Statistical Data Analysis

Jana de Wiljes

wiljes@uni-potsdam.de

www.dewiljes-lab.com

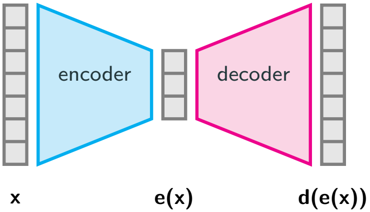
January 2, 2023

Dimension reduction

Dimension reduction

Goal: reducing the number of given features in a data set $x_i \in \mathcal{S}$ with $i \in \{1, \ldots, N\}$

- choose model class for the encoder $e \in \mathcal{E}$ and for the decoder $d \in \mathcal{D}$
- and appropriate loss functional $l(x, d(e(x)))$



Dimension reduction problem

For a given data $\mathcal{S}$ and fixed families of functions $\mathcal{E}$ and $\mathcal{D}$

$$(e^*, d^*) = \arg \min_{(e,d) \in \mathcal{E} \times \mathcal{D}} l(x, d(e(x))) \tag{1}$$

Orthogonal Projection

Orthogonal basis

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

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

Example

Example:

- the canonical unit vectors $e_j \in K^n$ are forming an orthonormal basis of the vector space K^n wrt. the standard scalar product.
- the columns v_i an orthogonal respectively unitary $n \times n$-matrix A form an orthogonal system, since:

$$(\langle v_j, v_k \rangle)_{j=1,\ldots,n,\ k=1,\ldots,n} = \overline{A}^t A = E.$$

Gram-Schmidt-Algorithm

Theorem: Let V be a K-vector space with scalar product and $w_1, \dots, w_n$ a family of linearly independent vectors. Particularly

$$\dim\langle w_1, \dots, w_k\rangle = k \ \forall k = 1, \dots, n.$$

Then vectors $v_1, \dots, v_n$ in V exist with $\langle v_i, v_j\rangle = \delta_{ij}$ für $i, j \in \{1, 2, \dots, n\}$, so that:

$$\langle v_1, \dots, v_k\rangle = \langle w_1, \dots, w_k\rangle \text{ für } k = 1, \dots, n.$$

Proof: we are showing the result via induction

- Choose $v_1 := \frac{w_1}{\|w_1\|}$. Then v_1 is an unity vector and $\langle v_1\rangle = \langle w_1\rangle$.
- Given $v_1, \dots, v_{k-1}$ we set

$$u_k := w_k - \sum_{j=1}^{k-1} \langle v_j, w_k\rangle v_j.$$

Proof

The following holds:

$$\langle v_l, u_k \rangle = \langle v_l, w_k \rangle - \sum_{j=1}^{k-1} \langle v_j, w_k \rangle \langle v_l, v_j \rangle = 0, \quad l = 1, 2, \ldots, k-1,$$

in other words $u_k \perp \langle v_1, \ldots, v_{k-1} \rangle$ and $u_k \neq 0$, since

$$w_k \notin \langle v_1, \ldots, v_{k-1} \rangle = \langle w_1, \ldots, w_{k-1} \rangle.$$

Then we define

$$v_k := \frac{u_k}{\|u_k\|},$$

so that $v_1, \ldots, v_k$ is an **orthonormal system**. Further:

$$\langle w_1, \ldots, w_k \rangle = \langle v_1, \ldots, v_k \rangle.$$

$\square$

Example

We consider $V = \mathbb{R}^4$ with the standard-scalar product and the basis:

$$w_1 = \begin{pmatrix}1\\1\\1\\1\end{pmatrix}, w_2 = \begin{pmatrix}1\\1\\0\\0\end{pmatrix}, w_3 = \begin{pmatrix}1\\-1\\1\\-1\end{pmatrix}, w_4 = \begin{pmatrix}1\\0\\0\\0\end{pmatrix}.$$

To obtain an **orthonormal basis** from the given basis we use the Gram-Schmidt-Method:

$$v_1 = \frac{1}{2} w_1 = w_1 = \begin{pmatrix}\frac{1}{2}\\\frac{1}{2}\\\frac{1}{2}\\\frac{1}{2}\end{pmatrix}, \text{ da } 2 = \sqrt{4} = \|w_1\|.$$

Now we can determine v_2:

$$u_2 = w_2 - \langle v_1, w_2 \rangle \cdot v_1 = \begin{pmatrix}1\\1\\0\\0\end{pmatrix} - 1 \cdot \begin{pmatrix}\frac{1}{2}\\\frac{1}{2}\\\frac{1}{2}\\\frac{1}{2}\end{pmatrix} = \begin{pmatrix}\frac{1}{2}\\\frac{1}{2}\\-\frac{1}{2}\\-\frac{1}{2}\end{pmatrix}, \text{ yielding } v_2 = \frac{u_2}{\|u_2\|}.$$

Example

The third vector v_3 is determined as follows:

$$u_3 = w_3 - \langle v_1, w_3 \rangle v_1 - \langle v_2, w_3 \rangle v_2, \text{ so } v_3 = \frac{u_3}{\|u_3\|} = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}.$$

Finally we determine

$$\begin{aligned} u_4 &= w_4 - \langle v_1, w_4 \rangle v_1 - \langle v_2, w_4 \rangle v_2 - \langle v_3, w_4 \rangle v_3 && (2) \\ &= \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} - \frac{1}{2} \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \\ -\frac{1}{2} \end{pmatrix} - \frac{1}{2} \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{4} \\ -\frac{1}{4} \\ -\frac{1}{4} \\ \frac{1}{4} \end{pmatrix} && (3) \end{aligned}$$

yielding

$$v_4 = \frac{u_4}{\|u_4\|} = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix}.$$

Example

Writing the vectors in the column in a matrix:

$$(v_1 \ v_2 \ v_3 \ v_4) = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \in \mathsf{SO}(4) \subseteq \mathsf{GL}(4, \mathbb{R}).$$

Representation via orthonormal basis

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

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

Proof: We compute the difference $u = w - \sum_{j=1}^{n} \langle v_j, w\rangle v_j$, and apply the scalar product $\langle v_k, \cdot\rangle$:

$$\begin{aligned}\langle v_k, u\rangle &= \langle v_k, w\rangle - \sum_{j=1}^{n} \langle v_k, \langle v_j, w\rangle v_j\rangle \\ &= \langle v_k, w\rangle - \sum_{j=1}^{n} \langle v_j, w\rangle \langle v_k, v_j\rangle = \langle v_k, w\rangle - \langle v_k, w\rangle = 0.\end{aligned}$$

It follows that u is orthogonal to $v_1, \dots, v_n$ and consequently orthogonal to every linear combination of j$v_1, \dots, v_n$. Since $v_1, \dots, v_n$ generates the vector space V the following holds

$$\langle u, u\rangle = 0 \Rightarrow \|u\| = 0 \Rightarrow u = 0.$$

$\square$

orthogonale Untervektorraum

Definition: Let V be a vector space with scalar product $\langle \cdot, \cdot \rangle$ and $U \subseteq V$ a subvector space. Then

$$U^{\perp} := \{v \in V \mid \langle v, u \rangle = 0 \; \forall u \in U\}$$

is called the **subvector space orthogonal to** U or **orthogonal complement** of U.

Remark: It holds that:

$$U^{\perp} \cap U = 0,$$

since only the vector of zeros is orthogonal to itself.
If $\dim V < \infty$, the following holds:

$$U^{\perp} \oplus U = V,$$

since $\dim U^{\perp} = \dim V - \dim U$. Further

$$(U^{\perp})^{\perp} = U.$$

Orthogonal subvector space

Definition: Let $U \subseteq V$ be a subvector space. A map $\varphi\colon V \to U$ is called a **projection of V on U**, falls für jedes $u \in U$ gilt: $\varphi(u) = u$. A projection is called an **orthogonal projection on the subvector space** U, if for every vector $v \in V$ the following holds:

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

Theorem: Let V be a K-vector space with scalar product and $U \subseteq V$ a finite dimensional subspace. Let$\{u_1, \dots, u_k\}$ be a orthonormal basis of U. Then the maps

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

and

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

are orthogonal Projektions onto the respective subspaces.

Proof

Proof: firstly we realise

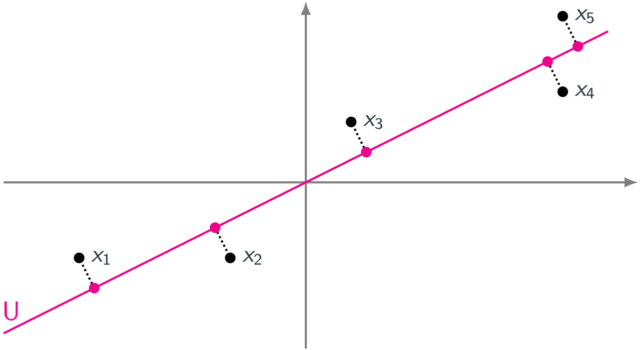
- $\mathrm{pr}_U(v) \in U$
- $v = \mathrm{pr}_U(v) + \mathrm{pr}_{U^\perp}(v)$
- $\mathrm{pr}_U(u) = u \; \forall u \in U$ and $\mathrm{pr}_{U^\perp}(w) = w \; \forall w \in U^\perp$

Then we consider

$$\langle u_l, w \rangle = \langle u_l, v \rangle - \sum_{j=1}^{k} \langle u_j, v \rangle \langle u_l, u_j \rangle = 0, \quad l = 1, \ldots, k.$$

which yields: $w = \mathrm{pr}_{U^\perp}(v) \in U^\perp = \langle u_1, \ldots, u_k \rangle^\perp$. Further $u := \mathrm{pr}_U(v)$ implies $v = u + w \in U \oplus U^\perp$ and $u \perp w$. pr_U. Then $\mathrm{pr}_{U^\perp}$ are orthogonal projections onto U and $U^\perp$ respectively. Further $U \oplus U^\perp = V$. $\square$

Visualisation in 2 dimensions



Example

1. Let $V = \mathbb{R}^n$ and $U = \langle e_1, \dots, e_k \rangle$. Then $U^\perp = \langle e_{k+1}, \dots, e_n \rangle$ and further the following holds for $x = (x_1, \dots, x_n)^t$:

$$\mathrm{pr}_U(x) = (x_1, \dots, x_k, 0, \dots, 0)^t = \sum_{j=1}^{k} \langle e_j, x \rangle e_j,$$
$$\mathrm{pr}_{U^\perp}(x) = (0, \dots, 0, x_{k+1}, \dots, x_n)^t = x - \mathrm{pr}_U(x).$$

2. Let $L \subset \mathbb{R}^n$ be a line through the origin in direction v. Without loss of generality we can choose $\|v\| = 1$. Then the orthogonal projektions onto L or $L^\perp$ respectively are:

$$\mathrm{pr}_L \colon \mathbb{R}^n \to L, \quad x \mapsto \langle v, x \rangle v,$$
$$\mathrm{pr}_{L^\perp} \colon \mathbb{R}^n \to L^\perp, \quad x \mapsto x - \langle v, x \rangle v.$$

Consider the concrete example:

$$L = \langle (1, \dots, 1)^t \rangle \subset \mathbb{R}, \text{ d.h. } v = \frac{1}{\sqrt{n}}(1, \dots, 1)^t.$$

Example

1. Es ergeben sich: $L^\perp = \{x \mid \sum x_i = 0\}$ und $\langle v, x\rangle = \frac{1}{\sqrt{n}} \sum x_i = \sqrt{n} \cdot \overline{x}$, wobei $\overline{x} = \frac{1}{n} \sum x_i$ das **arithmetische Mittel** der Komponenten von x ist. Die beiden Projektionen sind also:

$$\begin{aligned} \mathrm{pr}_L \colon \mathbb{R}^n &\to L, \quad x \mapsto (\overline{x}, \ldots, \overline{x}), \\ \mathrm{pr}_{L^\perp} \colon \mathbb{R}^n &\to L^\perp, \quad x \mapsto x - (\overline{x}, \ldots, \overline{x}). \end{aligned}$$

   Tatsächlich ist für $y = \mathrm{pr}_{L^\perp}(x)$:

$$\sum y_i = \sum x_i - n \cdot \overline{x} = \sum x_i - \sum x_i = 0,$$

   d.h. $y \in L^\perp$.

Approximation theorem

Theorem: Let V be a $\mathbb{R}$-vector space with a scalar product and the corresponding norm $\|\cdot\|$. Let U be a subvector space. For every $v \in V$ $\mathrm{pr}_U(v)$ is the best approximation of v in U, i.e.,:

$$\|v - \mathrm{pr}_U(v)\| < \|v - u\| \; \forall u \in U \text{ with } u \neq \mathrm{pr}_U(v).$$

Proof: Since the Pythagorean theorem holds for $x, y \in V$, i.e.,

$$\|x + y\|^2 = \|x\|^2 + 2\langle x, y\rangle + \|y\|^2,$$

follows that

$$\|v - u\|^2 = \| \underbrace{v - \mathrm{pr}_U(v)}_{\in U^\perp} + \underbrace{\mathrm{pr}_U(v) - u}_{\in U} \|^2 \tag{4}$$

$$= \|v - \mathrm{pr}_U(v)\|^2 + \| \mathrm{pr}_U(v) - u\|^2 \tag{5}$$

$$\geq \|v - \mathrm{pr}_U(v)\|^2. \tag{6}$$

Equality is obtain if $u = \mathrm{pr}_U(v)$. □

Matrix

Definition: Let K a field. A $m \times n$ **matrix** with entries in K ist a table

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \in K^{m \times n}$$

of elements $a_{ij} \in K$. m is the number of rows and n the number of columns of A. Let $A = (a_{ij}) \in K^{m \times n}$ and $B = (b_{jk}) \in K^{n \times r}$ be two matrices, so that the column number of A coincides with the number of rows of B. Then the product

$$C = A \cdot B = (c_{ik}) \in K^{m \times r}$$

is given via

$$c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}.$$

Principal Component Analysis

Principal Component Analysis (PCA)

Goal: reducing the number of given features in a data set $x_i \in \mathcal{S}$ with $i \in \{1, \dots, N\}$ via a linear projection

- choose model class such that the combination of the encoder and decoder is $\{d(e(x)) = \sum_{j=1}^{k} \langle u_j, x \rangle u_j | u_1, \dots, u_k$ orthonomal basis of U $\}$ where U is k-dimensional subspace
- loss functional $l(x, d(e(x))) = \|x - d(e(x))\|^2$

Optimisation Problem: For a given data $\mathcal{S} = \{x_1, \dots, x_N\}$ where $x_i \in \mathbb{R}^d$ the associated optimisation problem is defined by

$$Q^* = \arg \min_{Q \in \mathbb{R}^{d \times k} \text{ with } Q^\top Q = I} \frac{1}{N} \sum_{i=1}^{N} \|x_i - \sum_{j=1}^{k} \langle u_j, x_i \rangle u_j\|^2$$

where $Q = \begin{pmatrix} | & & | \\ u_1 & \dots & u_k \\ | & & | \end{pmatrix}$.

Principal Component Analysis (PCA)

Given: data $\mathcal{S} = \{x_1, \ldots, x_N\}$ where $x_i \in \mathbb{R}^d$

Consider: the following optimisation problem

$$\begin{aligned}
Q^* &= \arg \min_{Q \in \mathbb{R}^{d \times k},\ Q^\top Q = I} \frac{1}{N} \sum_{i=1}^{N} \|x_i - \sum_{j=1}^{k} \langle u_j, x_i \rangle u_j\|^2 \\
&= \arg \min_{Q \in \mathbb{R}^{d \times k},\ Q^\top Q = I} \frac{1}{N} \sum_{i=1}^{N} \|x_i - \sum_{j=1}^{k} u_j^\top x_i u_j\|^2 \\
&= \arg \min_{Q \in \mathbb{R}^{d \times k},\ Q^\top Q = I} \frac{1}{N} \sum_{i=1}^{N} \|x_i - \sum_{j=1}^{k} \langle u_j, u_j \rangle x_i\|^2 \\
&= \arg \min_{Q \in \mathbb{R}^{d \times k},\ Q^\top Q = I} \frac{1}{N} \sum_{i=1}^{N} \|x_i - (\sum_{j=1}^{k} u_j^T u_j) x_i\|^2 \\
&= \arg \min_{Q \in \mathbb{R}^{d \times k},\ Q^\top Q = I} \frac{1}{N} \sum_{i=1}^{N} \|x_i - QQ^T x_i\|^2
\end{aligned}$$