Machine Learning

Abdelhak Mahmoudi
abdelhak.mahmoudi@um5.ac.ma

INPT- 2020

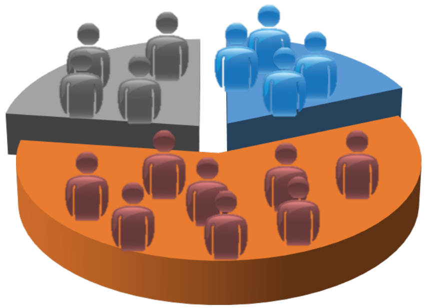
Content

1. The Big Picture
2. Supervised Learning
   - Linear Regression, Logistic Regression, Support Vector Machines, Trees, Random Forests, Boosting, Artificial Neural Networks
3. Unsupervised Learning
   - Principal Component Analysis, K-means, Mean Shift

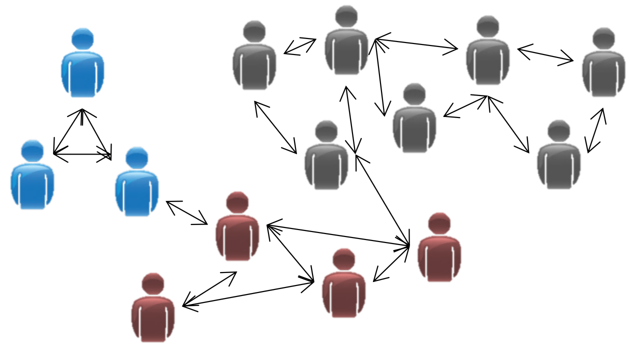
Unsupervised Learning

- **Dimensionality Reduction**
  - **Principal Component Analysis (PCA)**
- Clustering
  - K-Means
  - Mean-Shift

Unsupervised Learning

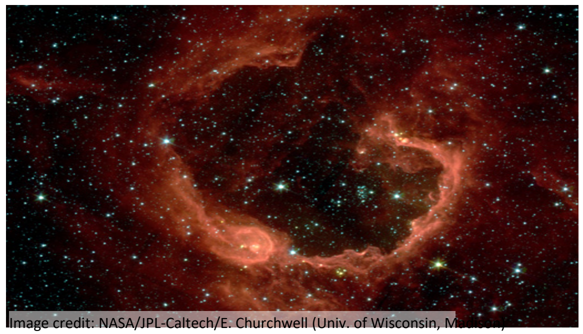


Market segmentation



Social network analysis

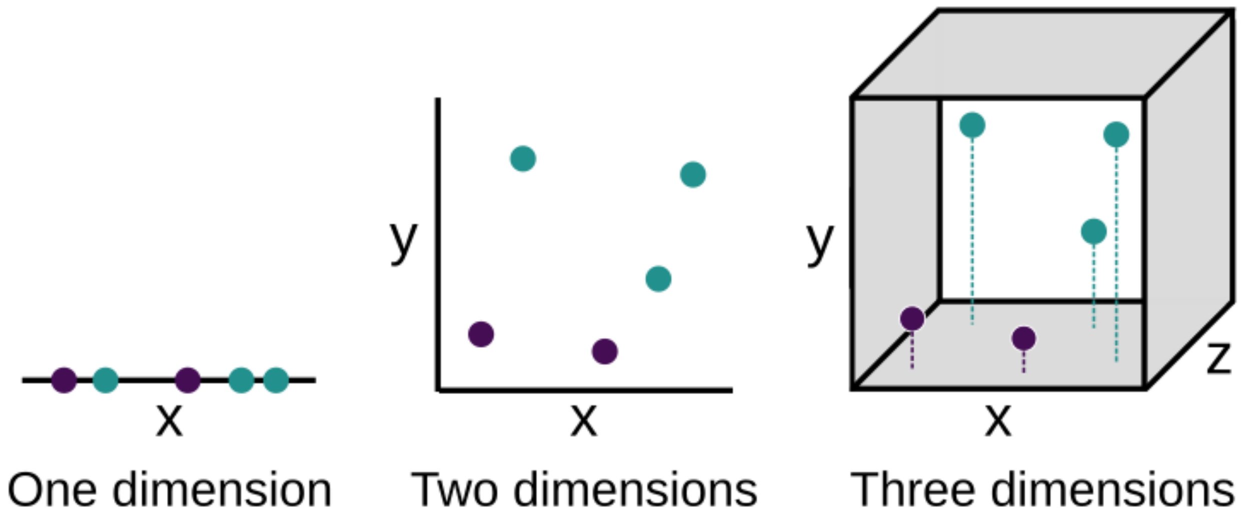
Organize computing clusters



Astronomical data analysis

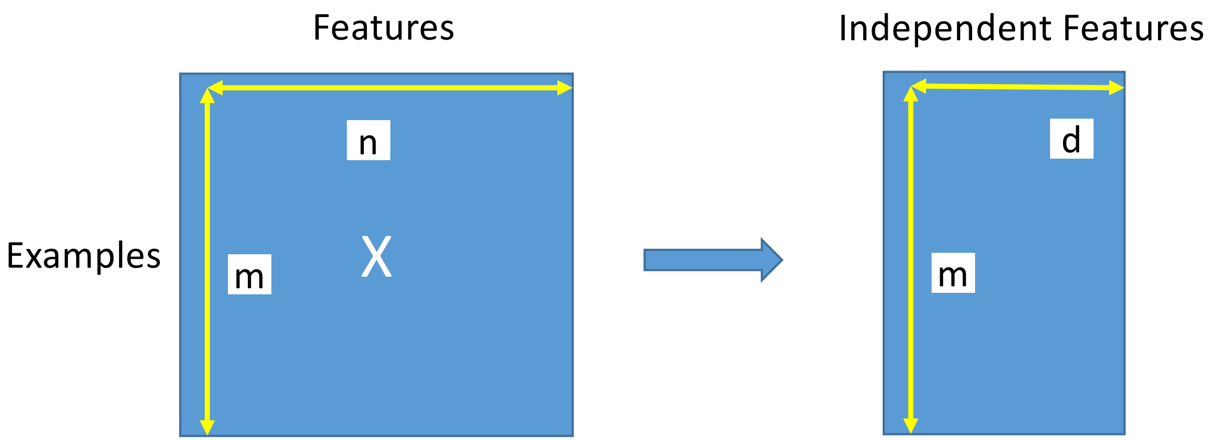
Dimensionality Reduction

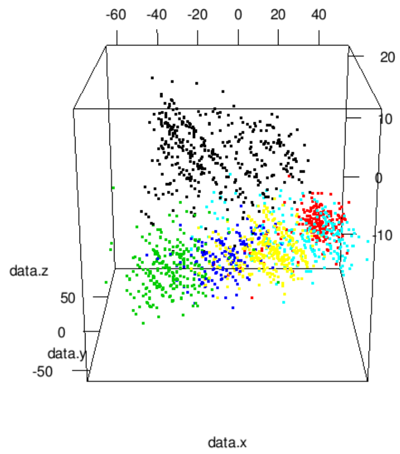
- Curse of dimensionality (n >> m)
  - Data are at risk of being very sparse in high dimensional space
  - High risk of overfitting



Dimensionality Reduction

- Transforms feature space from n to k (k<n)
  - Some features are probably corelated (dependent)
  - Some features are almost constant
  - Transform but preserve the maximum of variance

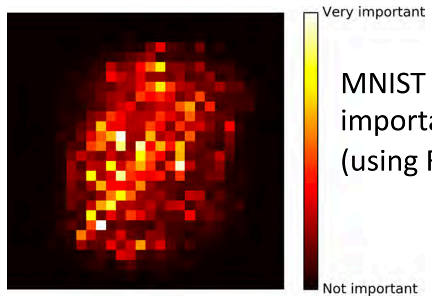




Dimensionality Reduction

- Transforms feature space from n to k
  - Some features are probably corelated (dependent)
  - Some features are almost constant
  - Transform but preserve the maximum of variance
- MNIST Example
  - Dropping some pixels without losing much information
  - Two neighboring pixels are often highly correlated -> use their average.

MNIST



MNIST pixel importance (using RF)

Dimensionality Reduction

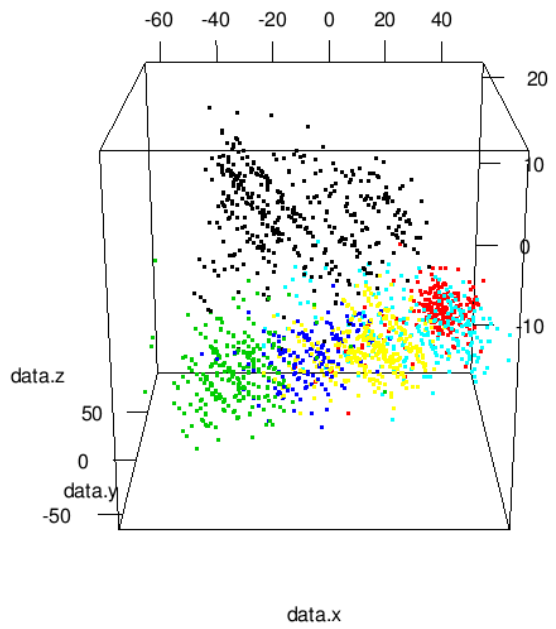
- Transforms feature space from n to k
  - Some features are probably corelated (dependent)
  - Some features are almost constant
  - Transform but preserve the maximum of variance
- EOCD example

Country	GDP (trillions of US$)	Per capita GDP (thousands of intl. $)	Human Dev Index	Life expectancy	…
Canada	1.577	39.17	0.908	80.7	…
China	5.878	7.54	0.687	73	…
India	1.632	3.41	0.547	64.7	…
Russia	1.48	19.84	0.755	65.5	…
Singapore	0.223	56.69	0.866	80	…
…	…	…	…	…	

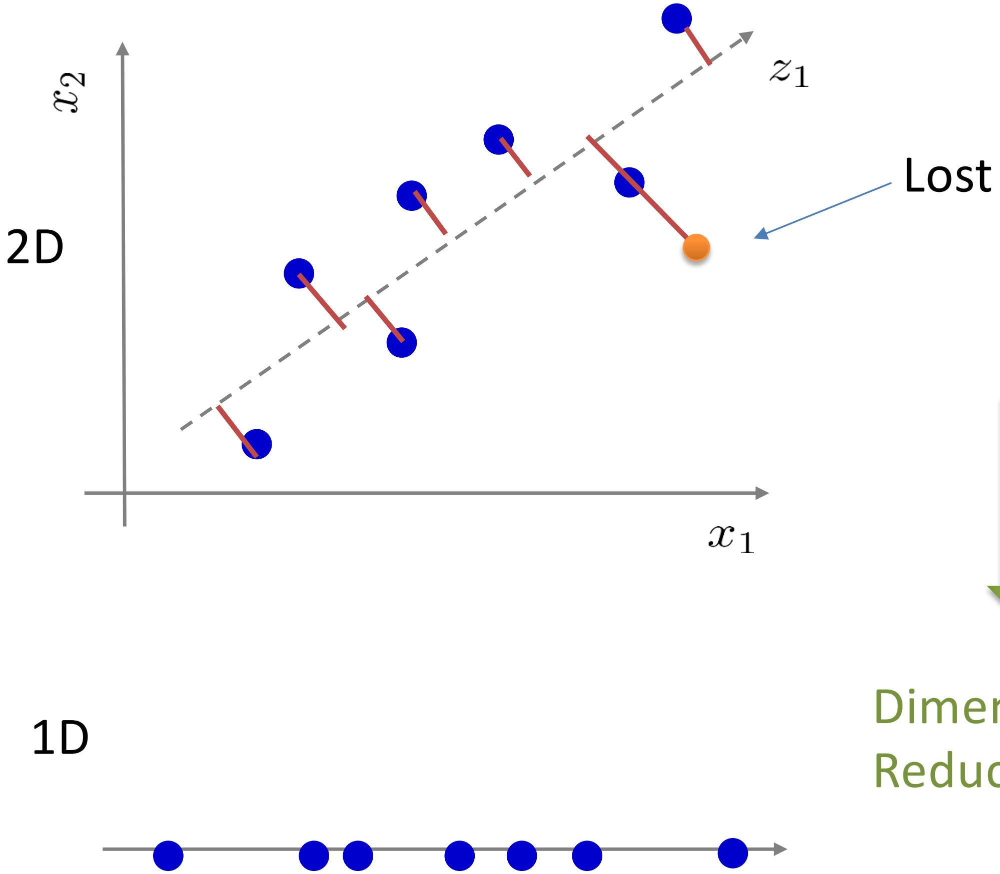
Country	z1	z2
Canada	1.6	1.2
China	1.7	0.3
India	1.6	0.2
Russia	1.4	0.5
Singapore	0.5	1.7
…	…	…

Dimensionality Reduction

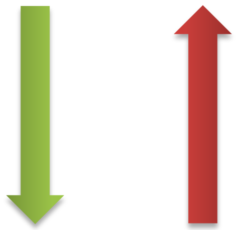
- Often
  - Not necessarily lead to better performance
  - Not the better way to address overfitting !
- Always
  - Speed up training
  - Allow data compression
  - Allow data exploration
  - Allow data visualization (DataViz)



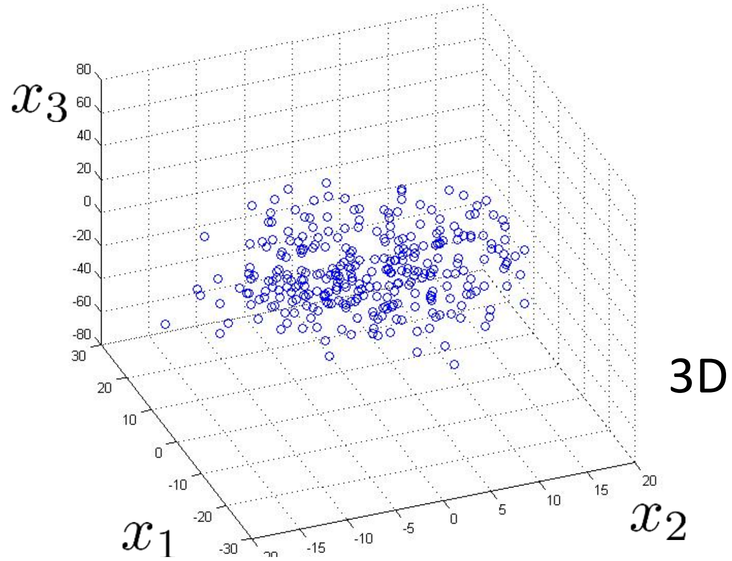
Principal Component Analysis

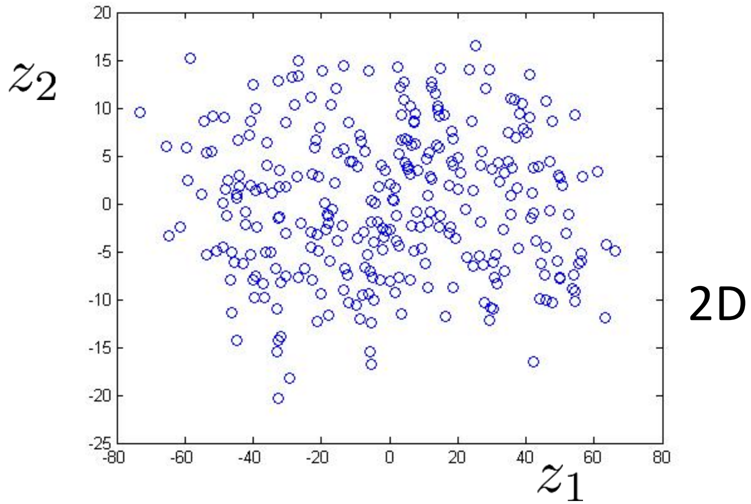


Reconstruction
With Loss

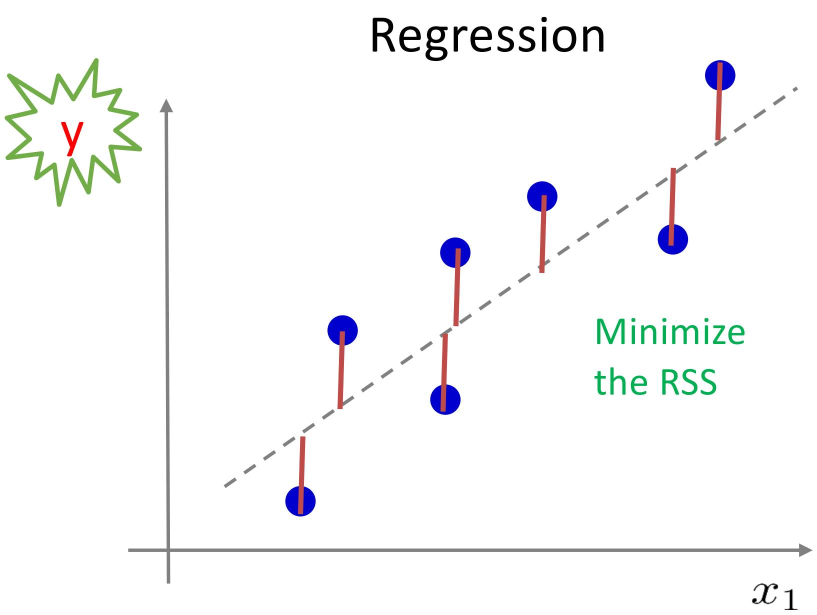


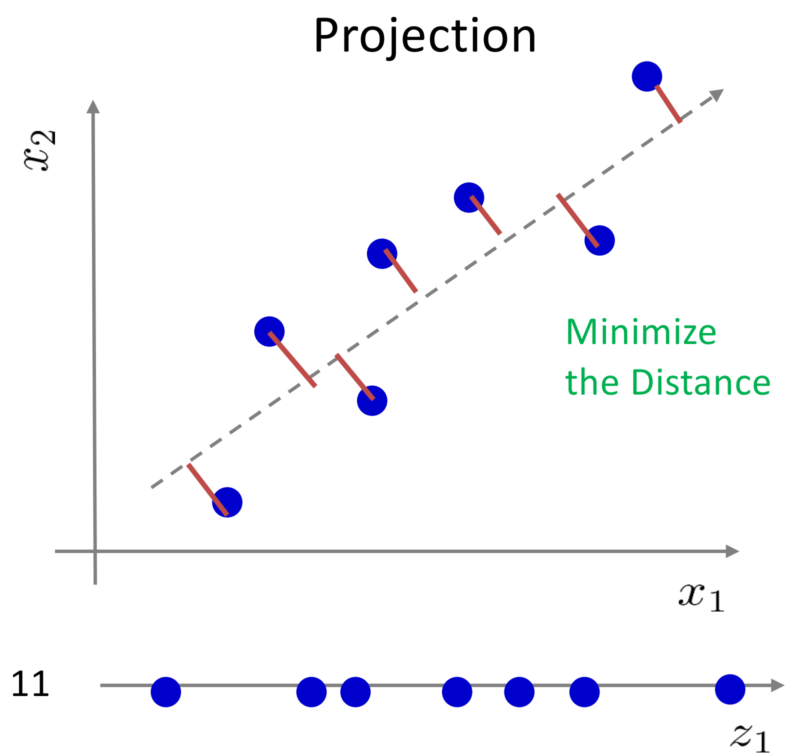
Dimensionality
Reduction





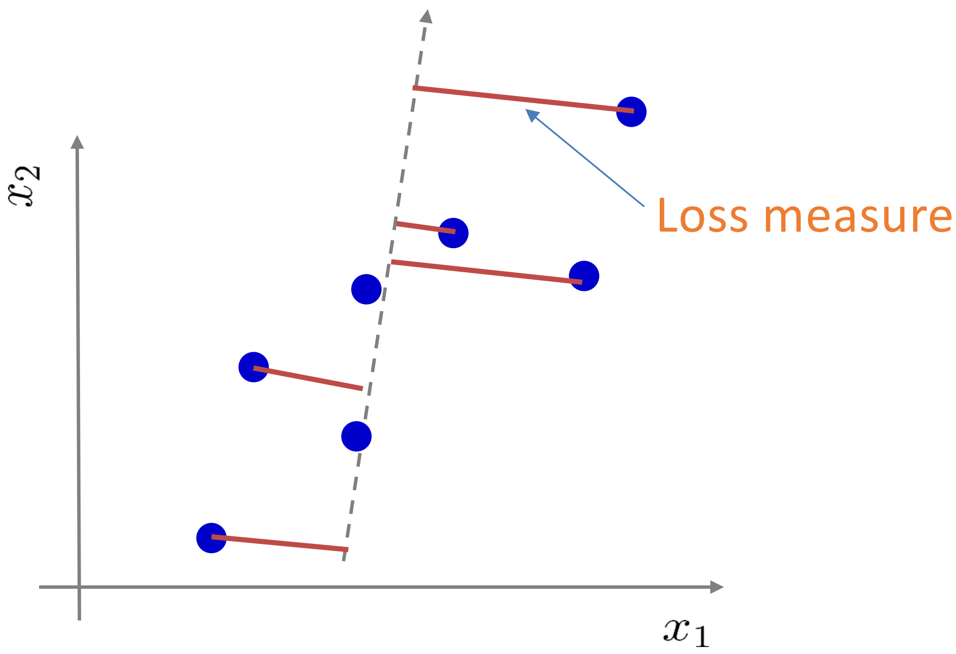
Principal Component Analysis



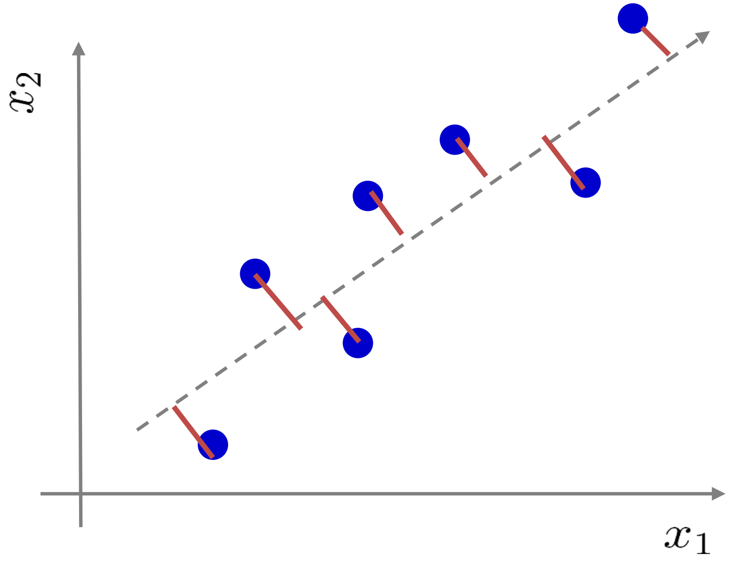


Don’t be confused !

Principal Component Analysis



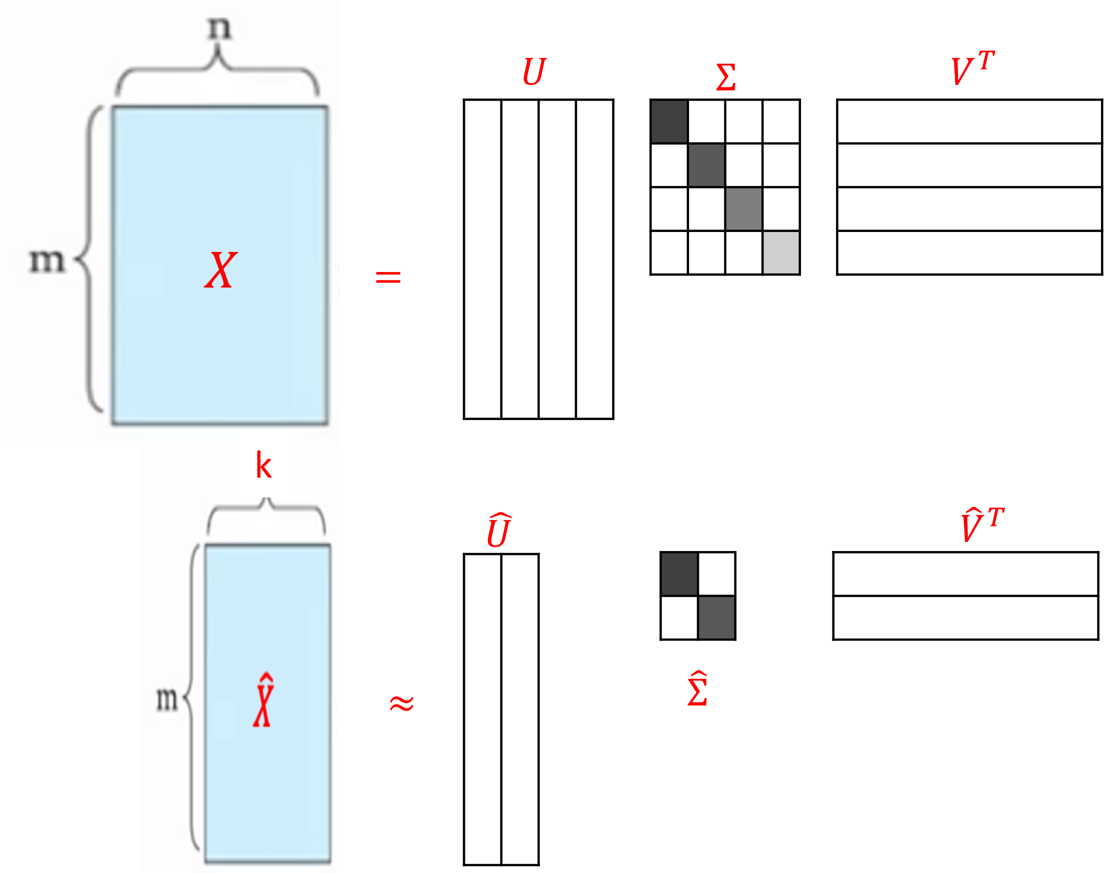
Maximum loss
Less variance



Minimum loss
More variance

Principal Component Analysis

- Singular Value Decomposition
  - $X = U\Sigma V^T$
  - Σ is diagonal, composed of ordered positive singular values
  - U and V form a basis (are orthonormal)
- k Principal Components
  - k first vectors of U
  - Corresponds to the first k singular values
- Explained Variance
  - Sum of the first k singular values
- Reconstruction
  - $\hat{X} = \hat{U}\hat{\Sigma}\hat{V}^T$



Principal Component Analysis

- Singular Value Decomposition (SVD) (very costly)
  - Parallelization: Incremental PCA (fast), Randomized PCA (faster)
- PCA assumes that the dataset is centered around the origin
- How many dimensions to preserve?
  - Reduce dimensions that add up to a sufficiently large portion of the variance (e.g., 99%)
- Kernel PCA (kPCA): use the kernel trick like SVM
- In practice, use kPCA to transform the feature space, then perform classification or regression.

Principal Component Analysis

- Hyper-Parameters Tuning
  - d: polynomial Kernel
  - γ: RBF kernel
  - K: Number of retained principal components
  - Etc.

Other Dimensionality Reduction Methods

- Multidimensional Scaling (MDS)
- Isomap
- t-Distributed Stochastic Neighbor Embedding (t-SNE)
- Manifold Learning
    - Locally Linear Embedding (LLE)
- Etc.