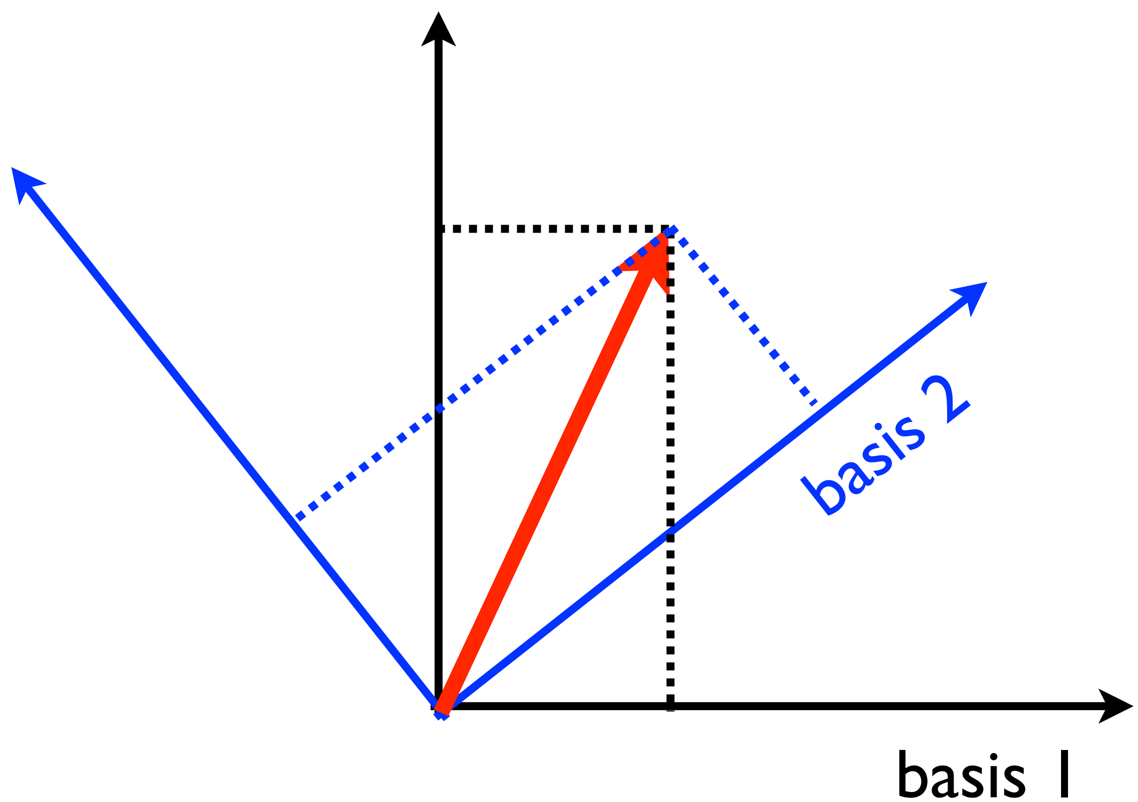
Changing basis in R^2

Orthonormal basis = coordinate system



Changing basis in R^2

Orthonormal basis = coordinate system



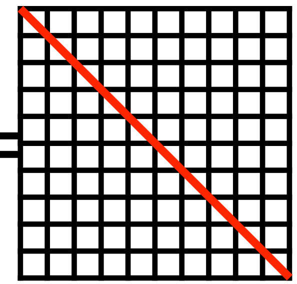
Eigenvectors and Eigenvalues

the vector **a** is an eigenvector of the matrix **M** with eigenvalue λ if

$$\mathbf{M}\mathbf{a} = \lambda\mathbf{a}$$

In words: the application of **M** to **a** amounts to changing the length of **a** by a factor of λ without changing **a**'s direction

Decomposing Symmetric Matrices

A symmetric matrix **M**= 

M can be written in the form

$$M = A^T \begin{pmatrix} \lambda_1, 0, 0, \ldots, 0 \\ 0, \lambda_2, 0, \ldots, 0 \\ 0, 0, \lambda_3, \ldots, 0 \\ \ldots \\ \ldots \\ 0, 0, 0, \ldots, \lambda_n \end{pmatrix} A$$

A is an orthonormal matrix consisting of the eigenvectors of **M**

Interpretation of the decomposition

Interpretation of the decomposition

1. the operation $A^T D A \mathbf{a}$ equals:

Interpretation of the decomposition

1. the operation $A^T D A \mathbf{a}$ equals:
2. transform $\mathbf{a}$ into basis defined by A

Interpretation of the decomposition

1. the operation $A^T D A \mathbf{a}$ equals:
2. transform $\mathbf{a}$ into basis defined by A
3. multiply each coordinate i by λ_i

Interpretation of the decomposition

1. the operation $A^T D A \mathbf{a}$ equals:
2. transform $\mathbf{a}$ into basis defined by A
3. multiply each coordinate i by λ_i
4. transform back to original basis

The covariance matrix

- A symmetric matrix that captures the pair-wise relations between observations.

The observations matrix

Suppose our data consists of n p-dimensional vectors. For example, for weather data, each observation can be a 365 dimensional vector. p=365

$$\begin{array}{cc} & \text{p variables} \\ \mathbf{X} = & \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} \text{ n observations} \end{array}$$

Subtracting the average observation vector

Subtracting the average observation vector

$$\text{observation matrix: } \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} = \begin{pmatrix} \vec{o}_1 \\ \vdots \\ \vec{o}_n \end{pmatrix}$$

Subtracting the average observation vector

$$\text{observation matrix: } \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} = \begin{pmatrix} \vec{o}_1 \\ \vdots \\ \vec{o}_n \end{pmatrix}$$

$$\text{The average observation vector: } \vec{\mu} = \begin{pmatrix} \mu_1 & \cdots & \mu_p \end{pmatrix} = \frac{1}{n}\sum_{i=1}^{n} \vec{o}_i = \frac{1}{n}\sum_{i=1}^{n} \begin{pmatrix} x_{i1} & \cdots & x_{ip} \end{pmatrix}$$

Subtracting the average observation vector

observation matrix: $$\begin{pmatrix} x_{11} & \dots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} = \begin{pmatrix} \vec{o}_1 \\ \vdots \\ \vec{o}_n \end{pmatrix}$$

The average observation vector: $\vec{\mu} = \begin{pmatrix} \mu_1 & \cdots & \mu_p \end{pmatrix} = \frac{1}{n}\sum_{i=1}^{n} \vec{o}_i = \frac{1}{n}\sum_{i=1}^{n} \begin{pmatrix} x_{i1} & \cdots & x_{ip} \end{pmatrix}$

Average corrected observation matrix: $$\begin{pmatrix} x_{11}-\mu_1 & \dots & x_{1p}-\mu_p \\ \vdots & \ddots & \vdots \\ x_{n1}-\mu_1 & \cdots & x_{np}-\mu_p \end{pmatrix} = \begin{pmatrix} \vec{o}_1-\vec{\mu} \\ \vdots \\ \vec{o}_n-\vec{\mu} \end{pmatrix}$$

Subtracting the average observation vector

observation matrix: $\begin{pmatrix} x_{11} & \dots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} = \begin{pmatrix} \vec{o}_1 \\ \vdots \\ \vec{o}_n \end{pmatrix}$

The average observation vector: $\vec{\mu} = \begin{pmatrix} \mu_1 & \cdots & \mu_p \end{pmatrix} = \frac{1}{n}\sum_{i=1}^{n} \vec{o}_i = \frac{1}{n}\sum_{i=1}^{n} \begin{pmatrix} x_{i1} & \cdots & x_{ip} \end{pmatrix}$

Average corrected observation matrix: $\begin{pmatrix} x_{11}-\mu_1 & \dots & x_{1p}-\mu_p \\ \vdots & \ddots & \vdots \\ x_{n1}-\mu_1 & \cdots & x_{np}-\mu_p \end{pmatrix} = \begin{pmatrix} \vec{o}_1 - \vec{\mu} \\ \vdots \\ \vec{o}_n - \vec{\mu} \end{pmatrix}$

The mean vector for the average corrected observation matrix is the zero vector.

self outer product

Row vector $\mathbf{a} = (a_1, a_2 \ldots, a_n)$

Self Inner Product

$$\mathbf{a} \cdot \mathbf{a} = \mathbf{a}\mathbf{a}^T = \sum_{i=1}^{n} a_i^2 \text{ is a scalar}$$

Self Outer product

$$\mathbf{a} \otimes \mathbf{a} = \mathbf{a}^T \mathbf{a} = \begin{pmatrix} a_1 a_1 & \ldots & a_n a_1 \\ \vdots & \ddots & \vdots \\ a_n a_1 & \cdots & a_n a_n \end{pmatrix} \text{ is an } n \times n \text{ matrix}$$

The covariance matrix

$$\mathbf{Cov}(\mathbf{X}) \doteq \frac{1}{n}\sum_{i=1}^{n} \left(\vec{x}_i - \vec{\mu}\right) \otimes \left(\vec{x}_i - \vec{\mu}\right)$$

As each outer product yields a symmetric matrix the Covariance matrix is also symmetric.

We can apply the orthonormal decomposition.

i.e. $\mathrm{Cov}(x) = A^T D A$

Interpretation of principle components

Interpretation of principle components

- distribution of random vector $\mathbf{x}$ defines covariance matrix $\text{cov}(\mathbf{x})$

Interpretation of principle components

- distribution of random vector $\mathbf{x}$ defines covariance matrix $\mathrm{cov}(\mathbf{x})$
- decomposition $\mathrm{cov}(\mathbf{x}) = A^T D A$ changes the coordinate system of $\mathbf{x}$

Interpretation of principle components

- distribution of random vector $\mathbf{x}$ defines covariance matrix $\mathrm{cov}(\mathbf{x})$
- decomposition $\mathrm{cov}(\mathbf{x}) = A^T D A$ changes the coordinate system of $\mathbf{x}$
- defines a new random vector $\mathbf{y} = A\mathbf{x}$

Interpretation of principle components

- distribution of random vector $\mathbf{x}$ defines covariance matrix $\mathrm{cov}(\mathbf{x})$
- decomposition $\mathrm{cov}(\mathbf{x}) = A^T D A$ changes the coordinate system of $\mathbf{x}$
- defines a new random vector $\mathbf{y}=A\mathbf{x}$
- $\mathrm{cov}(\mathbf{y}) = D$

Interpretation of principle components

- distribution of random vector $\mathbf{x}$ defines covariance matrix $\mathrm{cov}(\mathbf{x})$
- decomposition $\mathrm{cov}(\mathbf{x}) = A^T D A$ changes the coordinate system of $\mathbf{x}$
- defines a new random vector $\mathbf{y}=A\mathbf{x}$
- $\mathrm{cov}(\mathbf{y}) = D$
- $\mathrm{var}(y_i)= \lambda_i$, $\forall i\neq j$: $\mathrm{cov}(y_i,y_j)=0$

The best k-dimensional projection

Let Σ be the $p \times p$ covariance matrix of X. Its **eigendecomposition** can be computed in $O(p^3)$ time and consists of:

- real **eigenvalues** $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$
- corresponding **eigenvectors** $u_1, \ldots, u_p \in \mathbb{R}^p$ that are orthonormal: that is, each u_i has unit length and $u_i \cdot u_j = 0$ whenever $i \neq j$.

The best k-dimensional projection

Let Σ be the $p \times p$ covariance matrix of X. Its **eigendecomposition** can be computed in $O(p^3)$ time and consists of:

- real **eigenvalues** $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$
- corresponding **eigenvectors** $u_1, \ldots, u_p \in \mathbb{R}^p$ that are orthonormal: that is, each u_i has unit length and $u_i \cdot u_j = 0$ whenever $i \neq j$.

Theorem: Suppose we want to map data $X \in \mathbb{R}^p$ to just k dimensions, while capturing as much of the variance of X as possible. The best choice of projection is:

$$x \mapsto (u_1 \cdot x, u_2 \cdot x, \ldots, u_k \cdot x),$$

where u_i are the eigenvectors described above.

The best k-dimensional projection

Let Σ be the $p \times p$ covariance matrix of X. Its **eigendecomposition** can be computed in $O(p^3)$ time and consists of:

- real **eigenvalues** $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$
- corresponding **eigenvectors** $u_1, \ldots, u_p \in \mathbb{R}^p$ that are orthonormal: that is, each u_i has unit length and $u_i \cdot u_j = 0$ whenever $i \neq j$.

Theorem: Suppose we want to map data $X \in \mathbb{R}^p$ to just k dimensions, while capturing as much of the variance of X as possible. The best choice of projection is:

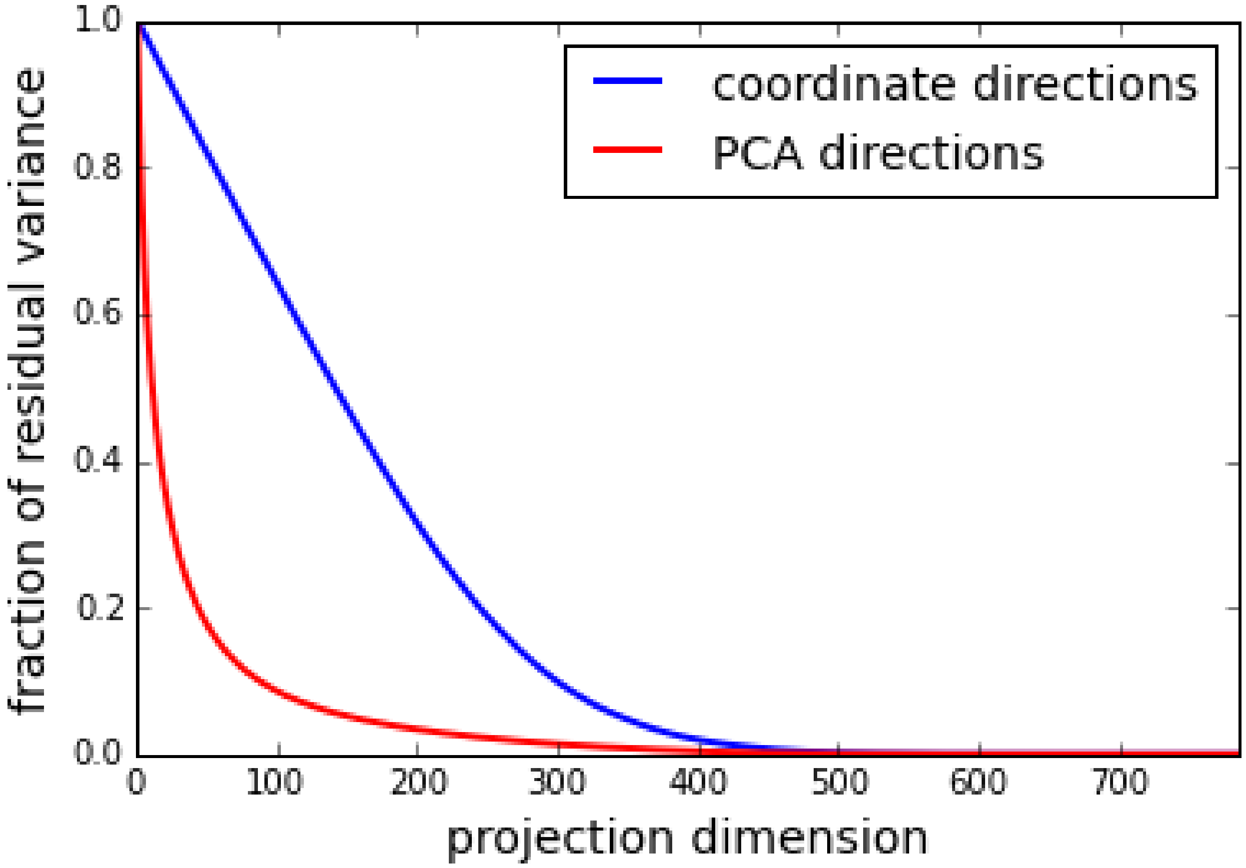
$$x \mapsto (u_1 \cdot x, u_2 \cdot x, \ldots, u_k \cdot x),$$

where u_i are the eigenvectors described above.

Projecting the data in this way is **principal component analysis** (PCA).

Example: MNIST

Contrast coordinate projections with PCA:



MNIST: image reconstruction

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

MNIST: image reconstruction

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

$k = 200$

$k = 150$

$k = 100$

$k = 50$

MNIST: image reconstruction

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

$k = 200$

$k = 150$

$k = 100$

$k = 50$

Q: What are these reconstructions exactly?

MNIST: image reconstruction

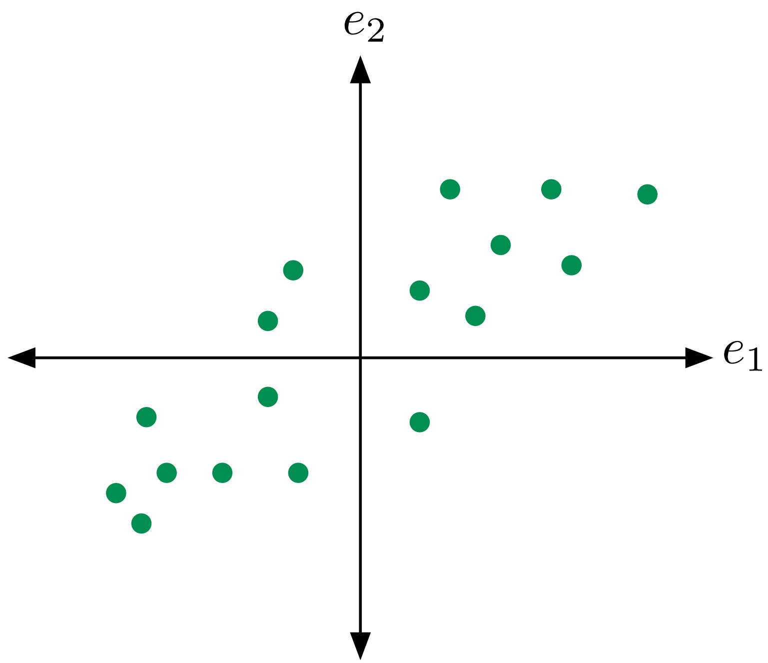
Reconstruct this original image from its PCA projection to k dimensions.

$k = 200$ $k = 150$ $k = 100$ $k = 50$

Q: What are these reconstructions exactly?
A: Image x is reconstructed as $UU^T x$, where U is a $p \times k$ matrix whose columns are the top k eigenvectors of Σ.

Principal component analysis: recap

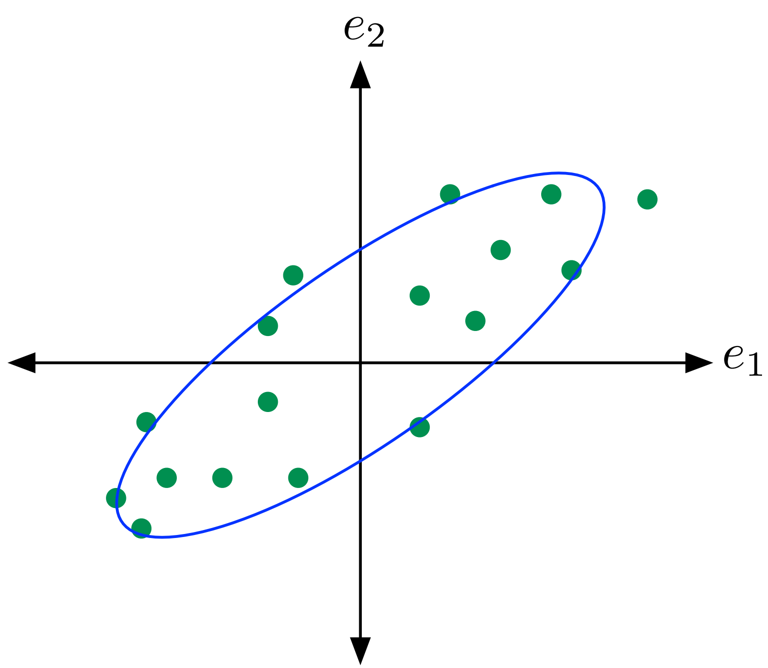
Consider data vectors $X \in \mathbb{R}^p$.



Principal component analysis: recap

Consider data vectors $X \in \mathbb{R}^p$.

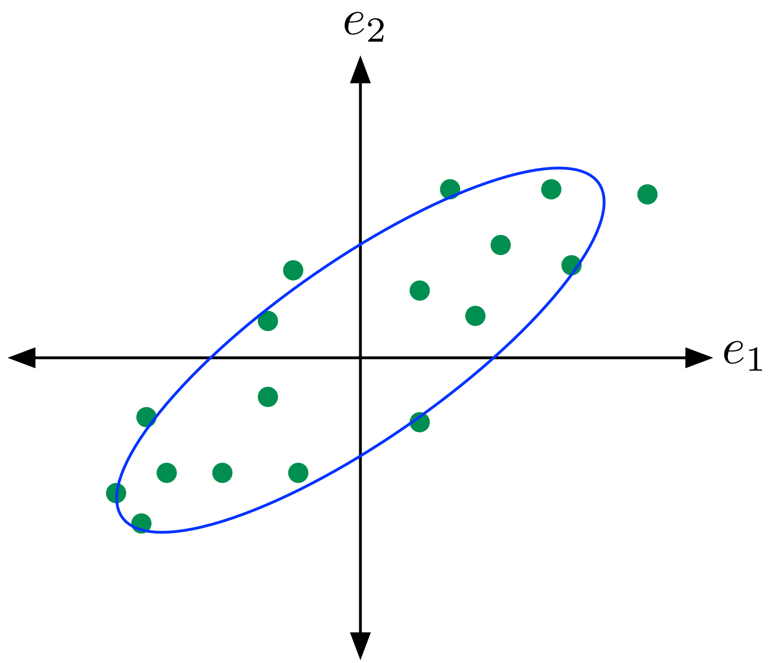
- The covariance matrix Σ is a $p \times p$ symmetric matrix.



Principal component analysis: recap

Consider data vectors $X \in \mathbb{R}^p$.

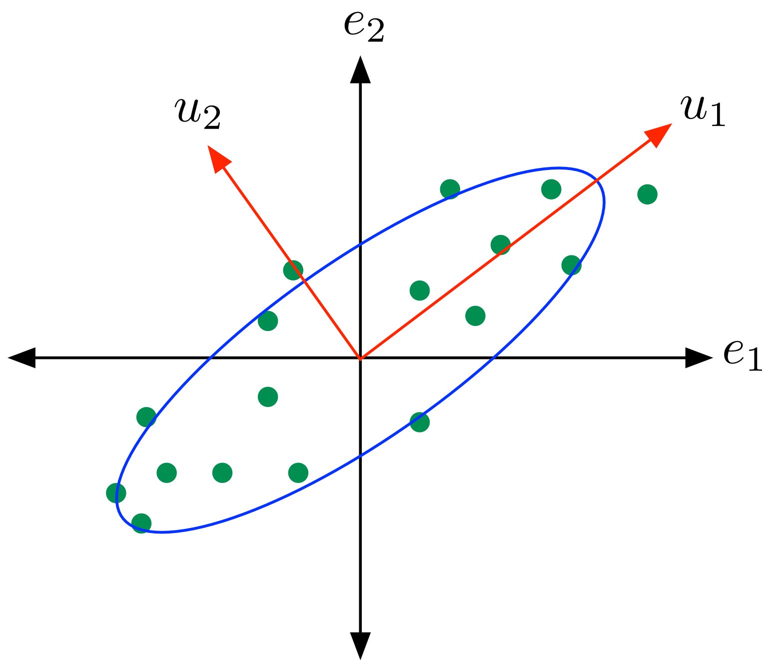
- The covariance matrix Σ is a $p \times p$ symmetric matrix.
- Get eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$, eigenvectors $u_1, \ldots, u_p$.



Principal component analysis: recap

Consider data vectors $X \in \mathbb{R}^p$.

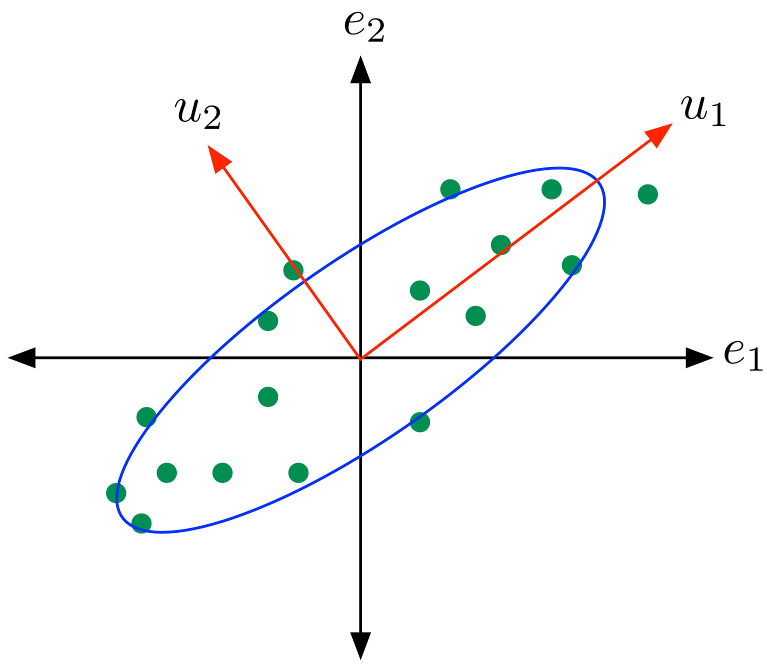
- The covariance matrix Σ is a $p \times p$ symmetric matrix.
- Get eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$, eigenvectors $u_1, \ldots, u_p$.
- $u_1, \ldots, u_p$ is an alternative basis in which to represent the data.
- The variance of X in direction u_i is λ_i.



Principal component analysis: recap

Consider data vectors $X \in \mathbb{R}^p$.

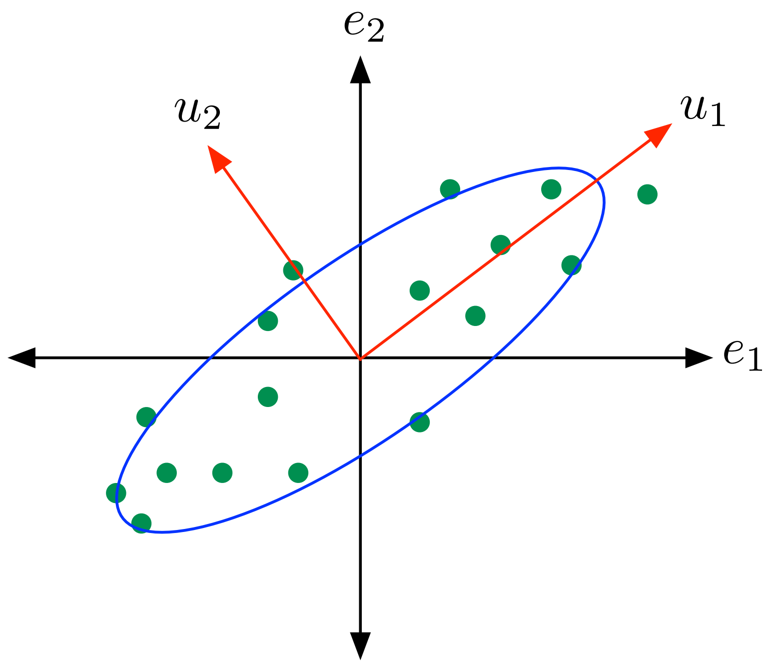
- The covariance matrix Σ is a $p \times p$ symmetric matrix.
- Get eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$, eigenvectors $u_1, \ldots, u_p$.
- $u_1, \ldots, u_p$ is an alternative basis in which to represent the data.
- The variance of X in direction u_i is λ_i.
- To project to k dimensions while losing as little as possible of the overall variance, use $x \mapsto (x \cdot u_1, \ldots, x \cdot u_k)$.



Principal component analysis: recap

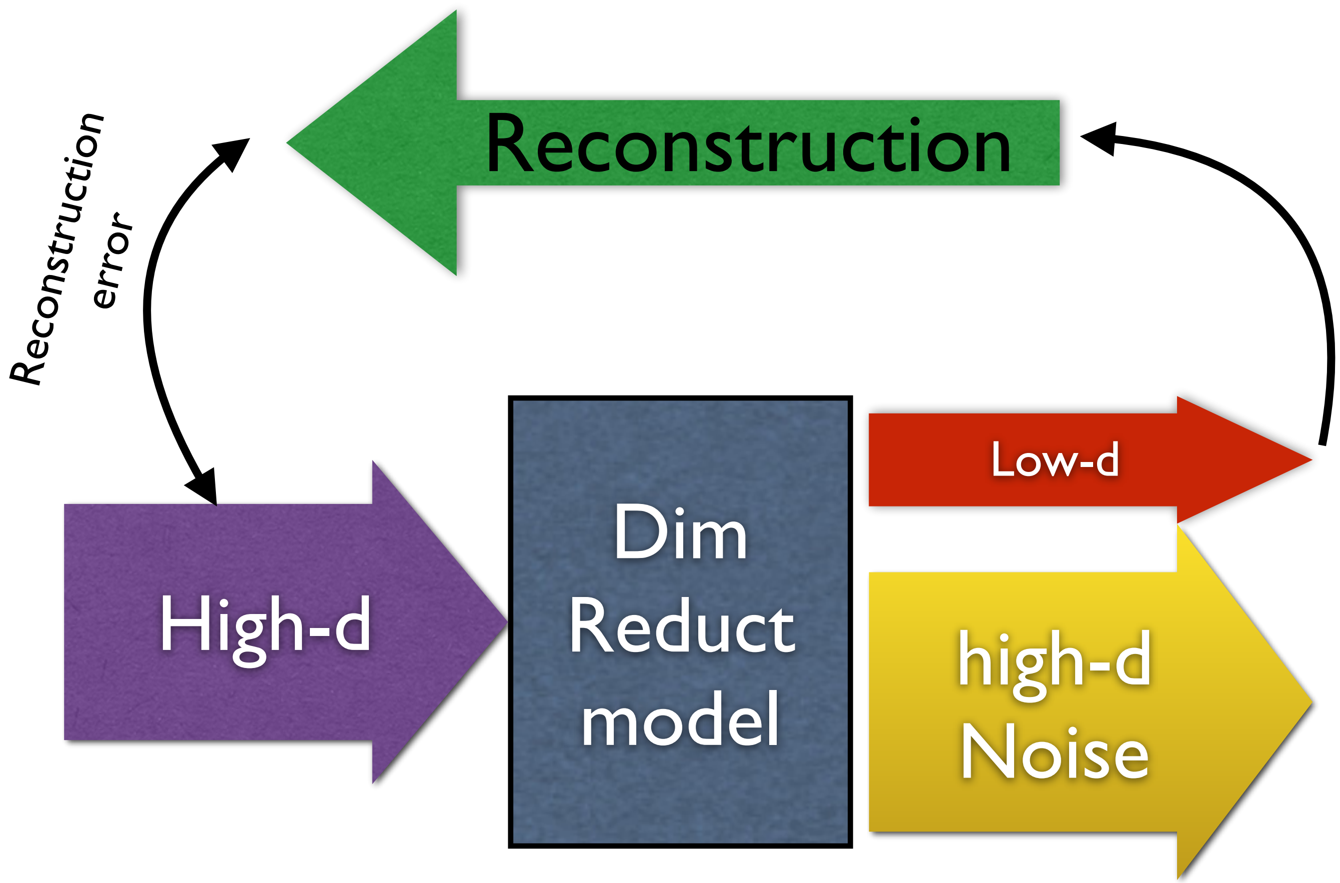
Consider data vectors $X \in \mathbb{R}^p$.

- The covariance matrix Σ is a $p \times p$ symmetric matrix.
- Get eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$, eigenvectors $u_1, \ldots, u_p$.
- $u_1, \ldots, u_p$ is an alternative basis in which to represent the data.
- The variance of X in direction u_i is λ_i.
- To project to k dimensions while losing as little as possible of the overall variance, use $x \mapsto (x \cdot u_1, \ldots, x \cdot u_k)$.



What is the covariance of the projected data?

dimensionality reduction



Example: personality assessment

What are the dimensions along which personalities differ?

Example: personality assessment

What are the dimensions along which personalities differ?

- *Lexical hypothesis*: most important personality characteristics have become encoded in natural language.

Example: personality assessment

What are the dimensions along which personalities differ?

- *Lexical hypothesis*: most important personality characteristics have become encoded in natural language.
- Allport and Odbert (1936): sat down with the English dictionary and extracted all terms that could be used to distinguish one person's behavior from another's. Roughly 18000 words, of which 4500 could be described as personality traits.

Example: personality assessment

What are the dimensions along which personalities differ?

- *Lexical hypothesis*: most important personality characteristics have become encoded in natural language.
- Allport and Odbert (1936): sat down with the English dictionary and extracted all terms that could be used to distinguish one person's behavior from another's. Roughly 18000 words, of which 4500 could be described as personality traits.
- Step: group these words into (approximate) synonyms. This is done by manual clustering. e.g. Norman (1967):

Spirit	Jolly, merry, witty, lively, peppy
Talkativeness	Talkative, articulate, verbose, gossipy
Sociability	Companionable, social, outgoing
Spontaneity	Impulsive, carefree, playful, zany
Boisterousness	Mischievous, rowdy, loud, prankish
Adventure	Brave, venturous, fearless, reckless
Energy	Active, assertive, dominant, energetic
Conceit	Boastful, conceited, egotistical
Vanity	Affected, vain, chic, dapper, jaunty
Indiscretion	Nosey, snoopy, indiscreet, meddlesome
Sensuality	Sexy, passionate, sensual, flirtatious

Example: personality assessment

What are the dimensions along which personalities differ?

- *Lexical hypothesis*: most important personality characteristics have become encoded in natural language.
- Allport and Odbert (1936): sat down with the English dictionary and extracted all terms that could be used to distinguish one person's behavior from another's. Roughly 18000 words, of which 4500 could be described as personality traits.
- Step: group these words into (approximate) synonyms. This is done by manual clustering. e.g. Norman (1967):

Spirit	Jolly, merry, witty, lively, peppy
Talkativeness	Talkative, articulate, verbose, gossipy
Sociability	Companionable, social, outgoing
Spontaneity	Impulsive, carefree, playful, zany
Boisterousness	Mischievous, rowdy, loud, prankish
Adventure	Brave, venturous, fearless, reckless
Energy	Active, assertive, dominant, energetic
Conceit	Boastful, conceited, egotistical
Vanity	Affected, vain, chic, dapper, jaunty
Indiscretion	Nosey, snoopy, indiscreet, meddlesome
Sensuality	Sexy, passionate, sensual, flirtatious

- Data collection: Ask a variety of subjects to what extent each of these words describes them.

Personality assessment: the data

Matrix of data (1 = strongly disagree, 5 = strongly agree)

	shy	merry	tense	boastful	forgiving	quiet
Person 1	4	1	1	2	5	5
Person 2	1	4	4	5	2	1
Person 3	2	4	5	4	2	2
		⋮				

How to extract important directions?

Personality assessment: the data

Matrix of data (1 = strongly disagree, 5 = strongly agree)

	shy	merry	tense	boastful	forgiving	quiet
Person 1	4	1	1	2	5	5
Person 2	1	4	4	5	2	1
Person 3	2	4	5	4	2	2
		$\vdots$				

How to extract important directions?

- Treat each column as a data point, find tight clusters
- Treat each row as a data point, apply PCA
- Other ideas: factor analysis, independent component analysis, ...

Personality assessment: the data

Matrix of data (1 = strongly disagree, 5 = strongly agree)

	shy	merry	tense	boastful	forgiving	quiet
Person 1	4	1	1	2	5	5
Person 2	1	4	4	5	2	1
Person 3	2	4	5	4	2	2
		⋮				

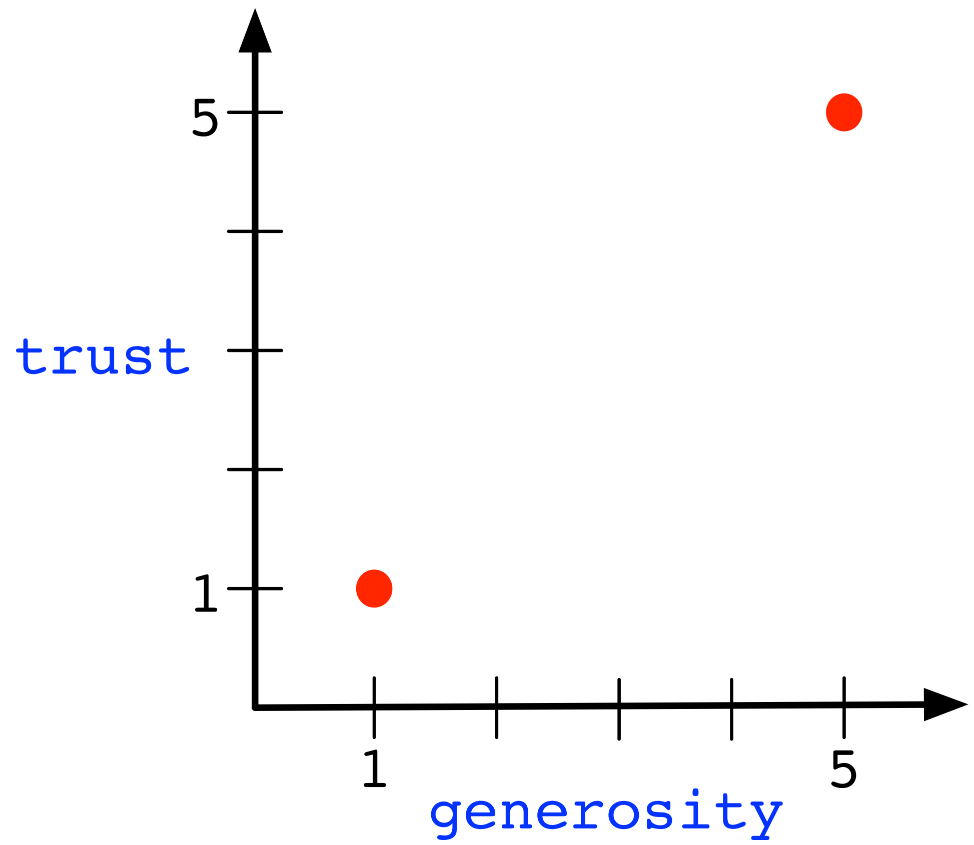
How to extract important directions?

- Treat each column as a data point, find tight clusters
- Treat each row as a data point, apply PCA
- Other ideas: factor analysis, independent component analysis, ...

Many of these yield similar results

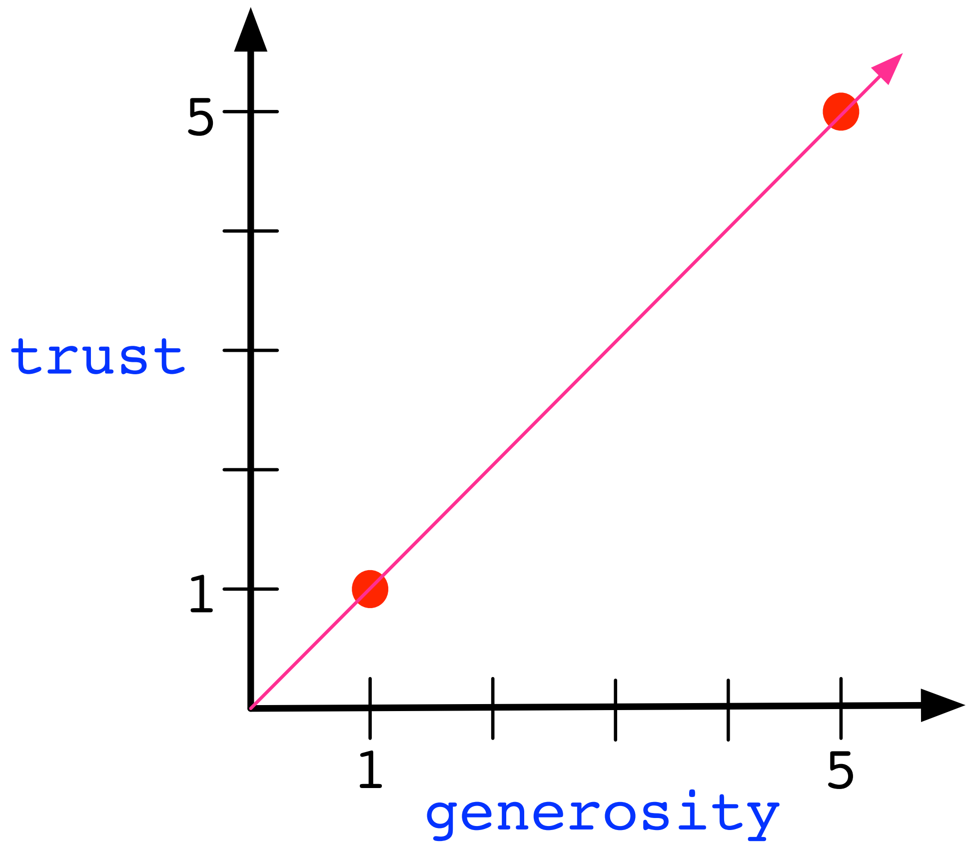
What does PCA accomplish?

Example: suppose two traits (generosity, trust) are highly correlated, to the point where each person either answers "1" to both or "5" to both.



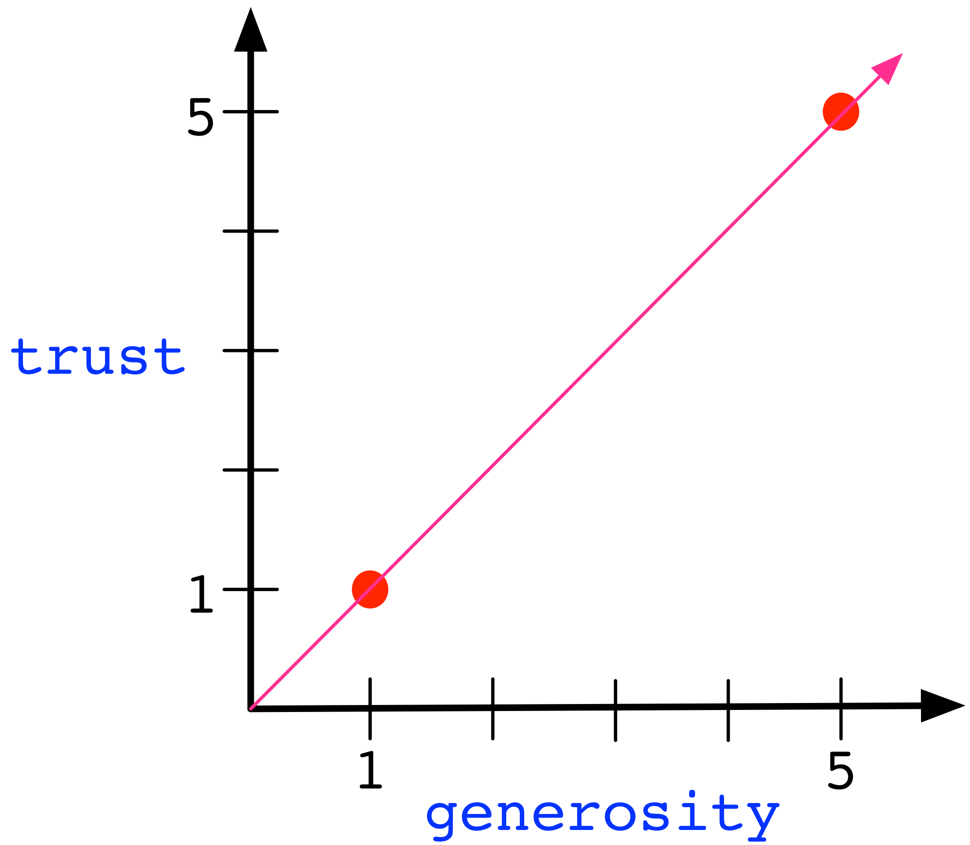
What does PCA accomplish?

Example: suppose two traits (generosity, trust) are highly correlated, to the point where each person either answers "1" to both or "5" to both.



What does PCA accomplish?

Example: suppose two traits (generosity, trust) are highly correlated, to the point where each person either answers "1" to both or "5" to both.



This single PCA dimension entirely accounts for the two traits.

The “Big Five” taxonomy

Extraversion		Agreeableness		Conscientiousness		Neuroticism		Openness/Intellect	
Low	High	Low	High	Low	High	Low	High	Low	High
-.83 Quiet	.85 Talkative	-.52 Fault-finding	.87 Sympathetic	-.58 Careless	.80 Organized	-.39 Stable*	.73 Tense	-.74 Commonplace	.76 Wide interests
-.80 Reserved	.83 Assertive	-.48 Cold	.85 Kind	-.53 Disorderly	.80 Thorough	-.35 Calm*	.72 Anxious	-.73 Narrow interests	.76 Imaginative
-.75 Shy	.82 Active	-.45 Unfriendly	.85 Appreciative	-.50 Frivolous	.78 Planful	-.21 Contented*	.72 Nervous	-.67 Simple	.72 Intelligent
-.71 Silent	.82 Energetic	-.45 Quarrelsome	.84 Affectionate	-.49 Irresponsible	.78 Efficient	.14 Unemotional*	.71 Moody	-.55 Shallow	.73 Original
-.67 Withdrawn	.82 Outgoing	-.45 Hard-hearted	.84 Soft-hearted	-.40 Slipshot	.73 Responsible		.71 Worrying	-.47 Unintelligent	.68 Insightful
-.66 Retiring	.80 Outspoken	-.38 Unkind	.82 Warm	-.39 Undependable	.72 Reliable		.68 Touchy		.64 Curious
	.79 Dominant	-.33 Cruel	.81 Generous	-.37 Forgetful	.70 Dependable		.64 Fearful		.59 Sophisticated
	.73 Forceful	-.31 Stern*	.78 Trusting		.68 Conscientious		.63 High-strung		.59 Artistic
	.73 Enthusiastic	-.28 Thankless	.77 Helpful		.66 Precise		.63 Self-pitying		.59 Clever
	.68 Show-off	-.24 Stingy*	.77 Forgiving		.66 Practical		.60 Temperamental		.58 Inventive
	.68 Sociable		.74 Pleasant		.65 Deliberate		.59 Unstable		.56 Sharp-witted
	.64 Spunky		.73 Good-natured		.46 Painstaking		.58 Self-punishing		.55 Ingenious
	.64 Adventurous		.73 Friendly		.26 Cautious*		.54 Despondent		.45 Witty*
	.62 Noisy		.72 Cooperative				.51 Emotional		.45 Resourceful*
	.58 Bossy		.67 Gentle						.37 Wise
			.66 Unselfish						.33 Logical*
			.56 Praising						.29 Civilized*
			.51 Sensitive						.22 Foresighted*
									.21 Polished*
									.20 Dignified*

Many applications, such as online match-making.

Extraversion	
Low	High
-.83 Quiet	.85 Talkative
-.80 Reserved	.83 Assertive
-.75 Shy	.82 Active
-.71 Silent	.82 Energetic
-.67 Withdrawn	.82 Outgoing
-.66 Retiring	.80 Outspoken
	.79 Dominant
	.73 Forceful
	.73 Enthusiastic
	.68 Show-off
	.68 Sociable
	.64 Spunky
	.64 Adventurous
	.62 Noisy
	.58 Bossy

Agreeableness	
Low	High
-.52 Fault-finding	.87 Sympathetic
-.48 Cold	.85 Kind
-.45 Unfriendly	.85 Appreciative
-.45 Quarrelsome	.84 Affectionate
-.45 Hard-hearted	.84 Soft-hearted
-.38 Unkind	.82 Warm
-.33 Cruel	.81 Generous
-.31 Stern*	.78 Trusting
-.28 Thankless	.77 Helpful
-.24 Stingy*	.77 Forgiving
	.74 Pleasant
	.73 Good-natured
	.73 Friendly
	.72 Cooperative
	.67 Gentle
	.66 Unselfish
	.56 Praising
	.51 Sensitive

Conscientiousness

Low	High
-.58 Careless	.80 Organized
-.53 Disorderly	.80 Thorough
-.50 Frivolous	.78 Planful
-.49 Irresponsible	.78 Efficient
-.40 Slipshot	.73 Responsible
-.39 Undependable	.72 Reliable
-.37 Forgetful	.70 Dependable
	.68 Conscientious
	.66 Precise
	.66 Practical
	.65 Deliberate
	.46 Painstaking
	.26 Cautious*

Neuroticism

Low	High
-.39 Stable*	.73 Tense
-.35 Calm*	.72 Anxious
-.21 Contented*	.72 Nervous
.14 Unemotional*	.71 Moody
	.71 Worrying
	.68 Touchy
	.64 Fearful
	.63 High-strung
	.63 Self-pitying
	.60 Temperamental
	.59 Unstable
	.58 Self-punishing
	.54 Despondent
	.51 Emotional

Oppenness/Intellect

Low	High
-.74 Commonplace	.76 Wide interests
-.73 Narrow interests	.76 Imaginative
-.67 Simple	.72 Intelligent
-.55 Shallow	.73 Original
-.47 Unintelligent	.68 Insightful
	.64 Curious
	.59 Sophisticated
	.59 Artistic
	.59 Clever
	.58 Inventive
	.56 Sharp-witted
	.55 Ingenious
	.45 Witty*
	.45 Resourceful*
	.37 Wise
	.33 Logical*
	.29 Civilized*
	.22 Foresighted*
	.21 Polished*
	.20 Dignified*