Singular Value Decomposition

-matrix of <u>n data points X m features</u>

-Goal: test this matrix and find its linear algebraic properties in order to

- -Approximate A with a smaller matrix B that is easier to store but contains similar information as A
- -Dimensionality Reduction / Feature Extraction
- -Anomaly Detection & Denoising

-Linear algebra review

- <u>no vector</u> in that set can be expressed as a <u>linear combination of other</u> vectors in the set

- determinant A:
  - <u>scalar value</u>
  - det(A) = ad – bc

-rank:
  - The dimension of the vector space spanned by its column space.
  This is equivalent to the maximal number of linearly independent columns / rows of A.
  - A matrix A is full-rank <u>iff rank(A) = min(m, n)</u>
  - We can get the rank of a matrix through the Gram-Schmidt process

-Approximation:

-Goal:

Approximate A with $A^{(k)}$ (low-rank matrix) such that

- $d(A, A^{(k)})$ is small
- k is small compared to m & n

-Frobenius Distance: the pairwise differences in values of A and B

$$d_F(A,B) = \|A-B\|_F = \sqrt{\sum_{i,j}(a_{ij}-b_{ij})^2}$$

-Approximation:

-When k < rank(A), the rank-k approximation of A (in the least squares sense) is

$$A^{(k)} = \underset{\{B|rank(B)=k\}}{\arg\min} \; d_F(A,B)$$

- The Singular Value Decomposition of a rank-r matrix A has the form $A = U\boldsymbol{\Sigma}V^T$

-Property: the larger **k** is, the smaller the distance.

$$d_F(A, A^{(k)})^2 = \sum_{i=k+1}^{r} \sigma_i^2$$

-Latent Semantic Analysis

- Inputs are documents. Each word is a feature. We can represent each document by:

- The presence of the word (0 / 1)
- Count of the word (0, 1, …)
- Frequency of the word (n_i / $\boldsymbol{\Sigma} n_i$)
- TfiDF:

  Term frequency in the document

  log (number of documents /number of documents that contain the term)

-see slides for e.g.

-Anomaly detection

-Define $O = A - A^{(k)}$

-The largest rows of O could be considered anomalies