Statistical Data Analysis

Dr. Jana de Wiljes

18. Januar 2022

Universität Potsdam

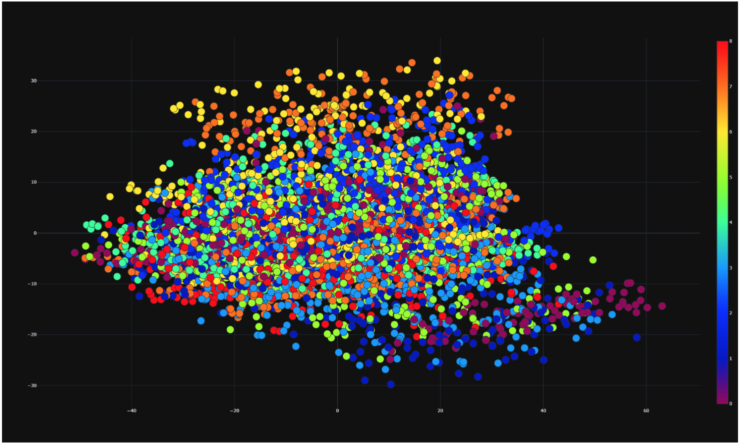
Principal Component Analysis

Principal Component Analysis (PCA)

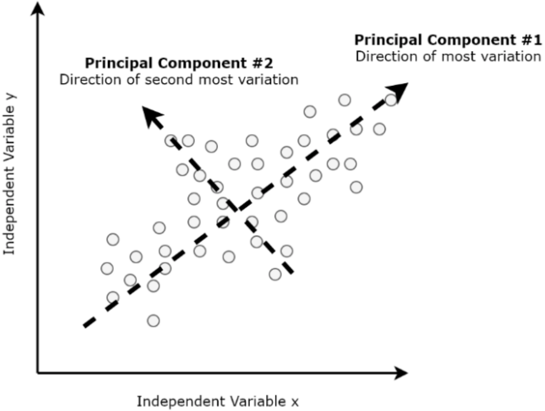
Snapshot information:

1. Dimensionality reduction, i.e., represent it in a more tractable, lower-dimensional form, without losing too much information
   - Data Compression (Save computation/memory) Noise Reduction
   - Noise Reduction/ avoid overfitting to noise
   - Data Visualization (e.g., in two dimensions)
2. unsupervised learning algorithm (Pearson 1901)
3. Idea: uses an orthogonal transformation to convert a set of observations of correlated variables into a set of linearly uncorrelated variables (called principal components)
4. Known under many different names:
   - Karhunen-Love transformation
   - Hotelling transformation
   - empirical orthogonal functions

Visualisation

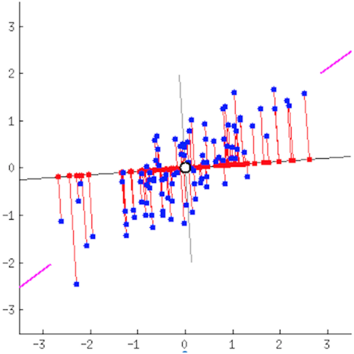


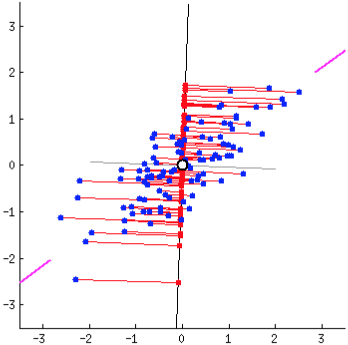
Motivation behind PCA



Idea: PCA finds directions with maximum variability

Motivation behind PCA





PCA:

- Identify a Hyperplane that lies closest to the data
- Project the data onto the hyperplane.

Background information

Orthogonal basis

Def: Let V be a vector space with scalar product $\langle \cdot, \cdot \rangle$ and $\{v_j\}_{j \in J}$ a family of vectors.

- $\{v_j\}_{j \in J}$ are a **orthogonal system**, if $\langle v_j, v_k \rangle = 0 \; \forall j \neq k \in J$ and $v_j \neq 0 \; \forall j \in J$.
- $\{v_j\}_{j \in J}$ is an **orthonormal system**, if additionally: $\langle v_j, v_j \rangle = 1 \; \forall j$ ($\Leftrightarrow \|v_j\| = 1$), in other words: $\langle v_j, v_k \rangle = \delta_{jk} \; \forall j, k \in J$.
- An Orthogonal- respectively. -normalsystem is called **orthogonal basis** bzw. **orthonormal basis**, if the die vectors of the systems form a basis.

Representation with respect to an orthonormal basis

Theorem

Let $\{v_j\}_{j=1,\dots,n}$ *be an orthonormal basis of the vector space* V *and* $w \in V$ *a second vector. Then the following holds:*

$$w = \langle v_1, w \rangle v_1 + \cdots + \langle v_n, w \rangle v_n.$$

Proof

Orthogonal Projection

Def: Let $U \subseteq V$ be a subspace. A map $\varphi\colon V \to$ is called **projection of** V **onto** U, if for every $u \in U$ gilt: $\varphi(u) = u$. A projection is called **orthogonal projection onto subspace** U, if for every vector $v \in V$ holds:

$$(\varphi(v) - v) \perp U.$$

Example: Hyperspace

Orthonormalbasis

Def: Let V be a K-vector space with scalar product and $U \subseteq V$ a finite-dimensional subspace. Furthermore let $\{u_1, \ldots, u_k\}$ be an orthonormal basis of U. Then the map

$$\mathrm{pr}_U \colon V \to U, \quad v \mapsto \sum_{j=1}^{k} \langle u_j, v \rangle u_j$$

is an orthogonal projection.

Approximation theorem

Theorem

Let V be a $\mathbb{R}$-vector space, with a scalar product and the corresponding norm $\|\cdot\|$. Let U be a subspace of V. Then for every $v \in V$ $pr_U(v)$ is the best approximation of v in U, i.e.,

$$\|v - pr_U(v)\| < \|v - u\| \; \forall u \in U \text{ mit } u \neq pr_U(v).$$

PCA

Given: data set of p dimensional vectors

Goal: Want to project them to q - dimensional subspace ($q << p$)

Principal components: q uncorrelated, orthogonal directions formed by projecting the original data

Derivation

Derivation

Derivation

Derivation

Derivation