



Deep Learning for Healthcare

3. Machine Learning Basics

Prof. Jimeng Sun

Unsupervised learning



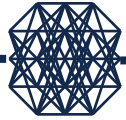
- Dimensionality Reduction
- Clustering methods

Dimensionality Reduction

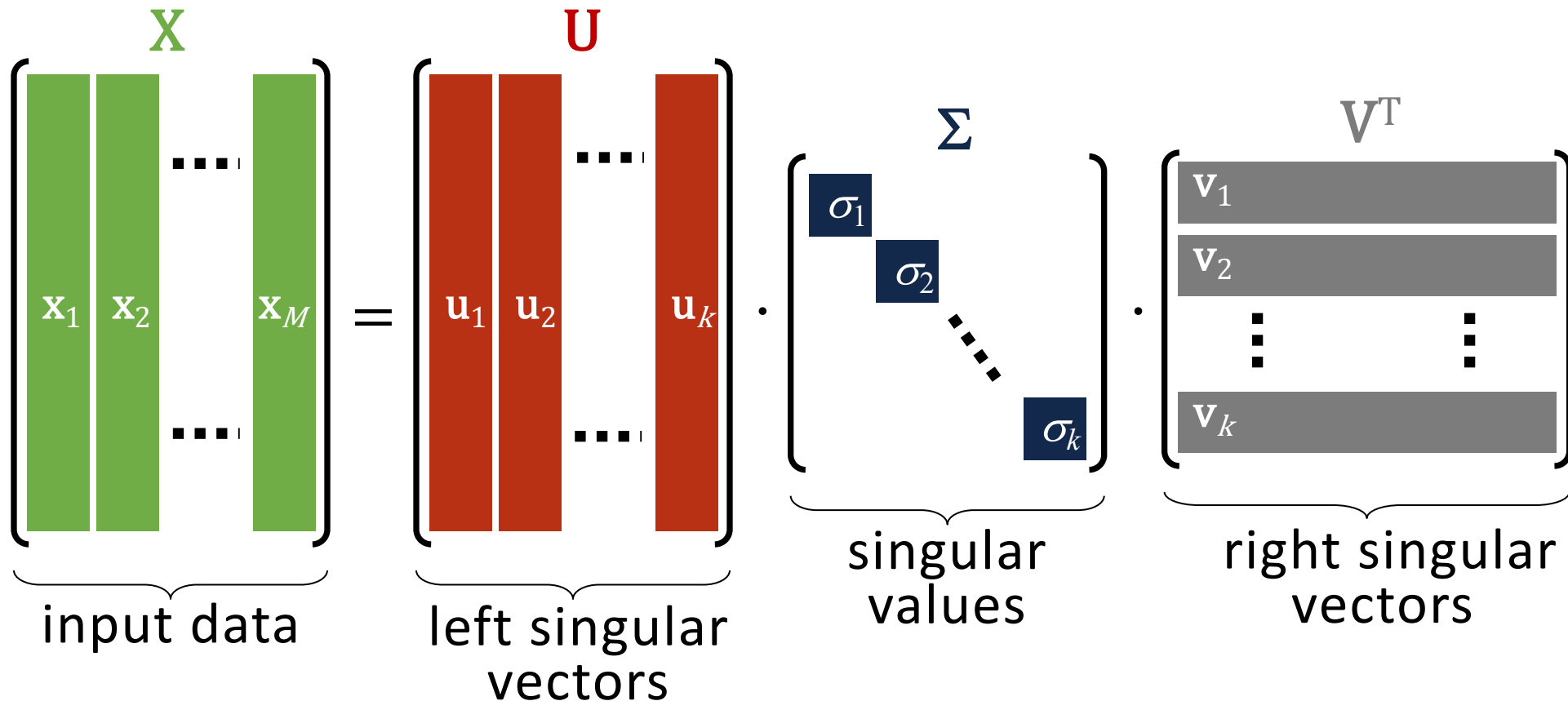


- Singular Value Decomposition (SVD)
- Principal Component Analysis (PCA)

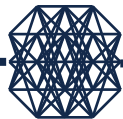
Singular Value Decomposition (SVD)



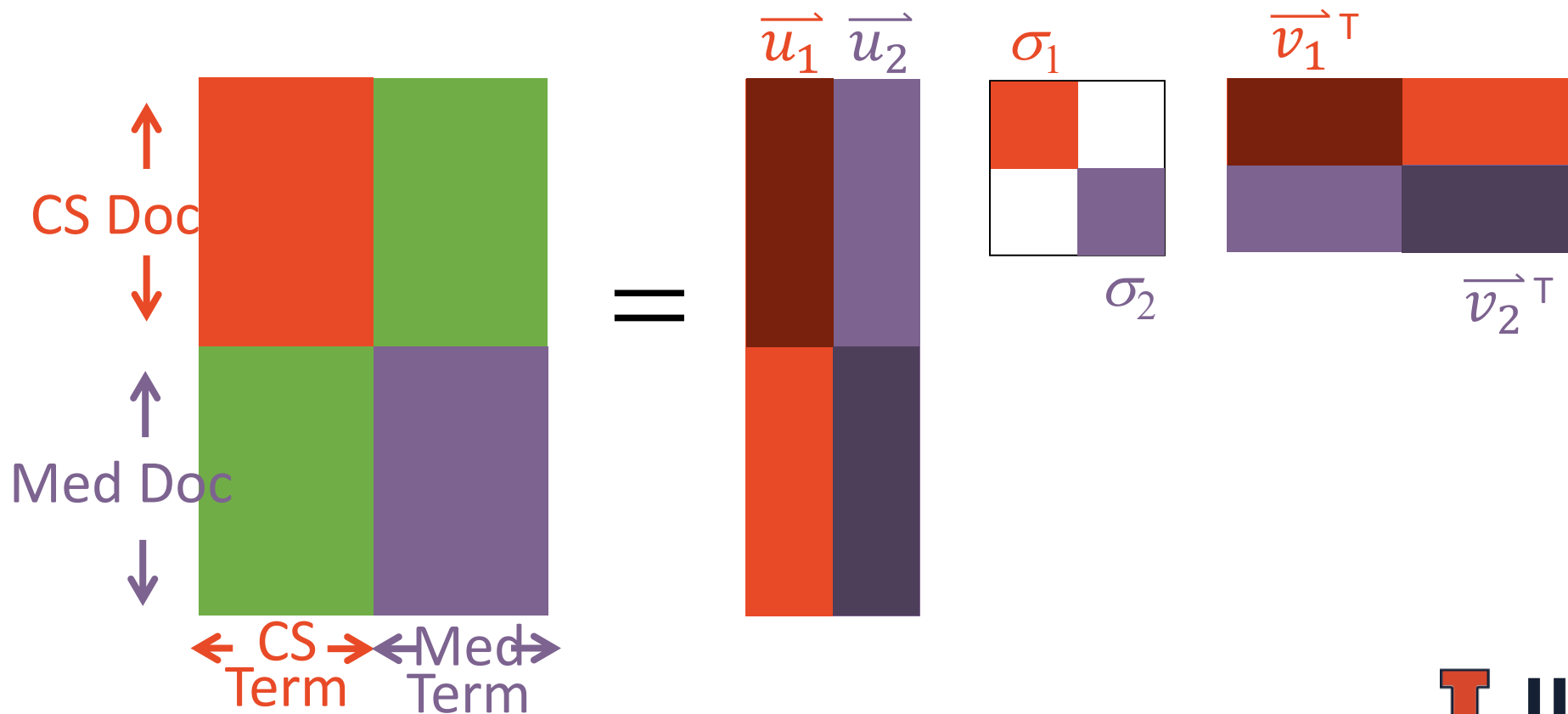
$$X = U \Sigma V^T \quad \text{where} \quad U^T U = V^T V = I$$



SVD Example



$$X = \underbrace{U\Sigma V^T}_{\text{matrix view}} = \underbrace{\sigma_1 \vec{u}_1 \vec{v}_1^T + \sigma_2 \vec{u}_2 \vec{v}_2^T + \dots}_{\text{spectral view}}$$



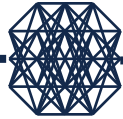
Quiz: SVD - Interpretation



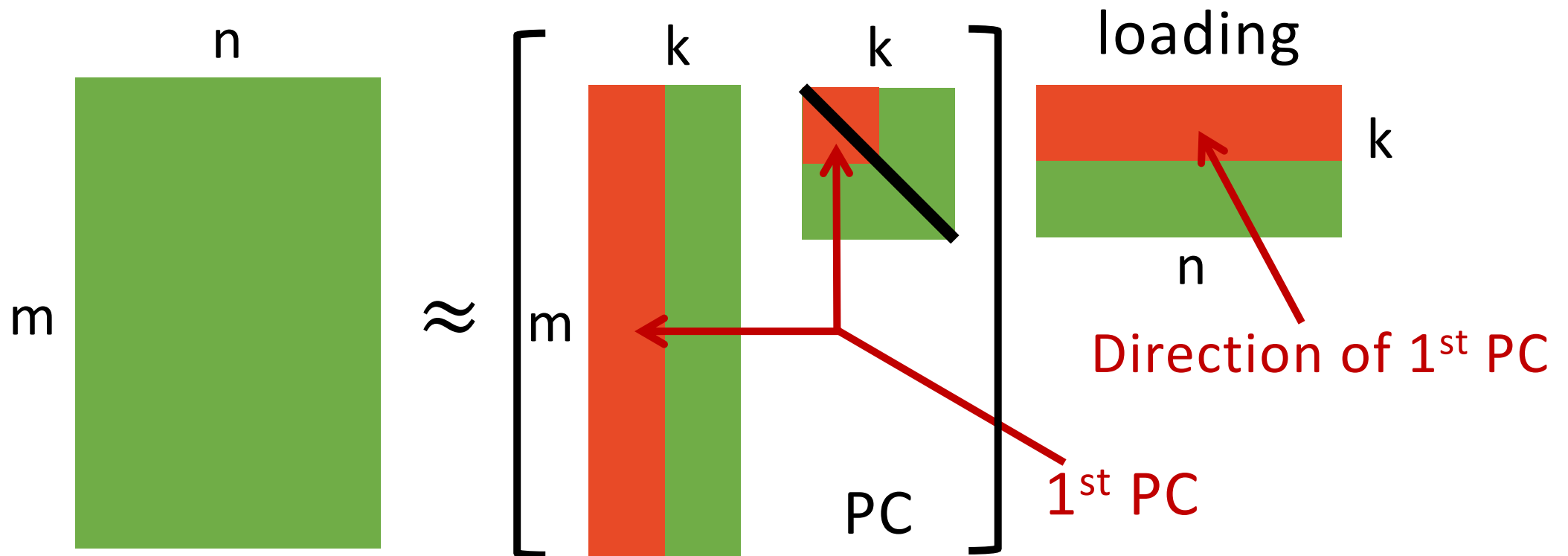
Given a document-by-term matrix A , what is $A^T A$?

- ☐ document-to-document similarity matrix
- ☒ term-to-term similarity matrix
- ☐ term-to-document similarity matrix

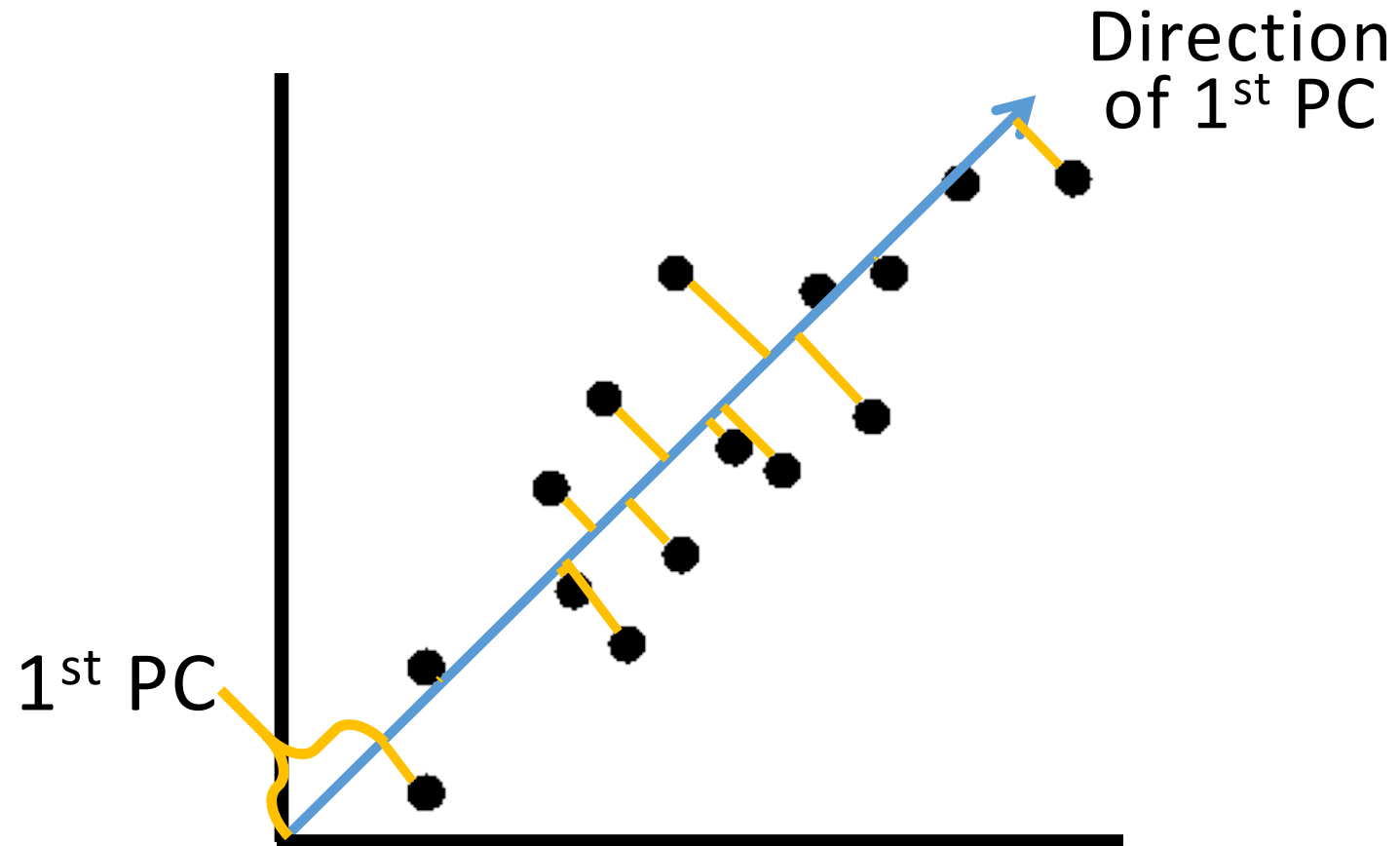
Principal Component Analysis (PCA)



$$X = (\underbrace{U}_{\text{PC}} \Sigma) \underbrace{V^T}_{\text{loading}}$$



PCA INTERPRETATION



What is Clustering?



Patient
(P)

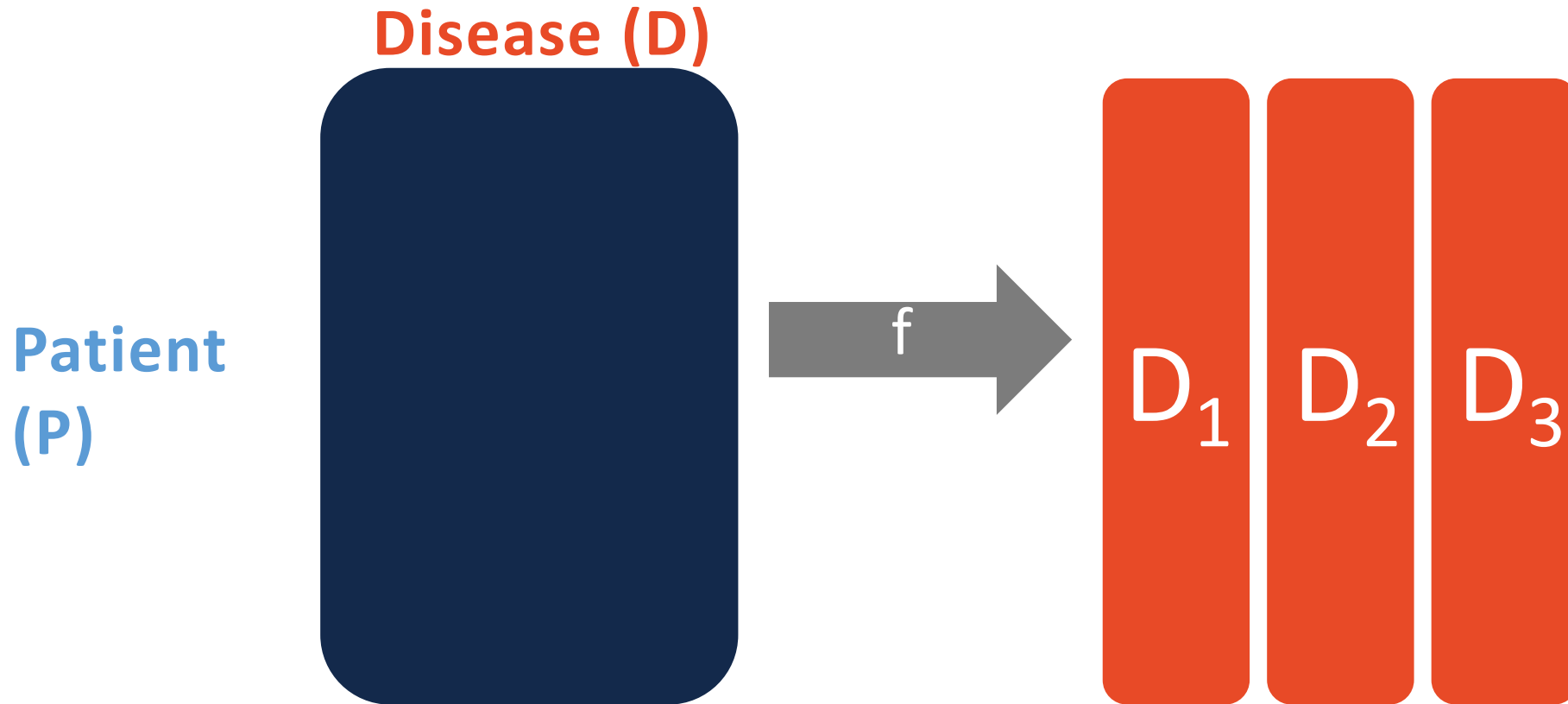


P_1

P_2

P_3

What is Clustering?



K-means



INPUT



Data points $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$,
of clusters k

k clusters S_1, S_2, \dots, S_k

OUTPUT



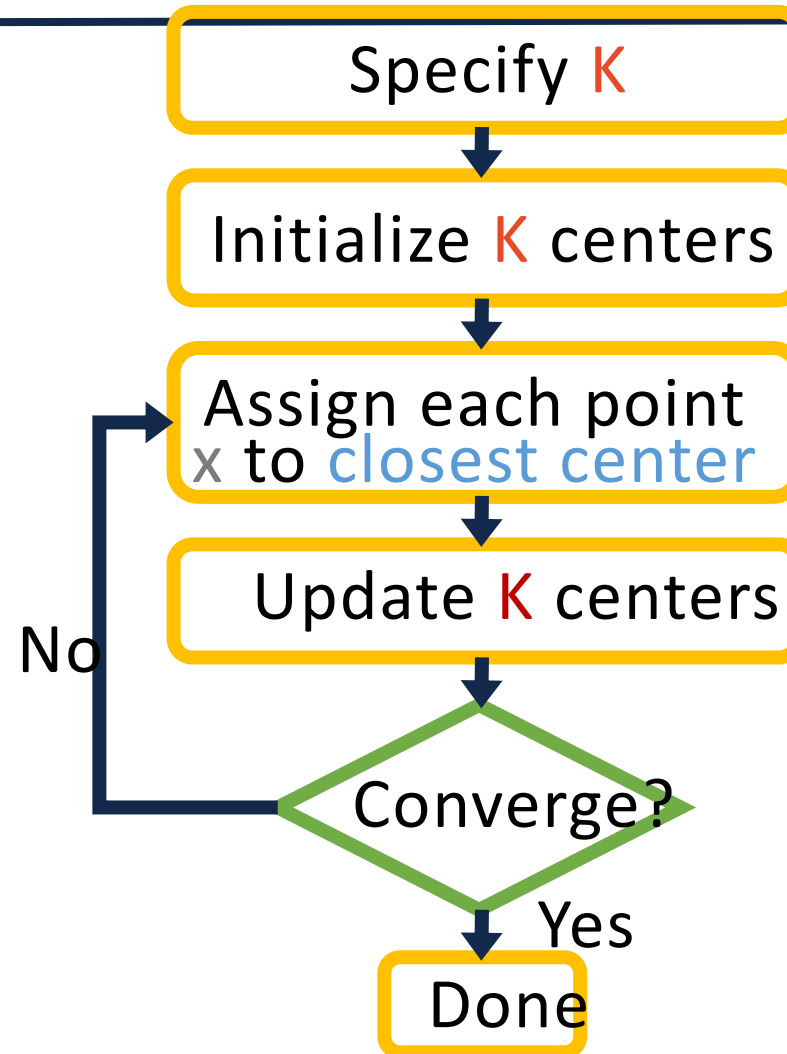
Objective

:

$$\min \sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_i\|^2$$

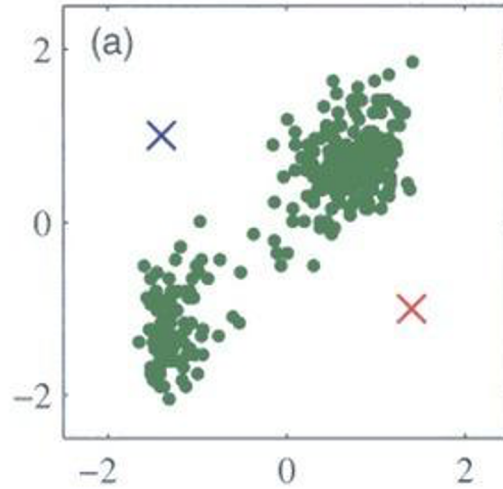
μ_i is the center in S_i

K-means Algorithm



$$\mu_i = \frac{1}{|S_i|} \sum_{x \in S_i} x$$

K-means Example



K-means Quiz



Given

- **n**: # of points
- **k**: clusters
- **d**: dimensionality of each point
- **i**: number of iterations

What is the computational complexity?

$$O(n \cdot k \cdot d \cdot i)$$

