Mean Values

Given a data set $\mathcal{D} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$, $\boldsymbol{x}_n \in \mathbb{R}^D$, we compute the mean of the data set as

$$\mathbb{E}[\mathcal{D}] = \frac{1}{N} \sum_{n=1}^{N} \boldsymbol{x}_n$$

Variances of 1D data sets

Given a data set $\mathcal{D} = \{x_1, \ldots, x_N\}$, $x_n \in \mathbb{R}$, we compute the variance of the data set as

$$\mathbb{V}[\mathcal{D}] = \frac{1}{N} \sum_{n=1}^{N} (x_n - \mu)^2$$

where μ is the mean value of the data set.

Variances of higher-dimensional data sets

Given a data set $\mathcal{D} = \{\boldsymbol{x}_1, \ldots, x_N\}$, $\boldsymbol{x}_n \in \mathbb{R}^D$, we compute the variance of the data set as

$$\mathbb{V}[\mathcal{D}] = \frac{1}{N} \sum_{n=1}^{N} (\boldsymbol{x}_n - \boldsymbol{\mu})(\boldsymbol{x}_n - \boldsymbol{\mu})^\top \in \mathbb{R}^{D \times D}$$

where $\boldsymbol{\mu} \in \mathbb{R}^D$ is the mean value of the data set.

Effect of Linear Transformations

Consider a data set $\mathcal{D} = \{\boldsymbol{x}_1, \ldots, x_N\}$, $\boldsymbol{x}_n \in \mathbb{R}^D$, with

$$\mathbb{E}[D] = \boldsymbol{\mu}$$
$$\mathbb{V}[D] = \boldsymbol{\Sigma}$$

If we now modify every $\boldsymbol{x}_i \in \mathcal{D}$ according to

$$\boldsymbol{x}_i' = \boldsymbol{A}\boldsymbol{x}_i + \boldsymbol{b}$$

for a given $\boldsymbol{A}, \boldsymbol{b}$, then

$$\mathbb{E}[\mathcal{D}'] = \boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{b}$$
$$\mathbb{V}[\mathcal{D}'] = \boldsymbol{A}\boldsymbol{Q}\boldsymbol{A}^\top$$

where $\mathcal{D}' = \{\boldsymbol{x}_1', \ldots, \boldsymbol{x}_N'\}$

Dot product

The **dot product** is defined as

$$\boldsymbol{x}^\top \boldsymbol{y} = \sum_{d=1}^{D} x_d y_d \,, \quad \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^D \,.$$

- The **length** of $\boldsymbol{x}$ is then

$$\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x}^\top \boldsymbol{x}} \,.$$

- The **angle** ω between two vectors $\boldsymbol{x}, \boldsymbol{y}$ can be computed using

$$\cos \omega = \frac{\boldsymbol{x}^\top \boldsymbol{y}}{\|\boldsymbol{x}\| \, \|\boldsymbol{y}\|}$$

Inner product

Consider a vector space V. A positive definite, symmetric bilinear mapping $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ is called an **inner product** on V.

- **symmetric**: For all $\boldsymbol{x}, \boldsymbol{y} \in V$ it holds that $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$
- **positive definite**: For all $\boldsymbol{x} \in V \backslash \{\boldsymbol{0}\}$ it holds that $\langle \boldsymbol{x}, \boldsymbol{x} \rangle > 0\,, \quad \langle \boldsymbol{0}, \boldsymbol{0} \rangle = 0$
- **bilinear**: For all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V, \lambda \in \mathbb{R}$

$$\langle \lambda \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{z} \rangle = \lambda \langle \boldsymbol{x}, \boldsymbol{z} \rangle + \langle \boldsymbol{y}, \boldsymbol{z} \rangle$$
$$\langle \boldsymbol{x}, \lambda \boldsymbol{y} + \boldsymbol{z} \rangle = \lambda \langle \boldsymbol{x}, \boldsymbol{y} \rangle + \langle \boldsymbol{x}, \boldsymbol{z} \rangle$$

Inner product: Lengths and distances

Consider a vector space V with an inner product $\langle \cdot, \cdot \rangle$.

- The **length** of a vector $\boldsymbol{x} \in V$ is

$$\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$$

- The **distance** between two vectors $\boldsymbol{x}, \boldsymbol{y} \in V$ is given by

$$d(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\| = \sqrt{\langle \boldsymbol{x} - \boldsymbol{y}, \boldsymbol{x} - \boldsymbol{y} \rangle}$$

Inner product: Angles

Consider a vector space V with an inner product $\langle \cdot, \cdot \rangle$.
The **angle** ω between two vectors $\boldsymbol{x}, \boldsymbol{y} \in V$ can be computed via

$$\cos \omega = \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{x}\| \, \|\boldsymbol{y}\|}$$

where the length/norm

$$\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$$

is defined via the inner product.

Projection onto 1D subspaces

Consider a vector space V with the dot product at the inner product and a subspace U of V. With a basis vector $\boldsymbol{b}$ of U, we obtain the **orthogonal projection** of any vector $\boldsymbol{x} \in V$ onto U via

$$\pi_U(\boldsymbol{x}) = \lambda \boldsymbol{b}, \quad \lambda = \frac{\boldsymbol{b}^\top \boldsymbol{x}}{\boldsymbol{b}^\top \boldsymbol{b}} = \frac{\boldsymbol{b}^\top \boldsymbol{x}}{\|\boldsymbol{b}\|^2}$$

where λ is the **coordinate** of $\pi_U(\boldsymbol{x})$ with respect to $\boldsymbol{b}$.
The **projection matrix** $\boldsymbol{P}$ is

$$\boldsymbol{P} = \frac{\boldsymbol{b}\boldsymbol{b}^\top}{\boldsymbol{b}^\top \boldsymbol{b}} = \frac{\boldsymbol{b}\boldsymbol{b}^\top}{\|\boldsymbol{b}\|^2}$$

such that

$$\pi_U(\boldsymbol{x}) = \boldsymbol{P}\boldsymbol{x}$$

for all $\boldsymbol{x} \in V$.

Projection onto k-dimensional subspaces

Consider an n-dimensional vector space V with the dot product at the inner product and a subspace U of V. With basis vectors $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k$ of U, we obtain the **orthogonal projection** of any vector $\boldsymbol{x} \in V$ onto U via

$$\pi_U(\boldsymbol{x}) = \boldsymbol{B}\boldsymbol{\lambda}\,, \quad \boldsymbol{\lambda} = (\boldsymbol{B}^\top \boldsymbol{B})^{-1}\boldsymbol{B}^\top \boldsymbol{x}$$

$$\boldsymbol{B} = (\boldsymbol{b}_1 | \cdots | \boldsymbol{b}_k) \in \mathbb{R}^{n \times k}$$

where $\boldsymbol{\lambda}$ is the **coordinate vector** of $\pi_U(\boldsymbol{x})$ with respect to the basis $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k$ of U.
The **projection matrix** $\boldsymbol{P}$ is

$$\boldsymbol{P} = \boldsymbol{B}(\boldsymbol{B}^\top \boldsymbol{B})^{-1}\boldsymbol{B}^\top$$

such that

$$\pi_U(\boldsymbol{x}) = \boldsymbol{P}\boldsymbol{x}$$

for all $\boldsymbol{x} \in V$.

Key steps of PCA algorithm

1. Compute the mean $\boldsymbol{\mu}$ of the data matrix $\boldsymbol{X} = [\boldsymbol{x}_1|...|\boldsymbol{x}_N]^\top \in \mathbb{R}^{N \times D}$
2. Mean subtraction: Replace all data points $\boldsymbol{x}_i$ with $\tilde{\boldsymbol{x}}_i = \boldsymbol{x}_i - \boldsymbol{\mu}$.
3. Divide the data by its standard deviation in each dimension: $\bar{\boldsymbol{X}}^{(d)} = \tilde{\boldsymbol{X}} / \sigma(\boldsymbol{X}^{(d)})$ for $d = 1, \ldots, D$.
4. Compute the eigenvectors (orthonormal) and eigenvalues of the data covariance matrix $\boldsymbol{S} = \frac{1}{N} \bar{\boldsymbol{X}}^\top \bar{\boldsymbol{X}}$
5. Choose the eigenvectors associated with the M largest eigenvalues to be the basis of the principal subspace.
6. Collect these eigenvectors in a matrix $\boldsymbol{B} = [\boldsymbol{b}_1, ..., \boldsymbol{b}_M]$
7. Orthogonal projection of the data onto the principal axis using the projection matrix $\boldsymbol{B}\boldsymbol{B}^\top$

PCA in high dimensions

- We need to solve the eigenvector/eigenvalue equation

$$\frac{1}{N}\bar{\boldsymbol{X}}\bar{\boldsymbol{X}}^\top \boldsymbol{c}_i = \lambda_i \boldsymbol{c}_i$$

  where $\boldsymbol{c}_i = \bar{\boldsymbol{X}}\boldsymbol{b}_i$
- We want to recover the original eigenvectors $\boldsymbol{b}_i$ of the data covariance matrix $\boldsymbol{S} = \frac{1}{N}\bar{\boldsymbol{X}}^\top\bar{\boldsymbol{X}}$
- Left-multiplying eigenvector equation by $\bar{\boldsymbol{X}}^\top$ yields

$$\underbrace{\frac{1}{N}\bar{\boldsymbol{X}}^\top\bar{\boldsymbol{X}}}_{=\boldsymbol{S}}\bar{\boldsymbol{X}}^\top \boldsymbol{c}_i = \lambda_i \bar{\boldsymbol{X}}^\top\boldsymbol{c}_i$$

  and we recover $\bar{\boldsymbol{X}}^\top\boldsymbol{c}_i$ as an eigenvector of $\boldsymbol{S}$ with (the same) eigenvalue λ_i

Note: To perform PCA as discussed in the lecture we need to make sure that $\|\bar{\boldsymbol{X}}^\top\boldsymbol{c}_i\| = 1$.