Orthogonal projections

- If $V \subseteq \mathbb{R}^d$, then $V^\perp := \{y \in \mathbb{R}^d \mid \forall x \in V : \langle y, x \rangle = 0\}$ is the orthogonal complement of V

Orthogonal projections

- If $V \subseteq \mathbb{R}^d$, then $V^\perp := \{y \in \mathbb{R}^d \mid \forall x \in V : \langle y, x \rangle = 0\}$ is the orthogonal complement of V
- $V \cap V^\perp = \{0\}$ and for all $x \in \mathbb{R}^d$ there exist unique $x' \in V, x'' \in V^\perp$ with $x = x' + x''$

Orthogonal projections

- If $V \subseteq \mathbb{R}^d$, then $V^\perp := \{y \in \mathbb{R}^d \mid \forall x \in V : \langle y, x \rangle = 0\}$ is the orthogonal complement of V
- $V \cap V^\perp = \{0\}$ and for all $x \in \mathbb{R}^d$ there exist unique $x' \in V, x'' \in V^\perp$ with $x = x' + x''$
- $\pi_V : \mathbb{R}^d \to V, \pi_V(x) = x'$, orthogonal projection onto V, x'' denoted $\pi_V(x)^\perp$.

Orthogonal projections

- If $V \subseteq \mathbb{R}^d$, then $V^\perp := \{y \in \mathbb{R}^d \mid \forall x \in V : \langle y, x \rangle = 0\}$ is the orthogonal complement of V
- $V \cap V^\perp = \{0\}$ and for all $x \in \mathbb{R}^d$ there exist unique $x' \in V, x'' \in V^\perp$ with $x = x' + x''$
- $\pi_V : \mathbb{R}^d \to V, \pi_V(x) = x'$, orthogonal projection onto V, x'' denoted $\pi_V(x)^\perp$.
- If $\dim(V) = 1$, $V = \text{span}(v)$, then

$$\pi_V(x) = \frac{\langle x, v \rangle}{\langle v, v \rangle} v \quad \text{and} \quad \pi_V(x)^\perp = x - \frac{\langle x, v \rangle}{\langle v, v \rangle} v$$

Problem 5.1 (k-variance problem)

Given $P \subset \mathbb{R}^d, |P| = n$ and $k \in \mathbb{N}$, Find the k-dimensional subspace V_k that minimizes

$$D(P, V) := \sum_{p \in P} \|p - \pi_V(p)\|^2.$$

The subspace V_k is called the (k-dimensional) singular value decomposition of P.

Characterization of optimal subspace

Lemma 5.2

For all $P \subset \mathbb{R}^d$

$$V_k = argmin_{V:\dim(V)=k}\{D(P,V)\}$$

$$\Leftrightarrow V_k = argmax_{V:\dim(V)=k}\left\{\sum_{p\in P}\|\pi_V(p)\|^2\right\}.$$

More generally, for every subspace $V \subseteq \mathbb{R}^d$

$$D(P,V) = \sum_{q\in P}\|q\|^2 - \sum_{q\in P}\|\pi_V(q)\|^2.$$

UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

Theorem 5.3

For every $P \subset \mathbb{R}^d$ and $k \in \mathbb{N}$ the subspace V_k minimizing $D(P, V)$ can be computed efficiently.

Complexity and relation to k-means

Theorem 5.3

For every $P \subset \mathbb{R}^d$ and $k \in \mathbb{N}$ the subspace V_k minimizing $D(P, V)$ can be computed efficiently.

Lemma 5.4

For every $P \subset \mathbb{R}^d$ and $k \in \mathbb{N}$

$$D(P, V_k) \leq opt_k(P).$$

Spectral algorithms

Given $P \subset \mathbb{R}^d$,

1. compute the singular value decomposition V_k, i.e. the subspace minimizing $D(P, V)$,
2. solve your favorite clustering problem with your favorite algorithm on input $\pi_{V_k}(P) := \{\pi_{V_k}(p) : p \in P\}$,
3. return the solution found in the previous step.

Definition 5.5

Let $V \subseteq \mathbb{R}^d$ be a k-dimensional subspace of $\mathbb{R}^d$ and let $B = \{v_1, \ldots, v_k\}$ be a basis of V. Basis B is an orthonormal basis (ONB) of V if

1. $\|v_i\| = 1, i = 1, \ldots, k$
2. $\langle v_i, v_j \rangle = 0$ *for* $i \neq j, i, j = 1, \ldots, n.$

Definition 5.5

Let $V \subseteq \mathbb{R}^d$ be a k-dimensional subspace of $\mathbb{R}^d$ and let $B = \{v_1, \ldots, v_k\}$ be a basis of V. Basis B is an orthonormal basis (ONB) of V if

1. $\|v_i\| = 1, i = 1, \ldots, k$
2. $\langle v_i, v_j \rangle = 0$ *for* $i \neq j, i, j = 1, \ldots, n$.

Theorem 5.6

Every subspace $V \subseteq \mathbb{R}^d$ has an orthonormal basis. Moreover, any orthonormal basis of V can be extended to an orthonormal basis of $\mathbb{R}^d$.

Length-preserving linear maps

- $V \subseteq \mathbb{R}^d$ subspace with orthonormal basis $B_V = \{v_1, \ldots, v_k\}$.

UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

Length-preserving linear maps

- $V \subseteq \mathbb{R}^d$ subspace with orthonormal basis $B_V = \{v_1, \dots, v_k\}$.
- $U \in \mathbb{R}^{k \times d}$ matrix with rows $v_1^T, \dots, v_k^T$.

UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

Length-preserving linear maps

- $V \subseteq \mathbb{R}^d$ subspace with orthonormal basis $B_V = \{v_1, \ldots, v_k\}$.
- $U \in \mathbb{R}^{k \times d}$ matrix with rows $v_1^T, \ldots, v_k^T$.
- Π_V denotes function $\Pi_V : \mathbb{R}^d \to \mathbb{R}^k, x \mapsto U \cdot x$

UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

Length-preserving linear maps

- $V \subseteq \mathbb{R}^d$ subspace with orthonormal basis $B_V = \{v_1, \ldots, v_k\}$.
- $U \in \mathbb{R}^{k \times d}$ matrix with rows $v_1^T, \ldots, v_k^T$.
- Π_V denotes function $\Pi_V : \mathbb{R}^d \to \mathbb{R}^k, x \mapsto U \cdot x$

UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

Length-preserving linear maps

- $V \subseteq \mathbb{R}^d$ subspace with orthonormal basis $B_V = \{v_1, \ldots, v_k\}$.
- $U \in \mathbb{R}^{k \times d}$ matrix with rows $v_1^T, \ldots, v_k^T$.
- Π_V denotes function $\Pi_V : \mathbb{R}^d \to \mathbb{R}^k, x \mapsto U \cdot x$

Theorem 5.7

The linear function Π_V has the following properties:

1. Π_V *is surjective.*
2. Π_V *is length-preserving on V, i.e. for all* $x \in V : \|x\| = \|\Pi_V(x)\|$.

UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

Spectral algorithms

Given $P \subset \mathbb{R}^d$,

1. compute the singular value decomposition V_k, i.e. the subspace minimizing $D(P, V)$,
2. solve your favorite clustering problem with your favorite algorithm on input $\pi_{V_k}(P) := \{\pi_{V_k}(p) : p \in P\}$,
3. return the solution found in the previous step.

Spectral algorithms

Given $P \subset \mathbb{R}^d$,

1. compute the singular value decomposition V_k, i.e. the subspace minimizing $D(P, V)$,
2. solve your favorite clustering problem with your favorite algorithm on input $\pi_{V_k}(P) := \{\pi_{V_k}(p) : p \in P\}$, i.e. compute an orthonormal basis for V_k and apply your favorite clustering algorithm on the set $\Pi_{V_k}(\pi_{V_k}(P))$
3. return the solution found in the previous step.

Lemma 5.8

Let $P \subset \mathbb{R}^d$ and let V be an arbitrary k-dimensional subspace of $\mathbb{R}^d$. Then

$$opt_k(\pi_V(P)) \leq opt_k(P),$$

where $opt_k(P)$ denotes the cost of an optimal solution of k-means with input P.

Lemma 5.9

Let $P \subset \mathbb{R}^d$ and let V be an arbitrary k-dimensional subspace of $\mathbb{R}^d$. Assume $\hat{\mathcal{C}} = \{\hat{C}_1, \ldots, \hat{C}_k\}$ is a k-clustering of $\pi_V(P)$ and denote by $\mathcal{C} := \{C_1, \ldots, C_k\}$ with $C_i := \{p \in P : \pi_V(p) \in \hat{C}_i\}$, the corresponding k-clustering of P. Then

$$cost(\pi_V(P), \hat{\mathcal{C}}) \leq cost(P, \mathcal{C}) \leq cost(\pi_V(P), \hat{\mathcal{C}}) + D(P, V).$$

Approximation guarantees for spectral algorithms

Spectral algorithms

Given $P \subset \mathbb{R}^d$,

1. compute the singular value decomposition V_k, i.e. the subspace minimizing $D(P, C)$,
2. solve your favorite clustering problem with your favorite algorithm on input $\pi_{V_k}(P) := \{\pi_{V_k}(p) : p \in P\}$,
3. return the solution found in the previous step.

Theorem 5.10

Let $P \subset \mathbb{R}^d$ and let V_k be the k-dimensional subspace of $\mathbb{R}^d$ minimizing $D(P, V)$. If $\hat{\mathcal{C}}$ is a γ-approximate k-clustering for $\pi_{V_k}(P)$, then the corresponding k-clustering $\mathcal{C}$ as defined in the previous lemma is a $(\gamma + 1)$-approximate k-clustering for P.

An excact algorithm for k-means

EXACT-K-MEANS(P, k)

Compute the set K of sets of t hyperplanes with $k \leq t \leq \binom{k}{2}$ where each hyperplane contains d affinely independent points from P;

for $S \in K$ **do**

  check that S defines an arrangement of exactly k cells;

  for *all assignments* a_S *of points of* P *on hyperplanes in* S *to cells* **do**

    for *all cells* **do**

      compute the centroid of points of P in the cell;

    end

    C_{S,a_s} := set of centroids computed in the previous step;

  end

  $C_S := \text{argmin}_{C_{S,a_S}} \{D(P, C_{S,a_S})\}$;

end

return $\text{argmin}_{C_S} \{D(P, C_S)\}$;

UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

An excact algorithm for k-means

Theorem 5.11

Algorithm EXACT-K-MEANS *solves the k-means problem optimally in time $O\left(n^{dk^2/2}\right)$.*

UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

A spectral approximation algorithm

SPECTRAL-K-MEANS(P, k)

Compute $V_k := \text{argmin}_{V:\dim(V)=k}\{D(P, V)\}$;

$\bar{C} :=$ EXACT-K-MEANS$(\pi_{V_k}(P), k)$;

return $\bar{C}$;

A spectral approximation algorithm

SPECTRAL-K-MEANS(P, k)

Compute $V_k := \text{argmin}_{V:\dim(V)=k}\{D(P, V)\}$;

$\bar{C} :=$ EXACT-K-MEANS$(\pi_{V_k}(P), k)$;

return $\bar{C}$;

Theorem 5.12

SPECTRAL-K-MEANS *is an approximation algorithm for the k-means problem with running time* $O\left(n \cdot d^2 + n^{k^3/2}\right)$ *and approximation factor* 2.

Matrix representation of point sets

- $P = \{p_1, \ldots, p_n\} \subset \mathbb{R}^d$

Matrix representation of point sets

- $P = \{p_1, \ldots, p_n\} \subset \mathbb{R}^d$
- matrix $A \in \mathbb{R}^{d \times n}$ with columns p_i called *m*atrix representation of P

Matrix representation of point sets

- $P = \{p_1, \ldots, p_n\} \subset \mathbb{R}^d$
- matrix $A \in \mathbb{R}^{d \times n}$ with columns p_i called *m*atrix representation of P
- rows of $A^T \in \mathbb{R}^{n \times d}$ are p_i^T

UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

Matrix representation of point sets

- $P = \{p_1, \ldots, p_n\} \subset \mathbb{R}^d$
- matrix $A \in \mathbb{R}^{d \times n}$ with columns p_i called *m*atrix representation of P
- rows of $A^T \in \mathbb{R}^{n \times d}$ are p_i^T
- for every $v \in \mathbb{R}^d$:

Matrix representation of point sets

- $P = \{p_1, \ldots, p_n\} \subset \mathbb{R}^d$
- matrix $A \in \mathbb{R}^{d \times n}$ with columns p_i called *m*atrix representation of P
- rows of $A^T \in \mathbb{R}^{n \times d}$ are p_i^T
- for every $v \in \mathbb{R}^d$:
  - $A^T \cdot v = \left(\langle p_1, v\rangle, \ldots, \langle p_n, v\rangle\right)^T \in \mathbb{R}^n$

Matrix representation of point sets

- $P = \{p_1, \dots, p_n\} \subset \mathbb{R}^d$
- matrix $A \in \mathbb{R}^{d \times n}$ with columns p_i called *m*atrix representation of P
- rows of $A^T \in \mathbb{R}^{n \times d}$ are p_i^T
- for every $v \in \mathbb{R}^d$:
  - $A^T \cdot v = \left(\langle p_1, v\rangle, \dots, \langle p_n, v\rangle\right)^T \in \mathbb{R}^n$
  - $\|A^T \cdot v\|^2 = v^T \cdot A \cdot A^T \cdot v = \sum_{i=1}^n \langle p_i, v\rangle^2$

Characterization of k-variance solutions

Theorem 5.13

For every set of points $P \subset \mathbb{R}^d, |P| = n$*, with matrix representation* $A \in \mathbb{R}^{d \times n}$ *:*

$$\mathrm{argmax}_{V:\dim(V)=k} \left\{ \sum_{p \in P} \|\pi_V(p)\|^2 \right\} =$$

$$\mathrm{argmax}_{ONB\ B\ :\ |B| = k} \left\{ \sum_{v \in B} v^T \cdot A \cdot A^T \cdot v \right\}$$

UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

Definition 5.14

Let $M \in \mathbb{R}^{d\times d}, \lambda \in \mathbb{R}$ *and* $v \in \mathbb{R}^d, v \neq 0$. *Then* λ *is called an eigenvalue of* M *to eigenvector* v *(and vice versa) if* $M \cdot v = \lambda \cdot v$.

Definition 5.14

Let $M \in \mathbb{R}^{d\times d}, \lambda \in \mathbb{R}$ and $v \in \mathbb{R}^d, v \neq 0$. Then λ is called an eigenvalue of M to eigenvector v (and vice versa) if $M \cdot v = \lambda \cdot v$.

Theorem 5.15

For every $A \in \mathbb{R}^{d\times n}$ the matrix $M = A \cdot A^T \in \mathbb{R}^{d\times d}$ has non-negative eigenvalues $\lambda_1 \geq \cdots \lambda_d \geq 0$. Moreover, there is an orthonormal basis $B = \{v_1, \ldots, v_d\}$ such that λ_i is an eigenvalue of M to eigenvector $v_i, i = 1, \ldots, d$.

UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

Theorem 5.16

Let $P \subset \mathbb{R}^d$ be a finite set of points with matrix representation $A \in \mathbb{R}^{d \times n}$ and $k \in \mathbb{N}$. If $A \cdot A^T$ has eigenvalues $\lambda_1 \geq \cdots \geq \lambda_d$ and $B = \{v_1, \ldots, v_d\}$ is an orthonormal basis consisting of eigenvectors, i.e. v_i is an eigenvector to eigenvalue $\lambda_i, i = 1 \ldots, d$, then

$$\text{span}\{v_1, \ldots, v_k\} = \mathit{argmin}_{V:\dim(V)=k}\{D(P, V)\}.$$

Singular values and vectors

- $M \in \mathbb{R}^{n \times d}$,

Singular values and vectors

- $M \in \mathbb{R}^{n \times d}$,
- case $d = n$: $v \in \mathbb{R}^d$ eigenvector to eigenvalue σ if
  $M \cdot v = \sigma \cdot v$

Singular values and vectors

- $M \in \mathbb{R}^{n \times d}$,
- case $d = n$: $v \in \mathbb{R}^d$ eigenvector to eigenvalue σ if $M \cdot v = \sigma \cdot v$
- generalization to $n \neq d$?

UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

- $M \in \mathbb{R}^{n \times d}$,
- case $d = n$: $v \in \mathbb{R}^d$ eigenvector to eigenvalue σ if $M \cdot v = \sigma \cdot v$
- generalization to $n \neq d$?
- can one compute eigenvectors and eigenvalues of $A \cdot A^T$ without computing the matrix product?

Singular values and vectors

- $M \in \mathbb{R}^{n \times d}$,
- case $d = n$: $v \in \mathbb{R}^d$ eigenvector to eigenvalue σ if $M \cdot v = \sigma \cdot v$
- generalization to $n \neq d$?
- can one compute eigenvectors and eigenvalues of $A \cdot A^T$ without computing the matrix product?

UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

Singular values and vectors

- $M \in \mathbb{R}^{n \times d}$,
- case $d = n$: $v \in \mathbb{R}^d$ eigenvector to eigenvalue σ if $M \cdot v = \sigma \cdot v$
- generalization to $n \neq d$?
- can one compute eigenvectors and eigenvalues of $A \cdot A^T$ without computing the matrix product?

Singular vectors and singular values

$\sigma \in \mathbb{R}$ is called singular value of M with corresponding singular vectors $v \in \mathbb{R}^d, u \in \mathbb{R}^n$ if

1. $M \cdot v = \sigma \cdot u$
2. $u^T \cdot M = \sigma \cdot v^T$.

Eigenvectors and singular vectors

Lemma 5.17

Let $M \in \mathbb{R}^{n \times d}$. Then $\sigma \in \mathbb{R}$ is a singular value of M with corresponding singular vectors $v \in \mathbb{R}^d$ and $u \in \mathbb{R}^n$ if and only if

1. *σ^2 is an eigenvalue of $M^T \cdot M$,*
2. *v is a right eigenvector of $M^T \cdot M$ to eigenvalue σ^2,*
3. *u^T is a left eigenvector of $M \cdot M^T$ to eigenvalue σ^2.*