Oct 6^{th}

Singular Value Decomposition

goal
- approx A w/ smaller B easier to store
- Dimensionality Reduction | Feature Extraction
- Anomaly Detection & Denoising

Linear algebra review

Linear independent - cannot represent any col as linear combo of other vectors in a set iff $A = [v_1 \cdots v_n]_{(n \times n)}$ has nonzero determinant vectors are linearly independent

Determinant - of a square matrix A is a scalar value that encodes properties of linear mapping

2×2 ad - bc

3×3 $a \cdot \det\begin{pmatrix} e & f \\ h & i \end{pmatrix} - b \cdot \det\begin{pmatrix} d & f \\ g & i \end{pmatrix} + c \cdot \det\begin{pmatrix} d & e \\ g & h \end{pmatrix}$

Note : m>n vectors in n-dim cannot be linearly independent

Basis - of a vector space is a linearly independent subset of V that spans V.

Rank - max number of linearly independent cols of A.

↳ Full Rank iff rank(A) = min(m,n)

↳ get the rank w/ Gram-Schmidt

Approximation

~ in practice dataset contains a lot of redundant data.

$mn \leftarrow A_{(n \times m)} = U_{(n \times k)} V_{(k \times m)} \rightarrow k(m+n)$

A w/ $A^{(k)}$ approx

- $d(A, A^{(k)})$ is small
- k is comparably smaller than m,n

FROBENIUS DISTANCE

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

* A, B must be same size.

when k < rank(A)

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

SVD of rank-r

$$A_{(m \times n)} = \underset{(n \times r)}{U} \; \underset{(r \times r)}{\Sigma} \; \underset{(r \times m)}{V^T}$$

U: orthogonal, Σ: diagonal, V^T: orthogonal

σ_i is the sq. rt of eigenvalues of A^TA

$$\begin{bmatrix} \sigma_1 & & 0 \\ & \sigma_2 \ddots & \\ 0 & & \sigma_r \end{bmatrix}$$

* Note! these are ordered.

$\sigma_1 \geq \sigma_2 \geq \sigma_r > 0$

↓ SINGULAR VALUES

PROPERTY

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

all the singular columns we dropped (r-k)

$$A = \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} \Sigma_1 & \\ & \Sigma_2 \end{pmatrix} \begin{pmatrix} V_1 \\ V_2 \end{pmatrix}$$

$U_2, \Sigma_2, V_2 \longrightarrow$ made 0.

The i^{th} singular vector represents the direction of i^{th} most variance

singular values express the importance / significance of a singular values.

Principal Component Analysis (PCA)

project data onto a subspace k

* SHOULD DEF NORMALIZE.

Latent Semantic Analysis

apply to specific domain e.g. DOCUMENTS

REPRESENT BY
- presence of the word (0/1)
- count of word (0, 1...)
- rel. freq of word ($n_i / \sum n_i$)
- TFiDF (term freq · log(num doc / num doc that contain term))

DOC TO CONCEPT SIMILARITY (U)

STRENGTH OF CONCEPT (Σ)

TERM TO CONCEPT SIM (V^T)