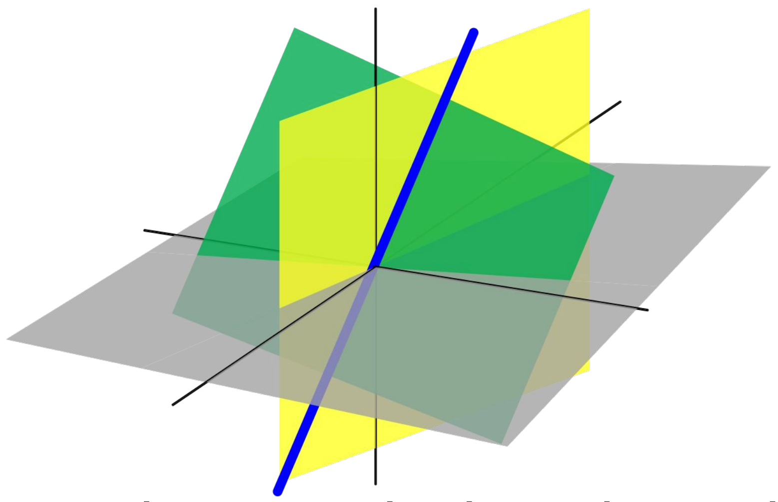


# Advanced Linear Algebra Cheat Sheet

V2020.10.21

(Dr Yan Xu)

## Vector Norm

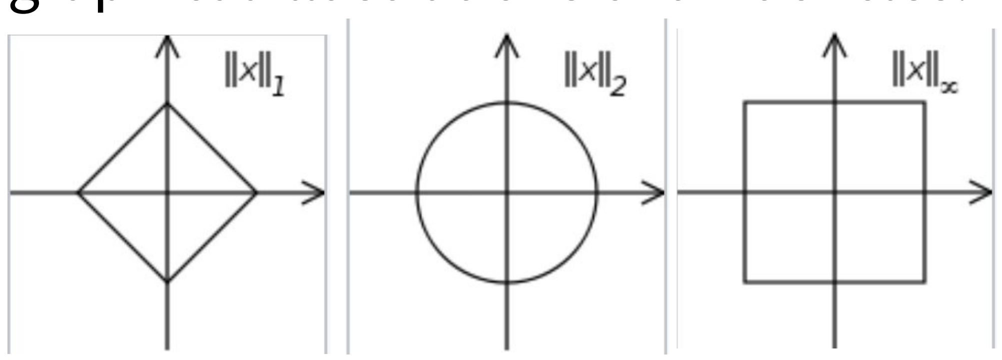
➔ ***p*-norm:** $\|\mathbf{x}\|_p := \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$

➔ ***p=1*** => **L1** norm: $\|\boldsymbol{x}\|_1 := \sum_{i=1}^{n} |x_i|$

➔ ***p=2*** => **L2** 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

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

where,

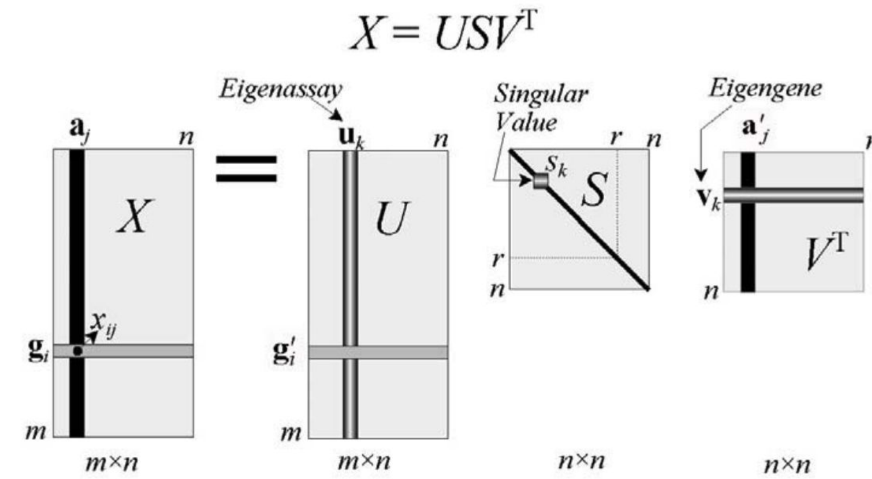
**A:** ***n* × *n*** matrix, with rank = n

**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




where,

**S:** ***n* × *n*** diagonal matrix, values as singular values of **X**

**U:** columns are eigenvectors of $\boldsymbol{XX^T}$ (orthonormal vectors)

**V:** columns are eigenvectors of $\boldsymbol{X^TX}$ (orthonormal vectors)

***r***: rank(X), elements in S with column > r are 0

1: Elements in **S** are also square roots of eigenvalues; when squared are proportional the amount of variance explained.
2: Given any matrix **X** (m * n), SVD always exist.
3: Uniqueness: **S:** Yes, **U** and **V:** not always
4: Geometric meaning:
**U** and **V:** rotation; **S:** scaling and dimension change
5: Computation complexity: $m^2n$ (m > n)
6: Assuming: m > n

Popular applications in ML:
1: Dimension Reduction (e.g. PCA)
2: Pseudoinverse of a matrix (in regression models)

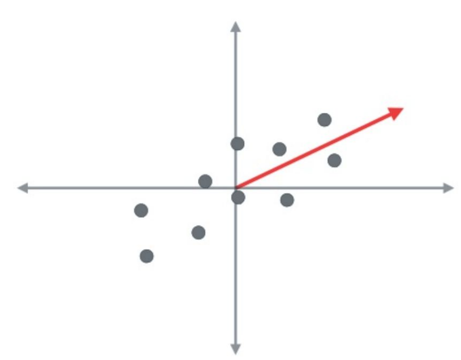
## Principal Components Analysis

➔ Problem:

$$M : X = \mathbb{R}^D \to \mathbb{R}^k, \quad k \ll D,$$

➔ Motivation:

1. min(projection-distance)
   or,
2. min(variance loss)



➔ Cost Function:

$$J(W, Z) = \left\|X - ZW^T\right\|_F^2$$

where,

W is top K principal components (or dimensions)
Z is the compressed (or low-dimensional) matrix
$WZ^T$ is the approximated original size matrix

## Principal Components Analysis (continued)

➔ SVD(X): $\underbrace{\mathbf{X}}_{N \times D} = \underbrace{\mathbf{U}}_{N \times D}\underbrace{\mathbf{S}}_{D \times D}\underbrace{\mathbf{V}^T}_{D \times D}$

➔ Rank *k* Approximation using SVD(X):

$$\underbrace{\mathbf{X}}_{N \times D} = \underbrace{\hat{\mathbf{U}}}_{N \times K}\underbrace{\hat{\mathbf{S}}}_{K \times K}\underbrace{\hat{\mathbf{V}}^T}_{K \times D}$$

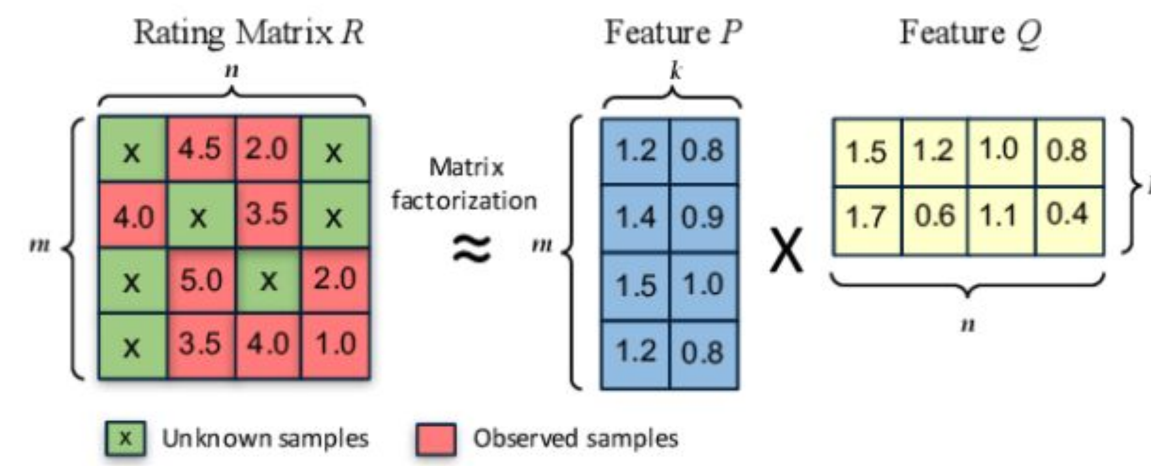
➔ Solution based on SVD:

$$W = \hat{V} \ and \ Z = \hat{U}\hat{S}$$

➔ **X** needs to be normalized before PCA, so that features' variances equally measured.

➔ Code Example

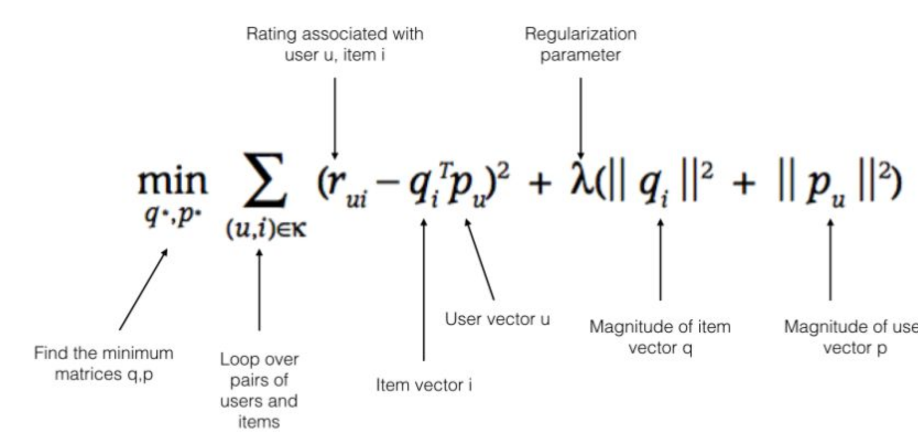
## Matrix Factorization




**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.




$$\min_{q^*,p^*} \sum_{(u,i)\in\kappa} (r_{ui} - q_i^T p_u)^2 + \lambda(\|q_i\|^2 + \|p_u\|^2)$$

Popular applications in ML:
1: Recommendation Systems