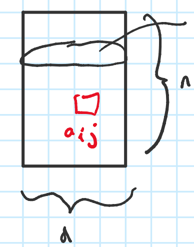
L16 SVD and Relatives

Sunday, April 5, 2020 8:54 AM

- Dimensionality Reduction techniques can be applied when working with a data matrix:

$A \in \mathbb{R}^{n \times d}$



- we will be using norms (for example $\|a_i\|$) which imply euclidean distance.
  - thus, it is important that every column has comparable units.
    - ie, things should be of similar type and scale.

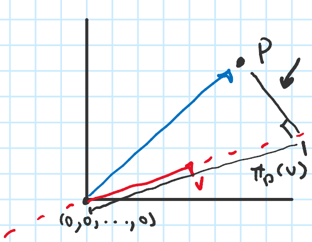
    Ex. Given n weather stations and d points in time, each column is of same type and scale.

    ↳ Thus, this satisfies the SVD assumptions.

- Challenge: what to do when d is very large??
  - Solution: reduce d to some k where $k < d$
    - The intuition is that there exists some subspace k that contains the information from d without the additional noise in d.

Projection:

- consider a unit vector $v \in \mathbb{R}^d$, where the L2-norm $\|v\| = 1$ (true by definition)
- then if I have some data point $p \in \mathbb{R}^d$ . . .



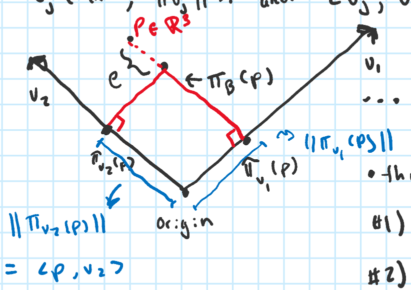
$\pi_v(p) = \langle v, p \rangle v$, where $\pi_v(p)$ is the closest point to p on a 1-dimensional subspace through v.

- the dot product $\langle v, p \rangle$ gives the length of p on v!
- multiplying by v scales $\langle v, p \rangle$ in the 1-D subspace.

However, in general we want a k-dimensional subspace.

Basis:

- A basis B is represented as $V_B = \{v_1, v_2, \ldots, v_k\}$, where each $v_j \in \mathbb{R}^d$, $\|v_j\| = 1$ and $\langle v_j, v_{j'} \rangle = 0$ (thus they are orthogonal),



. . . where $V_B = \{v_1, v_2\}$

- this projection can be thought of in two ways:

#1) $\pi_B(p) \in \mathbb{R}^d$

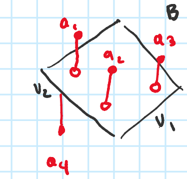
#2) $(\pi_{v_1}(p), \pi_{v_2}(p)) \in \mathbb{R}^k$

Our goal here is the "Sum of Squared Errors", SSE(A, B).

$$SSE(A,B) = \sum_{a_i \in A} \| a_i - \pi_B(a_i) \|^2$$

☆ Thus we want to find the k-dimensional subspace B s.t.

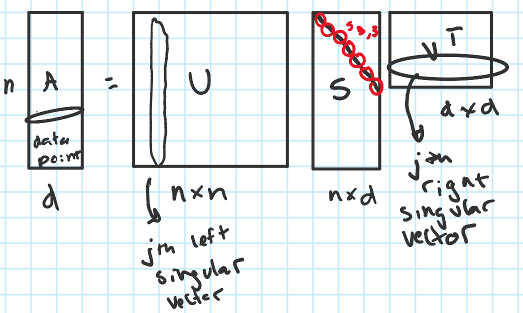
$$B^* = \operatorname{argmin}_B SSE(A,B)$$



- this is easy to solve for using SVD if we enforce that B^* contains the origin 0.
- note that multiple points p may project to the same spot.
  - thus some info can be lost

Singular Value Decomposition:

For $n \times d$ matrix A, define $[U, S, V] = \text{svd}(A)$ such that $USV^T = A$

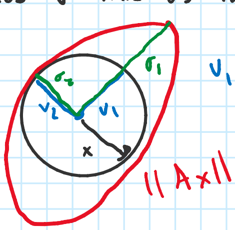


- some programming languages may give a more sparse representation of S because the only non-zero values are along the diagonal.
  - the i'th diagonal is the i'th singular value.
  - $\sigma_{i,i} = s_{i,i}$ = i'th singular value
  - $\sigma_j > \sigma_{j+1}$ and $\sigma_d \geq 0$

Orthonormal Properties:

$\|u_j\| = 1$, $\|v_j\| = 1$, $\langle v_j, v_j' \rangle = 0$ (thus, they form a basis)

- Thus V tells us about the d-dimensional space.



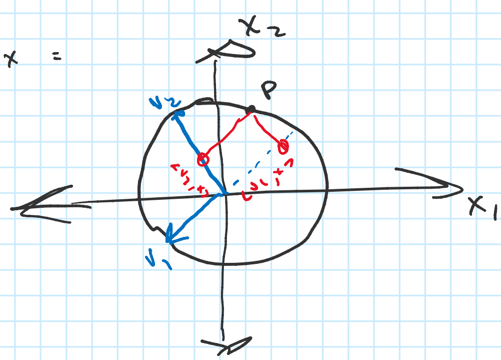
v_1, v_2 are the singular vectors

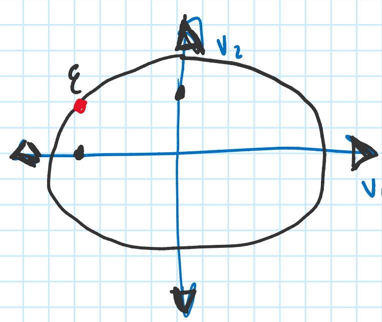
- there's more norm in the direction of v_1 associated with the shape A.
- these right singular vectors are capturing the directions which have the most variation about the origin of A
- S captures the amount of importance

Tracing a point through SVD:

$Ax = USV^Tx$

- V^Tx =





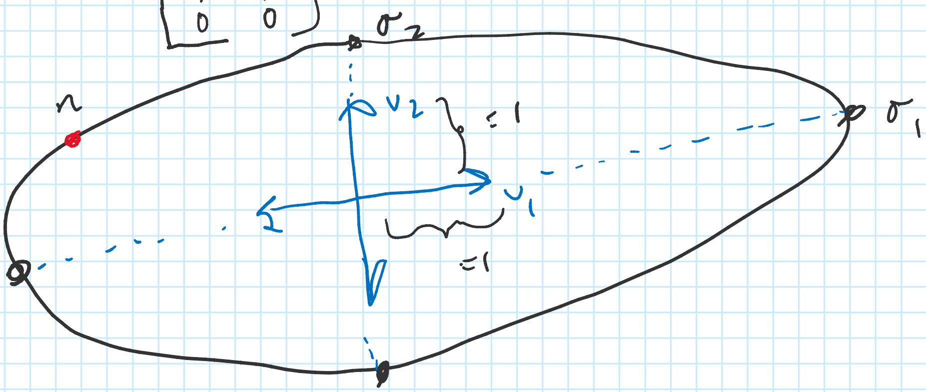
- this represents the change of basis

$$\begin{bmatrix} \langle v_1, x\rangle \\ \langle v_2, x\rangle \end{bmatrix} = \underbrace{\begin{bmatrix} V^T \end{bmatrix}}_{d\times d} \underbrace{\begin{bmatrix} x \end{bmatrix}}_{x\in\mathbb{R}^d}$$

- $SV^Tx = Sq$

$$S = \begin{bmatrix}\sigma_1 & 0\\ 0 & \sigma_2 \\ \vdots & \vdots \\ 0 & 0\end{bmatrix}$$

- S scales the ellipse in the correct directions of the basis



- U then rotates the final space into place.

This is the end result vector.

(but SV^Tx gives all aggregate information.