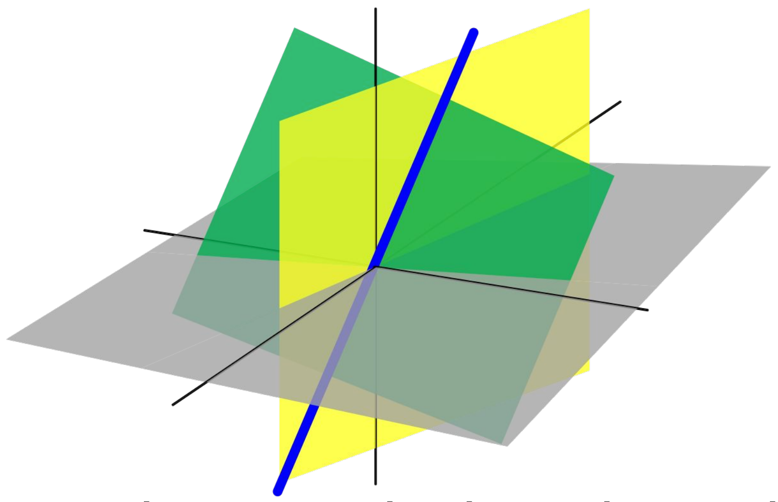


Advanced Linear Algebra Cheat Sheet

V2020.10.21

(Dr Yan Xu)

Vector Norm

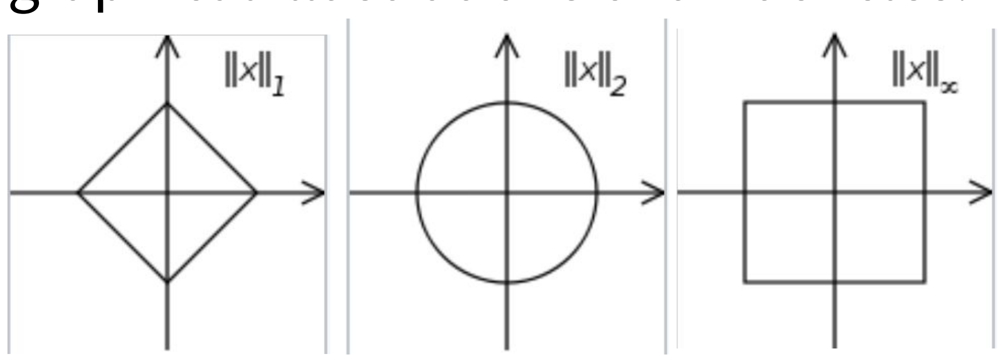
➔ ***p*-norm:** $\|\mathbf{x}\|_p := \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$

➔ ***p=1*** => **L**1 norm: $\|\boldsymbol{x}\|_1 := \sum_{i=1}^{n} |x_i|$

➔ ***p=2*** => **L**2 norm: $\|\boldsymbol{x}\|_2 := \sqrt{x_1^2 + \cdots + x_n^2}$

➔ ***P*->∞** => **Infinity** norm: $\|\mathbf{x}\|_\infty := \max\left(|x_1|, \ldots, |x_n|\right)$

➔ graphical illustrations of unit circles:



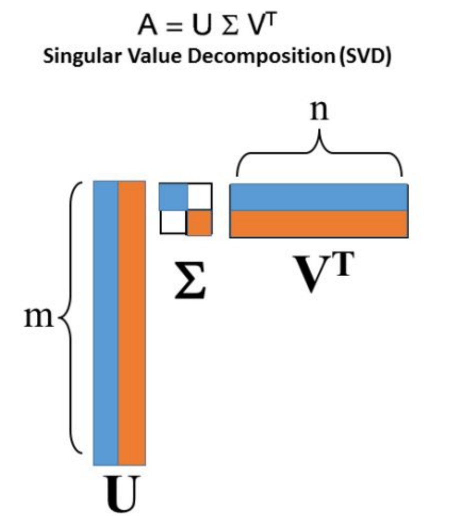
Eigen Decomposition

$$\mathbf{A} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^{-1}$$

where,
A: ***n* × *n*** matrix, with rank = n
Q: n × n matrix whose ith column is the eigenvector of A
Λ: diagonal matrix whose elements are the eigenvalues

1: Not all matrix (rank = n) could be eigen decomposed.
2: The product of the eigenvalues is equal to the determinant of A.
3: The sum of the eigenvalues is equal to the trace of A.
4: Not very popular in ML given it requires a *n x n* matrix.

Singular Value Decomposition



$\mathbf{A}_{[m \times n]} = \mathbf{U}_{[m \times r]}\, \Sigma_{[r \times r]}\, (\mathbf{V}_{[n \times r]})^T$

where,
r: rank of ***A***
Σ: ***r* × *r*** diagonal matrix, values as singular values of A
U: columns are eigenvectors of $\boldsymbol{AA}^T$ (orthonormal vectors)
V: columns are eigenvectors of $\boldsymbol{A}^T\boldsymbol{A}$ (orthonormal vectors)

1: Elements in **Σ** are also square roots of eigenvalues
2: Given any matrix **A** (m * n), SVD always exist.
3: Uniqueness: **Σ**: Yes, **U** and **V**: not always
4: Geometric meaning:
U and **V**: rotation; **Σ**: scaling and dimension change
5: Computation complexity: m^2n (m > n)

Popular applications in ML:
1: Pseudoinverse of a matrix (in regression models)
2: Low-rank matrix approximation (e.g. PCA)

Low-rank Matrix Approximation

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \ldots + \sigma_r u_r v_r^T$$

$$\mathbf{B} = \mathbf{U}\boldsymbol{\Sigma}_B\mathbf{V}^T = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \ldots + \sigma_k u_k v_k^T$$

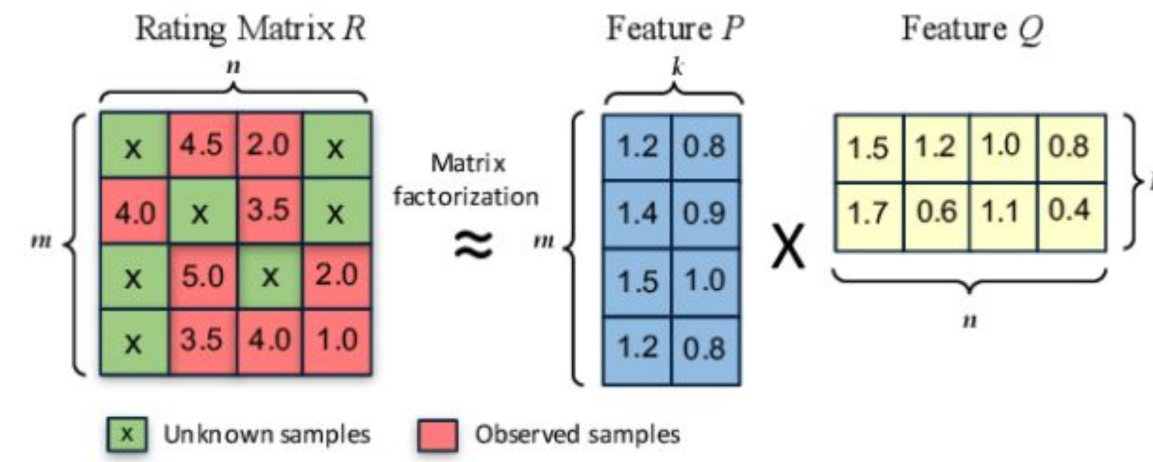
where, $\boldsymbol{\Sigma}_B$: only contains the ***k*** largest singular values in **Σ**

B is the optimal approximation of A with rank k (k <= r), which minimizes the L2 norm of the difference.

Pseudoinverse (Moore–Penrose inverse)

1: The pseudoinverse is defined and unique for all matrices
2: If A has linearly independent columns, $\mathbf{A}^+ = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$
3: The pseudoinverse of the pseudoinverse is the original matrix: $(\mathbf{A}^+)^+ = \mathbf{A}$
4: It could generated from its SVD, by taking the reciprocal of each non-zero singular values in **Σ**

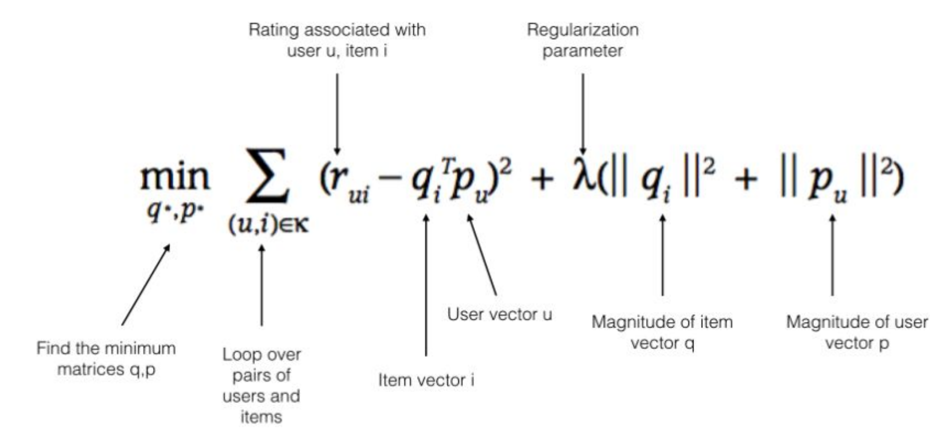
Matrix Factorization



R' = P*Q

where, k is the number of latent factors.

1: if k = 1, equivalent to recommends the items with the most interactions without any personalization.
2: if k is large (~n), it encourages overfit. So, it is critical to add regularization terms to the objective function.



Popular applications in ML:
1: Recommendation Systems