Lecture 8 Singular Value Decomposition

- Examine this matrix and uncover its linear algebraic properties to:
  - 1. Approximate A with a smaller matrix B that is easier to store but contains similar information as A
  - 2. Dimensionality Reduction / Feature Extraction
  - 3. Anomaly Detection & Denoising
- Linearly Independent vectors
  - Vecotrs V = { V1,… Vn}   are linearly independent if aV1 + … aVn = 0 vector
  - This can only be satisfied by ai= 0
  - This means that no vetor in that set can be expressed as a linear combinator of other vectors in that set
- Determinant
  - The determinant of a square matrix A is a scalar value that encodes properties about the linear mapping described by A.
  - $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ det(A) = ad - bc
  - $A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \quad det(A) = a \cdot det\begin{pmatrix} e & f \\ h & i \end{pmatrix} - b \cdot det\begin{pmatrix} d & f \\ g & i \end{pmatrix} + c \cdot det\begin{pmatrix} d & e \\ g & h \end{pmatrix}$
  - N vectors V = { V1,… Vn} in an n dimensional space are linearly independent iff the matrix A = = { V1,… Vn}(n x n) has a non -zero determinant
- Rank
  - The rank of a matrix A is the dimension of the vector space spanned by its column space. This is equal to the maximal number of linearly independent columns /rows of A
  - Full Rank : A Matrix A is full rank iff rank(A) = min(m,n)
    - Most datasets are full rank despite containing a lot of redundant /similar
  - You can calculate the rank of a matrix through the Gram-Schmidt Process
- Matrix Factorization
  - Any matrix A of rank k can be factored as A = UV
    - Where U is n x k
    - And Where V = k x m
- Forbeanius Distance
  - $$d_F(A, B) = \|A - B\|_F = \sqrt{\sum_{i,j}(a_{ij} - b_{ij})^2}$$
  - The pairwise sum of squares difference in values of A and B
- Approximation
  - When k < rank (A) the rank -k approximation of A is
    - $$A^{(k)} = \underset{\{B|rank(B)=k\}}{\arg\min} \; d_F(A, B)$$
- Matrix Factorization Improved

- Not only can we factorize a matrix A of rank k as A = UV. But we can factorize A using a process called Singular Value Decomposition where $A = U\mathbf{\Sigma}V^T$
- where U is n x r
  - The columns of U are orthogonal & unit length ($U^TU = I$)
- Where V is m x r
  - The columns of V are orthogonal & unit length ($V^T V = I$)

- SVD
  - Data reduction tool -> helping reduce data into key features analyzing and describing data that can then be used to model data
  - Data driven generalization
  - Google search , facial recognition , Netflix (recommending shows -> correlations)
  - Reshaping data into column vectors to create a matrix
    - $U\Sigma V^T$