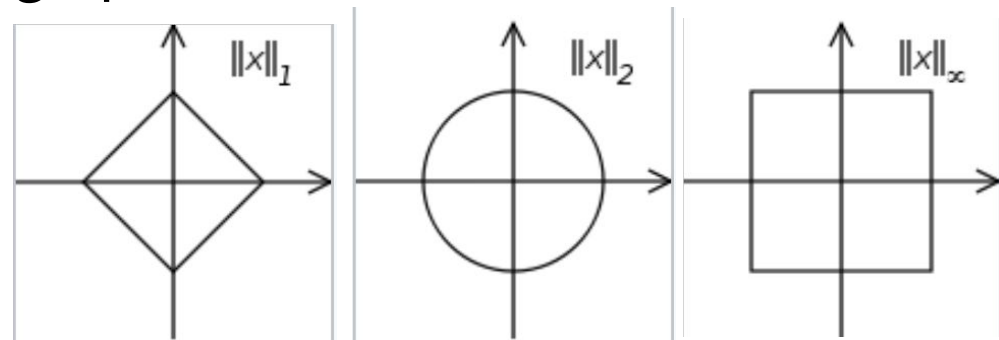


## Advanced Linear Algebra Cheat Sheet

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:**  $\|\mathbf{x}\|_1 := \sum_{i=1}^n |x_i|$
- **p=2 => L2 norm:**  $\|\mathbf{x}\|_2 := \sqrt{x_1^2 + \dots + x_n^2}$
- **p->∞ => Infinity norm:**  $\|\mathbf{x}\|_\infty := \max(|x_1|, \dots, |x_n|)$
- graphical illustrations of unit circles:



### Eigen Decomposition

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

where,

**A:**  $n \times n$  matrix, with rank = n

**Q:**  $n \times n$  matrix whose  $i$ th 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  $n \times n$  matrix.

### Singular Value Decomposition

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T$$

where,

**S:**  $n \times n$  diagonal matrix, values as singular values of **X**

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

**V:** columns are eigenvectors of  $\mathbf{X}^T\mathbf{X}$  (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** ( $m \times n$ ), 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)

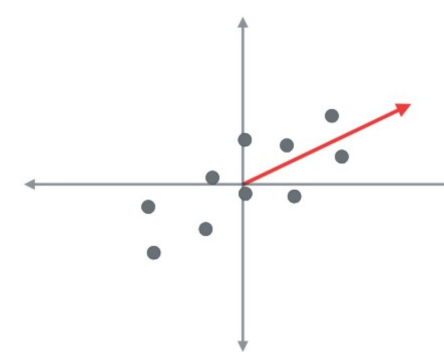
### Principal Components Analysis

→ Problem:

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

→ Motivation:

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



→ Cost Function:

$$J(\mathbf{W}, \mathbf{Z}) = \|\mathbf{X} - \mathbf{Z}\mathbf{W}^T\|_F^2$$

where,

**W** is top K principal components (or dimensions)

**Z** is the compressed (or low-dimensional) matrix

$\mathbf{W}\mathbf{Z}^T$  is the approximated original size matrix

### Principal Components Analysis (continued)

→ SVD(X):  $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T$

$N \times D \quad N \times D \quad D \times D \quad D \times D$

→ Rank k Approximation using SVD(X):

$$\mathbf{X} = \hat{\mathbf{U}}\hat{\mathbf{S}}\hat{\mathbf{V}}^T$$

$N \times D \quad N \times K \quad K \times K \quad K \times D$

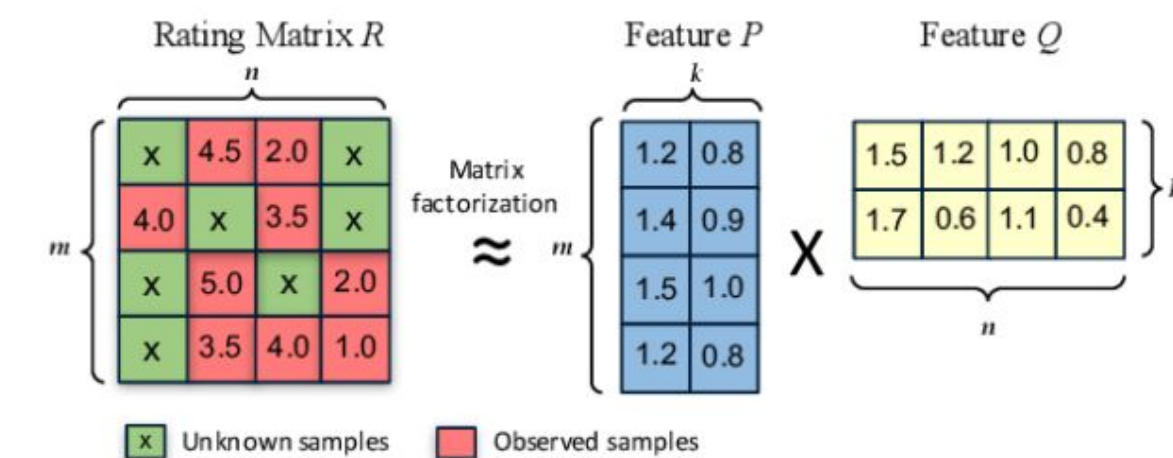
→ Solution based on SVD:

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

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

→ [Code Example](#)

### Matrix Factorization



$$\mathbf{R}' = \mathbf{P} * \mathbf{Q}$$

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 ( $\sim n$ ), it encourages overfit. So, it is critical to add regularization terms to the objective function.

$$\min_{\mathbf{q}, \mathbf{p}} \sum_{(u,i) \in K} (r_{ui} - q_i^T \mathbf{p}_u)^2 + \lambda (\|\mathbf{q}_i\|^2 + \|\mathbf{p}_u\|^2)$$

Rating associated with user u, item i      Regularization parameter  
 Find the minimum matrices q, p      Loop over pairs of users and items      Item vector i      User vector u      Magnitude of item vector q      Magnitude of user vector p

#### Popular applications in ML:

- 1: Recommendation Systems