MATHEMATICS

Principal Component Analysis Cheat Sheet

MEAN AND VARIANCE OF HIGH-DIMENSIONAL DATASETS

Variance of High-Dimensional Datasets

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

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

Linear Transformations

If we now modify every $x_i \in \mathcal{D}$ according to $\boldsymbol{x}'_i = \boldsymbol{A}\boldsymbol{x}_i + \boldsymbol{b}$, then the transformed mean and variance are:

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

INNER PRODUCT

Definition

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 $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V, \lambda \in \mathbb{R}$

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

In the canonical Euclidean context (no abstract spaces), the inner product coincides with the dot product between vectors. However, nonstandard definitions of the inner product may generate different results.

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

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

The length/norm $\|\boldsymbol{x}\|$ is defined via the inner product.

PROJECTIONS

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 $\boldsymbol{b}_1, \dots, \boldsymbol{b}_k$ of U (which are concatenated in the matrix $\boldsymbol{B}$), we obtain the **orthogonal projection** of any vector $\boldsymbol{x} \in V$ onto U via

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

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

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

such that $\pi_U(\boldsymbol{x}) = \boldsymbol{P}\boldsymbol{x}$ for all $\boldsymbol{x} \in V$.

PCA ALGORITHM

Key Steps

1. **Compute the mean $\boldsymbol{\mu}$** of the data matrix $\boldsymbol{X} = [\boldsymbol{x}_1 | \dots | \boldsymbol{x}_N]^\top \in \mathbb{R}^{N \times D}$
2. **Mean subtraction**: Replace all data points $\boldsymbol{x}_i$ with $\tilde{\boldsymbol{x}}_i = \boldsymbol{x}_i - \mu$.
3. **Divide the data by its standard deviation** in each dimension: $\overline{\boldsymbol{X}}^{(d)} = \tilde{\boldsymbol{X}} / \sigma\left(\boldsymbol{X}^{(d)}\right)$ for $d = 1, \dots, D$.
4. **Compute the eigenvectors** (orthonormal) **and eigenvalues** of the data covariance matrix $\boldsymbol{S} = \frac{1}{N} \overline{\boldsymbol{X}}^\top \overline{\boldsymbol{X}}$
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 $\boldsymbol{B} = [\boldsymbol{b}_1, \dots, \boldsymbol{b}_M]$
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

$$\underbrace{\frac{1}{N} \overline{\boldsymbol{X}}^\top \overline{\boldsymbol{X}}}_{=\boldsymbol{S}} \overline{\boldsymbol{X}}^\top \boldsymbol{c}_i = \lambda_i \overline{\boldsymbol{X}}^\top \boldsymbol{c}_i \qquad (\text{where } \boldsymbol{c}_i = \overline{\boldsymbol{X}} \boldsymbol{b}_i)$$

We multiply the canonical eigenvectors equation by $\overline{\boldsymbol{X}}^\top$ so that we can recover $\overline{\boldsymbol{X}}^\top \boldsymbol{c}_i$ as the eigenvector of $\boldsymbol{S}$ with (the same) eigenvalue λ_i.