- Hotteling transformation
- Karhunen-Loeve decomposition
- Proper orthogonal decomposition
- singular value decomposition.

Principal Components Analysis (PCA)

Setup: Given data $(x_1, x_2, \dots, x_n)$, $x_i \in \mathbb{R}^d$

Goal: Encode the data in a low-dimensional representation:

$$\underset{\text{latent variables}}{z_i} = \overset{\text{projection/encoding}}{f(x_i)}, \quad z_i \in \mathbb{R}^q, \quad q \ll d.$$

- Maximum variance formulation (Hotteling 1933):

To begin with, consider a 1-dimensional subspace, i.e. $q = 1$

We can define a coordinate vector for this sub-space $u_1 \in \mathbb{R}^d$,

$$\|u_1\|_2 = u_1^T u_1 = 1$$

Each data point x_i can be projected onto the sub-space span

by u_1 as: $\underset{1\times d}{u_1^T}\underset{d\times 1}{x_i}$ (scalar quantity) $\quad u_1 = [0\ 0\ 0\ 1\ 0 \dots 0$

Therefore, the mean of the projected data is:

$$\mu := u_1^T \bar{x}, \quad \bar{x} := \frac{1}{n}\sum_{i=1}^{n} x_i \longrightarrow \text{sample mean}$$

and the variance of the projected data is:

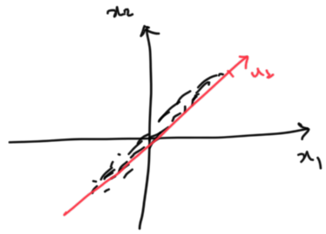
$$var := \frac{1}{n}\sum_{i=1}^{n}\{u_1^T x_n - u_1^T \bar{x}\}^2 = u_1^T S u_1, \quad \text{where}$$

$$S := \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^T \longrightarrow \text{sample covariance matrix.}$$

We seek to identify a direction u that captures the most

We seek to identify a direction u_1 that captures the

variance in the data (i.e. that "best" summarizes the data).



$\Longrightarrow$ Find u_1^* s.t.:

Lagrange multiplier

$$u_1^* = \underset{u_1}{\arg\max} \left\{ u_1^T S u_1 + \lambda_1 (1 - u_1^T u_1) \right\}$$

Critical points:

$$\frac{\partial L}{\partial u_1} = 0 \quad \Longrightarrow \quad S u_1 = \lambda_1 u_1$$

This implies that u_1^* (i.e the direction that captures the most variance in the data should be an eigenvector of S.

$$\Longrightarrow S u_1 = \lambda_1 u_1 \Longrightarrow u_1^T S u_1 = \lambda_1 \cancelto{1}{u_1^T u_1} \Rightarrow u_1^T S u_1 = \lambda_1$$

The variance will be maximum if we set the eignevector u_1 to be the one that corresponds to the maximum eigenvalue of S.

u_1^* is Known as the first principal component.

What about u_2 (i.e. $q=2$) ?

$$u_2^* = \underset{u_2}{\arg\max} \left\{ u_2^T S u_2 + \lambda_2 (1 - u_2^T u_2) \right\}$$

$$\boxed{\text{s.t. } \quad u_1 \perp u_2}$$

u_2^* should correspond to the eignvector of S that corresponds to the second largest eigenvalue of S.

Practical Implementation:

Given a data matrix $\underset{n \times d}{X}$

1.) Normalize the data to have a zero mean: $\underset{n\times d}{X} = X - \mathbb{E}[X]$

2.) Compute the sample covariance: $\underset{d\times d}{S} = \frac{1}{n-1} \underset{d\times n}{X^T} \underset{n\times d}{X}$ (Symmetric PSD)

3.) Compute the SVD of S:

$$\underset{}{S} = \underset{d \times d}{\underline{\underline{W}}} \; \underset{d \times d}{\Lambda} \; \underset{d \times d}{W^T} \longrightarrow d \text{ eig-pairs}$$

4.) Choose the dimension of the latent space (q), and keep the q eigenvectors that correspond q largest eigenvalues of S.

5.) <u>Encode</u>: $\underset{n \times q}{Z} = \underset{n \times d}{X} \; \underset{d \times q}{W}$

6.) <u>Reconstruct</u>: $\underset{n \times q}{X} = \underset{n \times q}{Z} \; \underset{q \times d}{W^T}$