Principal components analysis

Introduction

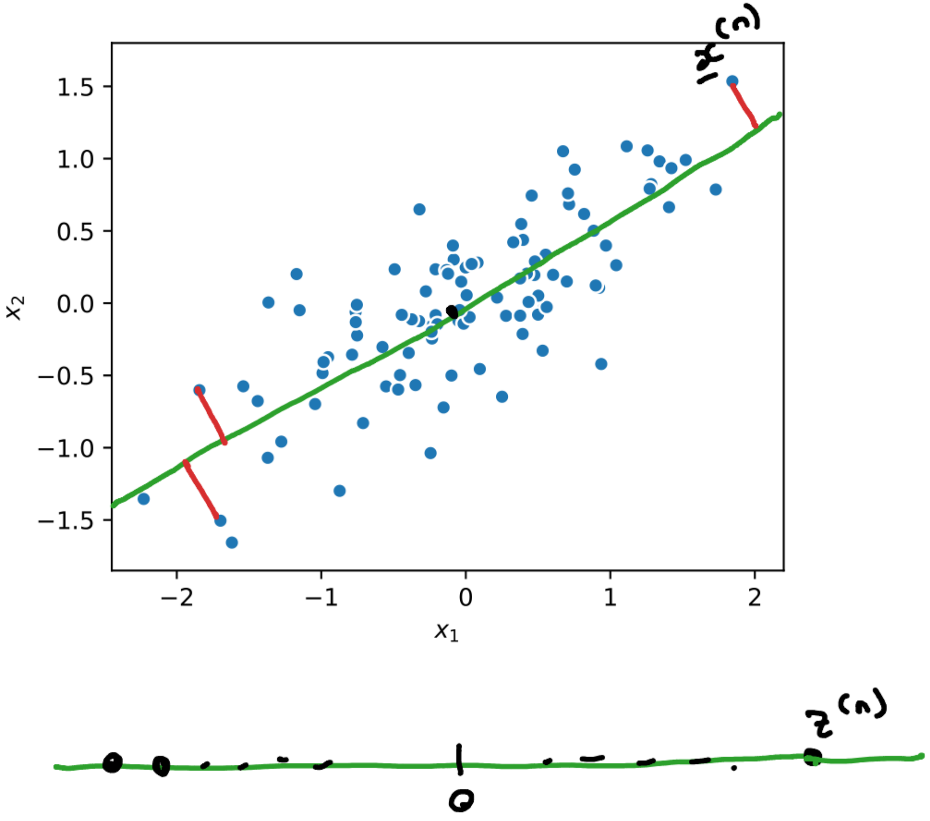
Herman Kamper

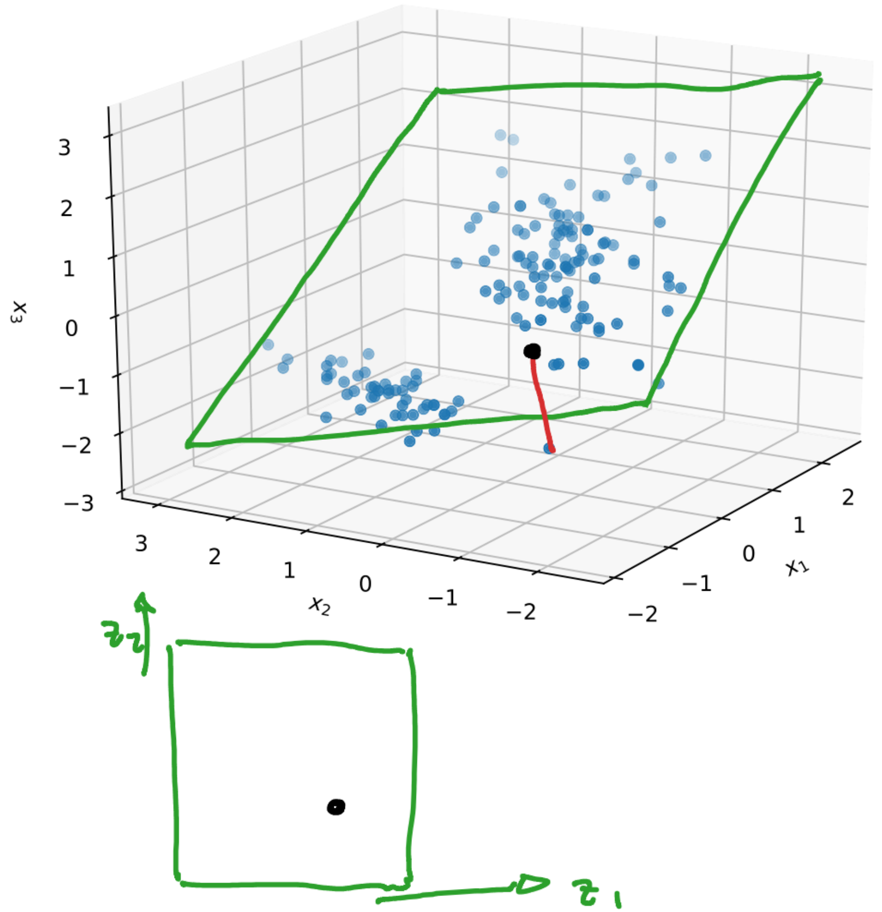
http://www.kamperh.com/

Linear projection

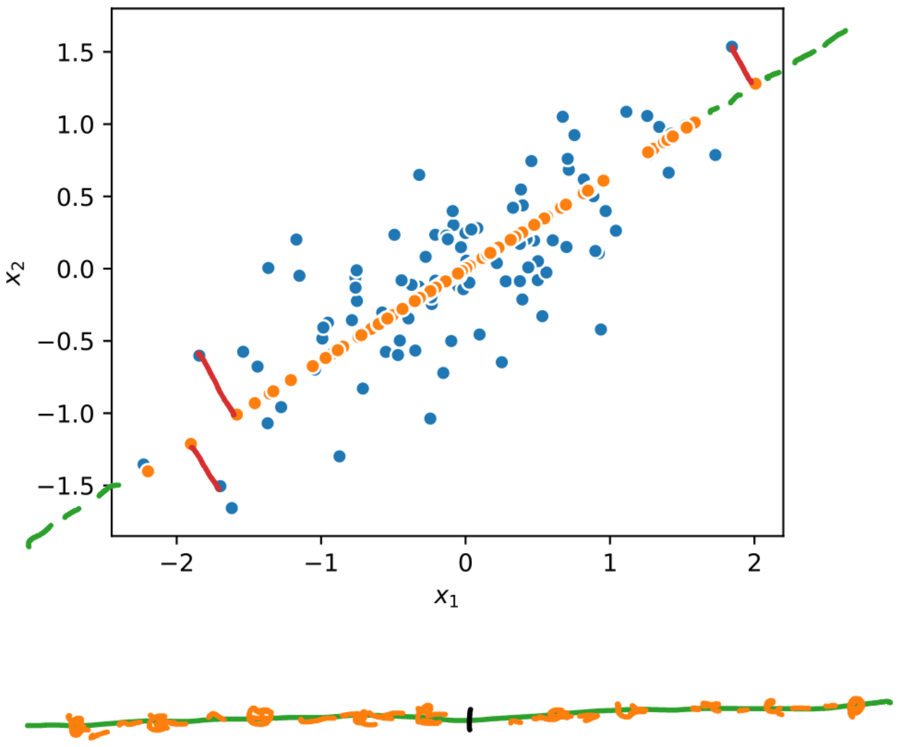
$\underline{x} \in \mathbb{R}^2 \longrightarrow z \in \mathbb{R}$

$\underline{x} \in \mathbb{R}^3 \longrightarrow \underline{z} \in \mathbb{R}^2$

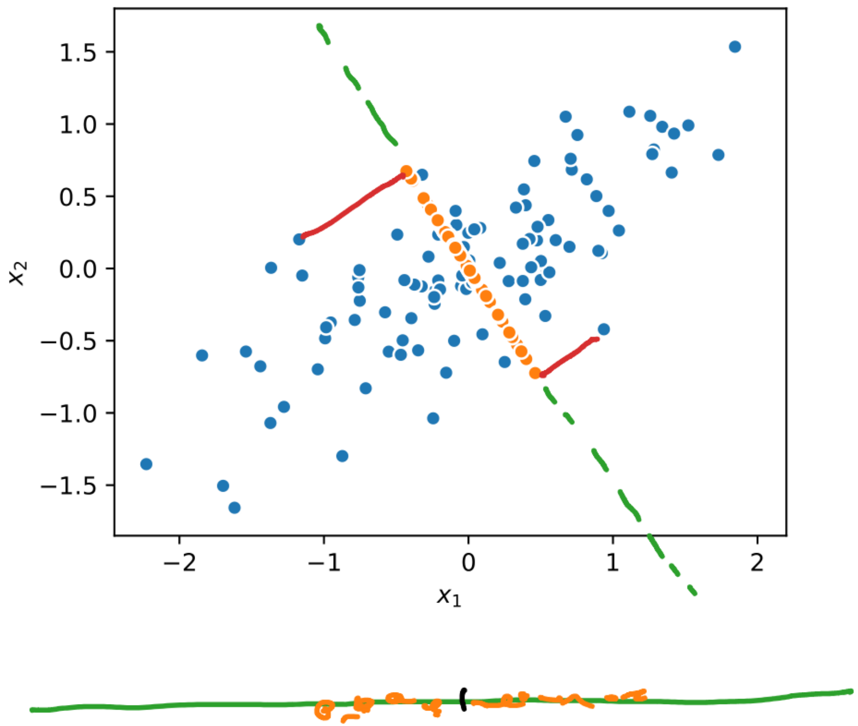




View 1: Maximising variance

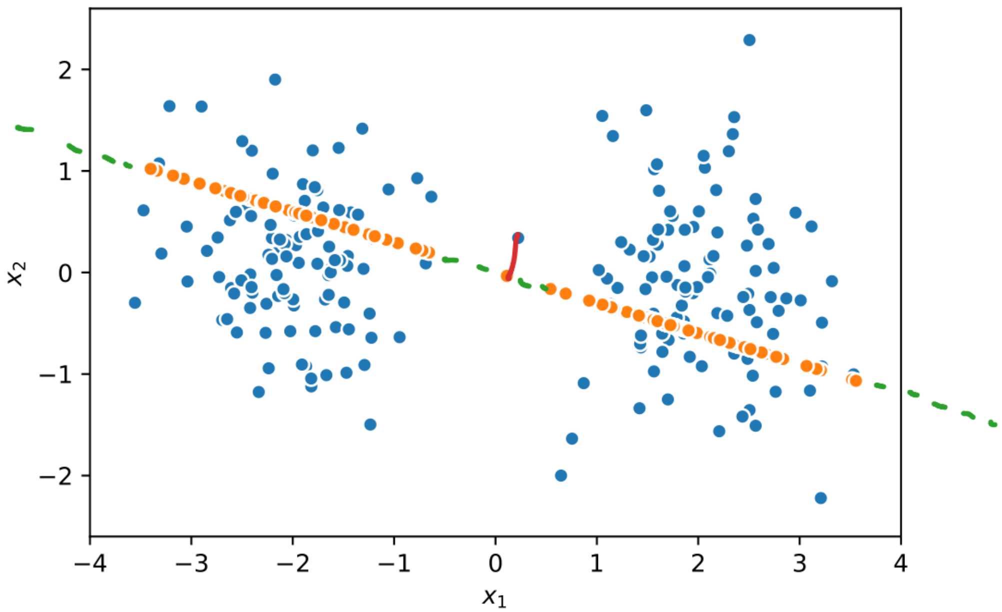


$\hat{\sigma}_z = 0.9493$

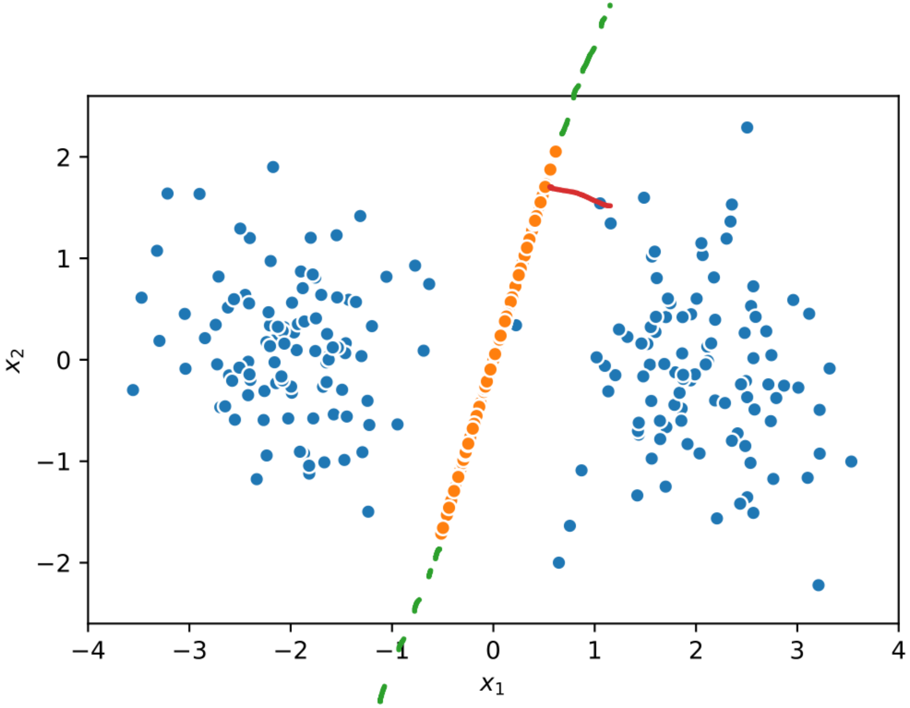


$\hat{\sigma}_z = 0.1017$

View 1: Maximising variance

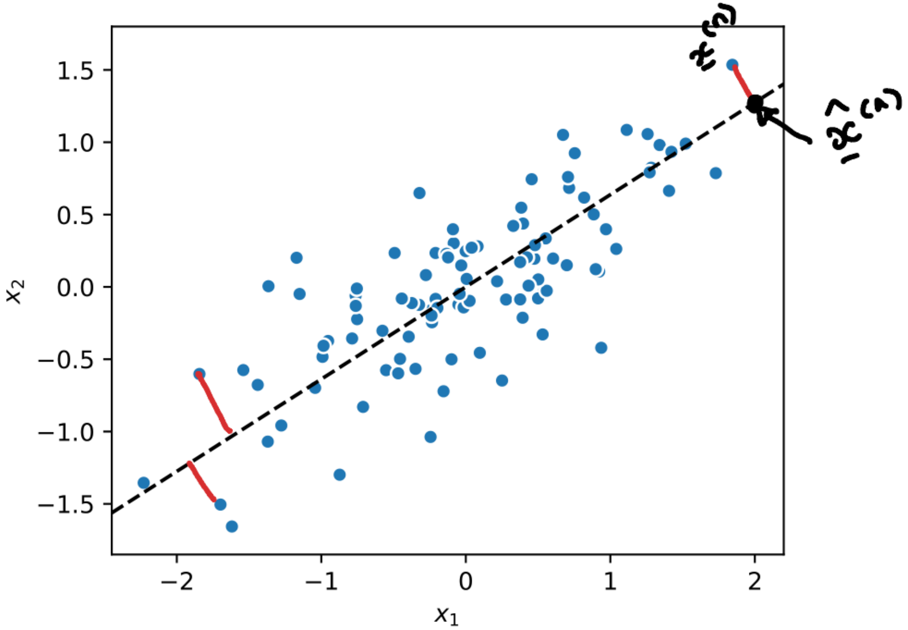


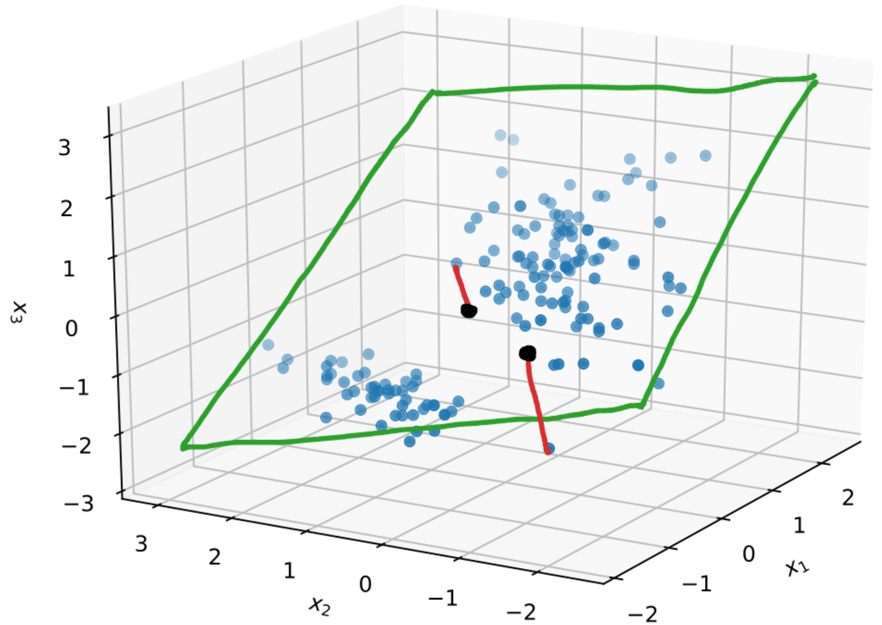
$\hat{\sigma}_z = 4.3187$



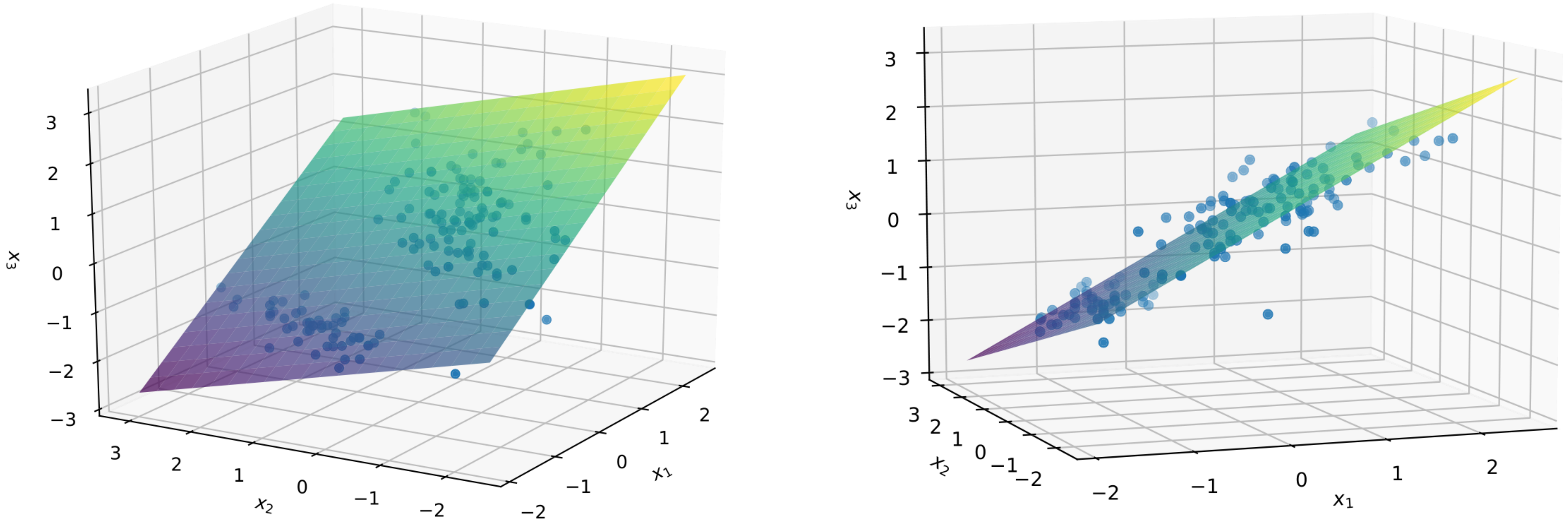
$\hat{\sigma}_z = 0.7331$

View 2: Minimising reconstruction error





View 2: Minimising reconstruction error



Principal components analysis

Mathematical background

Herman Kamper

http://www.kamper.com/

PCA: Mathematical background

Lagrange multipliers:

Want to optimise $f(x)$ subject to some constraint $g(x) = 0$.
Then we define a new objective:

$$J(x, \lambda) = f(x) + \lambda g(x)$$

and optimise w.r.t. both x and λ.

Eigenvalues and eigenvectors:

For a square matrix $\underline{A}$:

$$\underline{A}\,\underline{u} = \lambda \underline{u}$$

The solutions to this equation are pairs of eigenvalues (λ) with eigenvectors ($\underline{u}$)

Vector derivatives:

$$\frac{\partial f(\underline{x})}{\partial \underline{x}} \triangleq \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_D} \end{bmatrix}$$

D-dimensional

Identities:

- $$\frac{\partial \underline{x}^T \underline{A}\, \underline{x}}{\partial \underline{x}} = 2 \underline{A}\, \underline{x}$$
  if $\underline{A}$ is symmetrical
- $$\frac{\partial \underline{x}^T \underline{x}}{\partial \underline{x}} = 2 \underline{x}$$

(See "Matrix calculus" on Wikipedia.)

Principal components analysis

Setup

Herman Kamper

http://www.kamperh.com/

PCA : Setup

We want to project $\underline{x}^{(n)} \in \mathbb{R}^D$ to $\underline{z}^{(n)} \in \mathbb{R}^M$, with $M < D$.

(Normally assume data have been normalised to have zero-mean.)

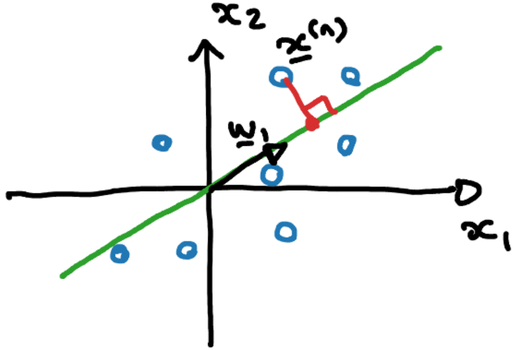
Use M "projection vectors" $\underline{w}_m \in \mathbb{R}^D$.

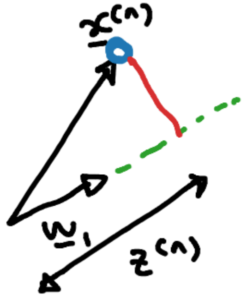
Projection vectors $\underline{w}_1, \dots, \underline{w}_M$ are unit length and orthogonal, i.e.

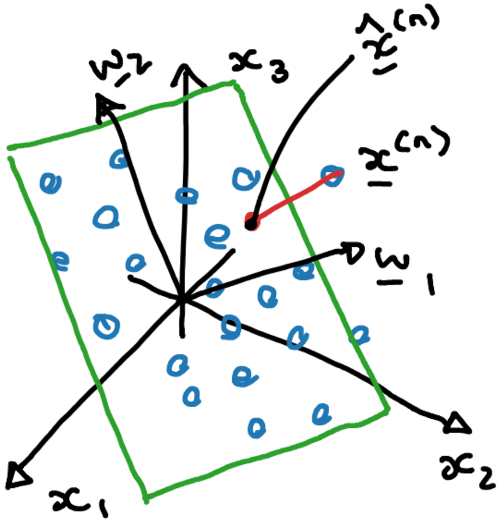
$\|\underline{w}_m\| = 1$ and $\underline{w}_i^T \underline{w}_j = 0 \quad \forall i \neq j$.

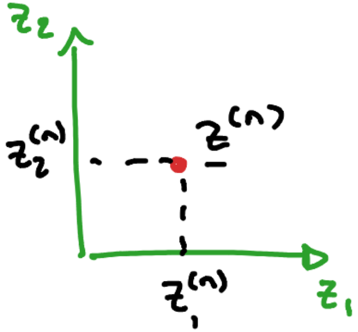
The projection of the n^{th} item $\underline{x}^{(n)}$ onto the m^{th} dimension is

$$z_m^{(n)} = \underline{w}_m^T \underline{x}^{(n)}$$









$$z_1^{(n)} = \underline{w}_1^T \underline{x}^{(n)}$$

$$z_2^{(n)} = \underline{w}_2^T \underline{x}^{(n)}$$

Projection:

So $\underline{x}^{(n)}$ is mapped to

$$\underset{M\times 1}{\underline{z}^{(n)}} = \begin{bmatrix} z_1^{(n)} \\ z_2^{(n)} \\ \vdots \\ z_M^{(n)} \end{bmatrix} = \begin{bmatrix} \underline{w}_1^T \underline{x}^{(n)} \\ \underline{w}_2^T \underline{x}^{(n)} \\ \vdots \\ \underline{w}_M^T \underline{x}^{(n)} \end{bmatrix}$$

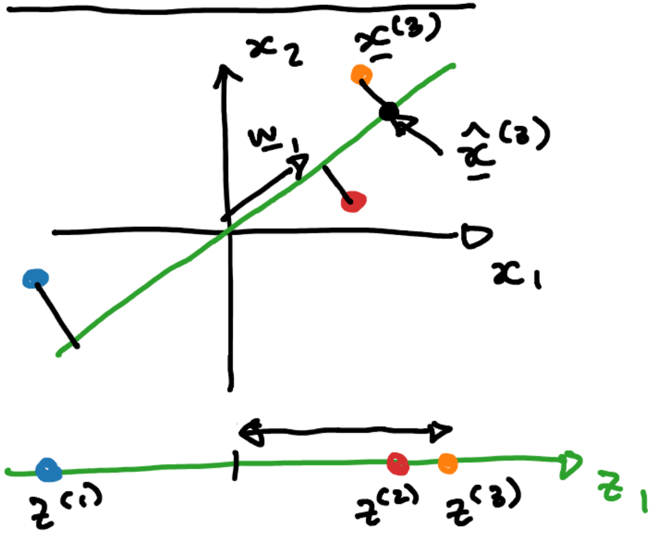
$$= \underset{M\times D}{\begin{bmatrix} — & \underline{w}_1^T & — \\ — & \underline{w}_2^T & — \\ & \vdots & \\ — & \underline{w}_M^T & — \end{bmatrix}} \underset{D\times 1}{\begin{bmatrix} x_1^{(n)} \\ x_2^{(n)} \\ \vdots \\ x_D^{(n)} \end{bmatrix}}$$

$$= \underbrace{\underline{W}}^{T} \underline{x}^{(n)}$$

$$\underline{W} = \begin{array}{c|c} & M \\ \hline D & \begin{bmatrix} | & | & & | \\ \underline{w}_1 & \underline{w}_2 & \cdots & \underline{w}_M \\ | & | & & | \end{bmatrix} \end{array}$$

$D\times M$

Reconstruction:



$$\hat{\underline{x}}^{(3)} = z^{(3)} \underline{w}_1 \qquad \hat{\underline{x}}^{(n)} \approx \underline{x}^{(n)}$$

In general: $\underset{D\times 1}{\hat{\underline{x}}^{(n)}} = \underset{D\times M}{\underline{W}} \; \underset{M\times 1}{\underline{z}^{(n)}}$

Principal components analysis

Finding the projection vectors

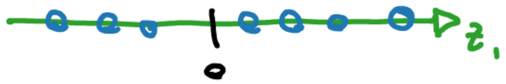
Herman Kamper

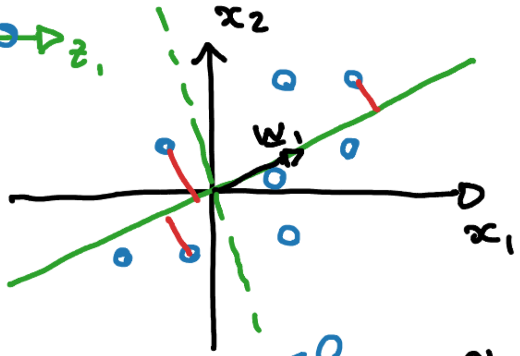
http://www.kamperh.com/

PCA: Learning the projection vectors

Setup:

- Data $\underline{x}^{(1)}, \dots, \underline{x}^{(N)}$ have been mean normalised (zero-mean)
- Want to find $\underline{w}_1, \dots, \underline{w}_M$
- $\|\underline{w}_m\| = 1 \quad \forall m$
- $\underline{w}_i^T \underline{w}_j = 0 \quad \forall i \neq j$





Problem:

Want to find $\underline{w}_1, \underline{w}_2, \dots, \underline{w}_M$ so that (sample) variance is maximised.

Let's first just look at one dimension.

$$\hat{\sigma}_{z_1}^2 = \frac{1}{N}\sum_{n=1}^{N}\Big(z_1^{(n)} - \underbrace{\bar{z}_1}_{=0}\Big)^2 = \frac{1}{N}\sum_{n=1}^{N}\left(z_1^{(n)}\right)^2$$

$$= \frac{1}{N}\sum_{n=1}^{N}\left(\underline{w}_1^T \underline{x}^{(n)}\right)^2$$

$$= \frac{1}{N}\sum_{n=1}^{N}\left(\underline{w}_1^T \underline{x}^{(n)}\right)\left(\underline{w}_1^T \underline{x}^{(n)}\right)^T$$

$$= \underline{w}_1^T \left[\frac{1}{N}\sum_{n=1}^{N}\underline{x}^{(n)}\left(\underline{x}^{(n)}\right)^T\right]\underline{w}_1$$

$$= \underline{w}_1^T \hat{\underline{\Sigma}}\, \underline{w}_1$$

Sample covariance matrix if $\underline{x}$ is zero-mean

Want to maximise $\hat{\sigma}_{z_1}^2$ subject to $\|\underline{w}_1\|^2 = 1$, i.e. $\underline{w}_1^T \underline{w}_1 = 1$

Loss (minimise)

Use Lagrange multiplier:

$$J(\underline{w}_1) = -\hat{\sigma}_{z_1}^2 + \lambda(\underline{w}_1^T \underline{w}_1 - 1)$$
$$= -\underline{w}_1^T \hat{\underline{\Sigma}} \underline{w}_1 + \lambda(\underline{w}_1^T \underline{w}_1 - 1)$$

Minimise w.r.t. $\underline{w}_1$:

$$\frac{\partial J(\underline{w}_1)}{\partial \underline{w}_1} = -\cancel{2}\hat{\underline{\Sigma}} \underline{w}_1 + \cancel{2}\lambda \underline{w}_1 = \underline{0}$$

$$\underbrace{\hat{\underline{\Sigma}} \underline{w}_1 = \lambda \underline{w}_1}_{\text{Eigenvalue / eigenvector equation}} \quad \ldots \text{(1)}$$

Which eigenvector/value do we use?

From (1): $\underline{w}_1^T \hat{\underline{\Sigma}} \underline{w}_1 = \lambda \underline{w}_1^T \underline{w}_1$

$$\underline{w}_1^T \hat{\underline{\Sigma}} \underline{w}_1 = \lambda$$

Want this maximised $\rightarrow$ $\hat{\sigma}_{z_1}^2 = \lambda$

So pick eigenvector corresponding to largest eigenvalue.

How do we find $\underline{w}_2$, with $\|\underline{w}_2\|^2 = 1$ and $\underline{w}_1^T \underline{w}_2 = 0$?

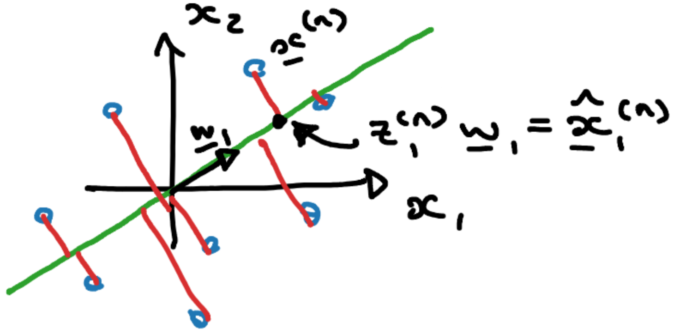
Repeat above steps:

$$\hat{\underline{\Sigma}} \underline{w}_2 = \lambda_2 \underline{w}_2$$

Pick eigenvector corresponding to 2nd highest eigenvalue, etc.

PCA: Another view

Instead of maximising variance, we think of PCA as minimising reconstruction loss:



$$J(\underline{w}_1) = \sum_{n=1}^{N} \| \underline{x}^{(n)} - \hat{\underline{x}}^{(n)} \|^2$$ [Looking at 1-dim. projection]

$$= \sum_{n=1}^{N} \| \underline{x}^{(n)} - z_1^{(n)} \underline{w}_1 \|^2 = \sum_{n=1}^{N} \overbrace{\left(\underline{x}^{(n)} - z_1^{(n)} \underline{w}_1 \right)^T}^{1 \times D} \overbrace{\left(\underline{x}^{(n)} - z_1^{(n)} \underline{w}_1 \right)}^{D \times 1}$$

$$= \sum_{n=1}^{N} \left[(\underline{x}^{(n)})^T \underline{x}^{(n)} - (\underline{x}^{(n)})^T z_1^{(n)} \underline{w}_1 - z_1^{(n)} \underline{w}_1^T \underline{x}^{(n)} + (z_1^{(n)})^2 \overbrace{\underline{w}_1^T \underline{w}_1}^{=1} \right]$$

$==$

$$= \sum_{n=1}^{N} \left[(\underline{x}^{(n)})^T \underline{x}^{(n)} - 2 z_1^{(n)} \overbrace{\underline{w}_1^T \underline{x}^{(n)}}^{z_1^{(n)}} + (z_1^{(n)})^2 \right]$$

Minimising reconstruction

==

Maximising variance

$$= \sum_{n=1}^{N} \left[(\underline{x}^{(n)})^T \underline{x}^{(n)} - (z_1^{(n)})^2 \right]$$

$$= C - \sum_{n=1}^{N} (z_1^{(n)})^2 = C - N \left[\frac{1}{N} \sum_{n=1}^{N} (z_1^{(n)})^2 \right] = C - N \hat{\sigma}_{z_1}^2$$

Principal components analysis

Relationship to singular value decomposition

Herman Kamper

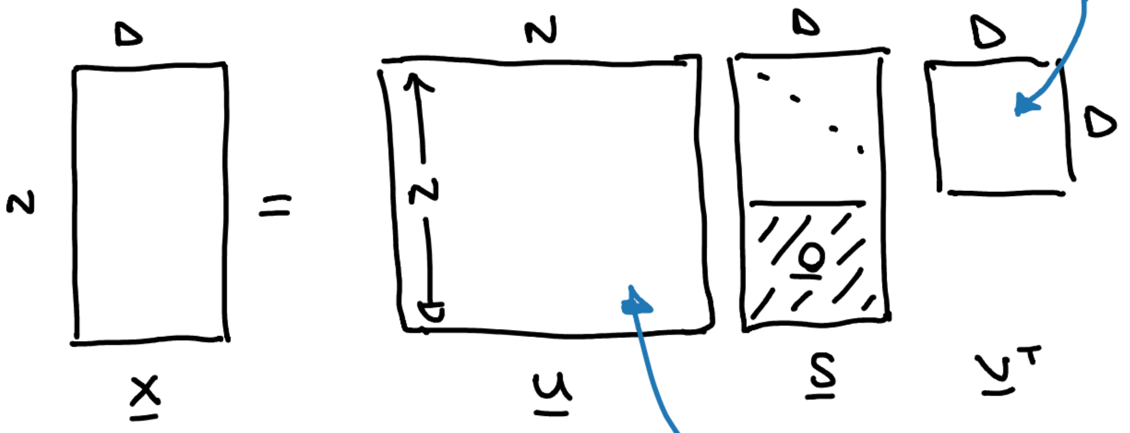
http://www.kamperh.com/

PCA: Relationship to SVD

Singular value decomposition:

$$\underset{N\times D}{\underline{X}} = \underset{N\times N}{\underline{U}}\;\underset{N\times D}{\underline{S}}\;\underset{D\times D}{\underline{V}^T}$$

Rows and columns orthonormal: $\underline{V}^T\underline{V} = \underline{I}$, $\underline{V}\underline{V}^T = \underline{I}$



Columns orthonormal: $\underline{U}^T\underline{U} = \underline{I}$

Relationship to PCA:

Take SVD of the design matrix $\underline{X}$:

$$\underline{X} = \underline{U}\,\underline{S}\,\underline{V}^T$$

Then $\underline{X}^T\underline{X} = \underline{V}\,\underline{S}^T\underline{U}^T\underline{U}\,\underline{S}\,\underline{V}^T$

$$= \underline{V}\,\underline{S}^T\underline{S}\,\underline{V}^T = \underline{V}\,\underline{D}\,\underline{V}^T$$

$\underline{D} = \underline{S}^T\underline{S}$ Diagonal with squares of singular values

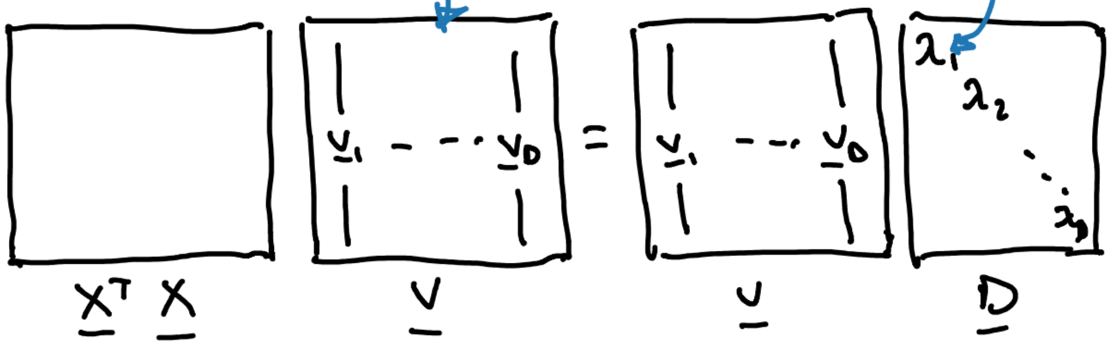
$$(\underline{X}^T\underline{X})\,\underline{V} = \underline{V}\,\underline{D}$$

$$\underline{\hat{\Sigma}} = \frac{1}{N}\sum_{n=1}^{N}\underline{x}^{(n)}(\underline{x}^{(n)})^T$$

$$= \frac{1}{N}\underline{X}^T\underline{X}$$

Eigenvectors

Eigenvalues



Principal components analysis

Steps

Herman Kamper

http://www.kamperh.com/

PCA steps

1 $\underline{x}^{(1)}, \underline{x}^{(2)}, \dots, \underline{x}^{(N)}$ where $\underline{x}^{(n)} \in \mathbb{R}^D$

1. Normalise the data to be zero-mean. $\bar{\underline{x}} = \underline{0}$

2. Calculate the sample covariance matrix. 2 $\hat{\underline{\Sigma}} = \frac{1}{N} \sum_{n=1}^{N} \underline{x}^{(n)} (\underline{x}^{(n)})^T = \frac{1}{N} \underline{X}^T \underline{X}$

3. Find the D eigenvector-eigenvalue pairs of the sample covariance matrix. $\underline{U}\,\underline{S}\,\underline{V}^T$

   3 Python: np.linalg.eig, np.linalg.svd — Matlab: eigs, svd

4. Choose the M eigenvectors corresponding to the highest eigenvalues. 4 $\underline{w}_1, \underline{w}_2, \dots, \underline{w}_M$

5. Project the data to the lower-dimensional space. 5 $\underline{z}^{(n)} = \underline{W}^T \underline{x}^{(n)}$ $\quad \underline{z}^{(n)} \in \mathbb{R}^M$

$$\underline{Z} = \begin{bmatrix} — (\underline{z}^{(1)})^T — \\ \vdots \\ — (\underline{z}^{(N)})^T — \end{bmatrix} \qquad \underset{N \times M}{\underline{Z}} = \underset{N \times D}{\underline{X}} \; \underset{D \times M}{\underline{W}}$$