Feature Engineering

Papangkorn Inkeaw, Ph.D.

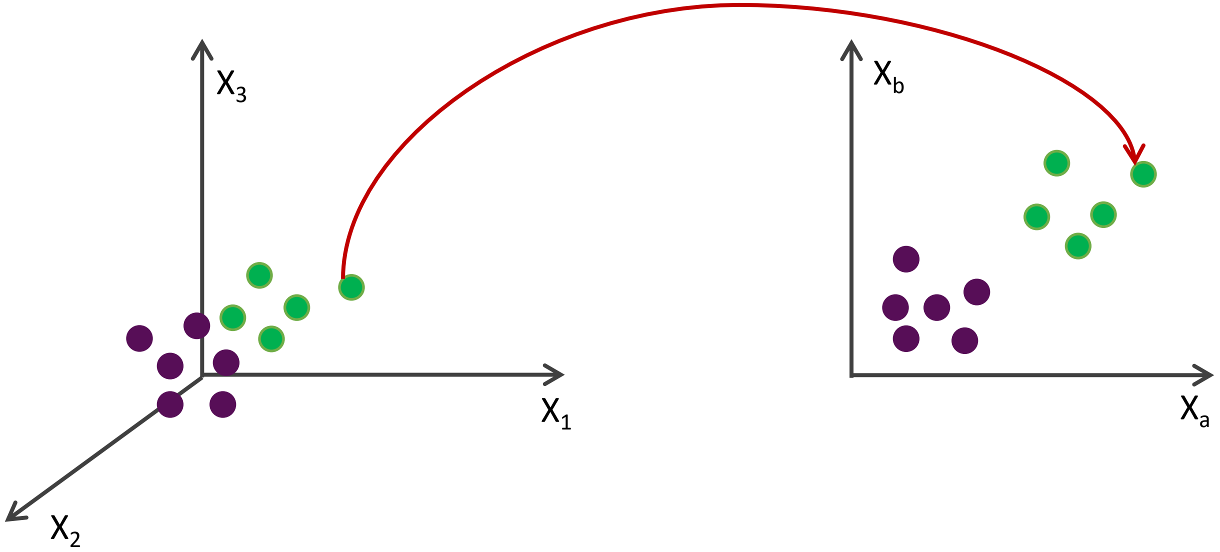
Dimensionality Reduction

Chapter 6 (Part II) - Feature Projection

Feature Projection

Transforms the data in the high-dimensional space to a space of fewer dimensions.

$$\begin{bmatrix} x_a \\ x_b \end{bmatrix} = f\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right)$$



Singular Value Decomposition (SVD)

- A factorization of a large matrix X of dimension $m \times n$ into the multiplication of three matrices:

$$X = U \Sigma V^T$$

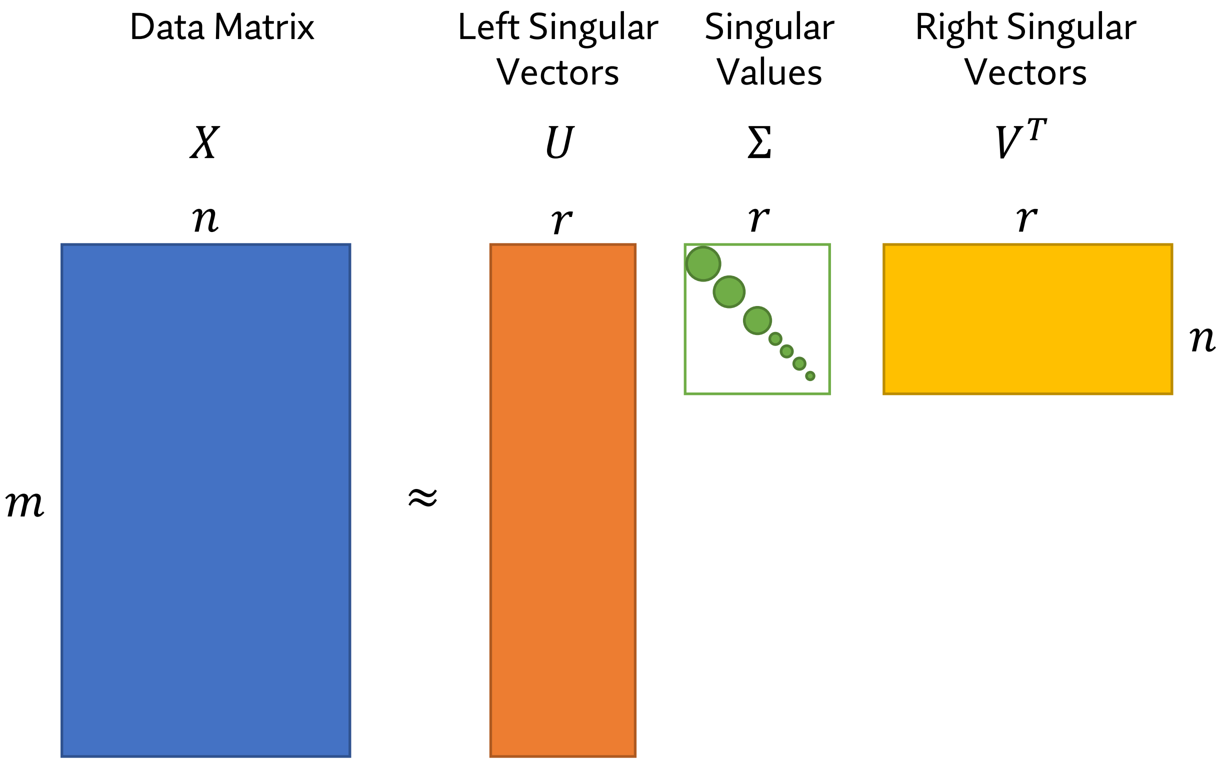
where U (Left Singular Vectors) is a $m \times r$ unitary matrix,

V (Right Singular Vectors) is a $r \times r$ matrix,

Σ (Singular Values) is a $r \times n$ matrix with nonnegative real numbers in the diagonal.

The diagonal values σ_i of Σ are the singular values of M (usually listed in order).

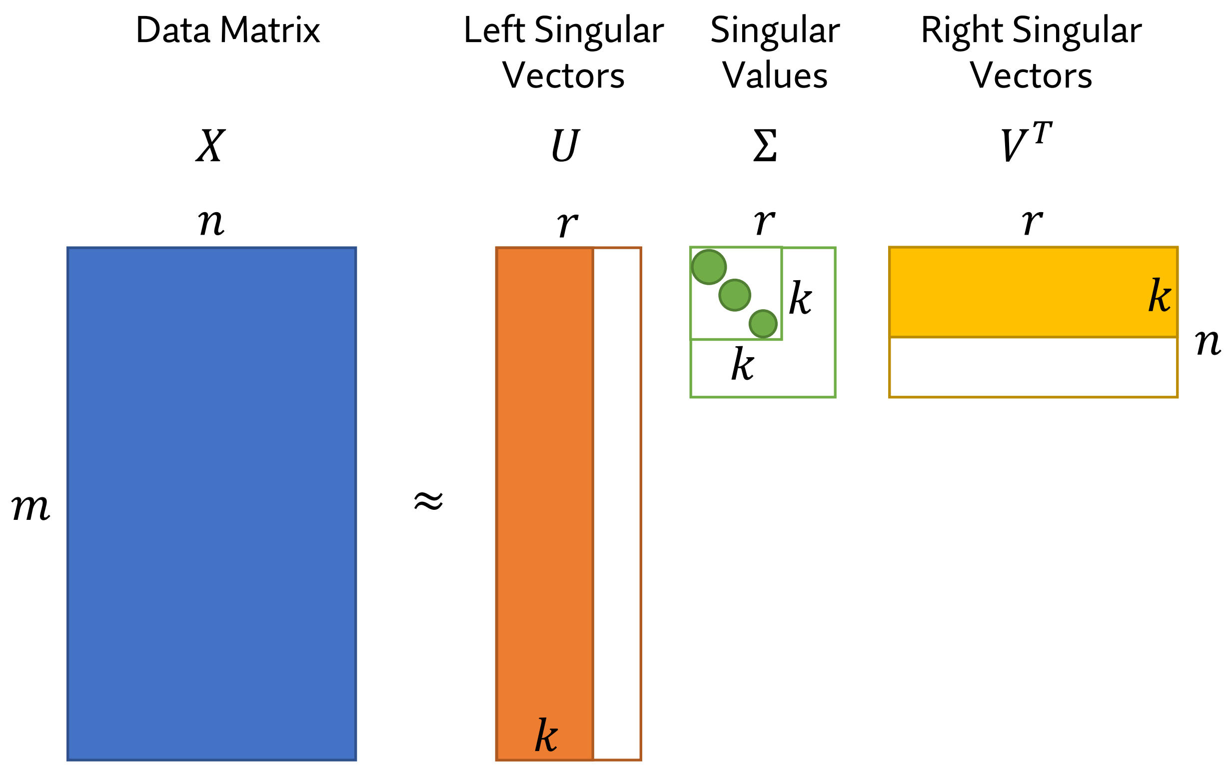
Singular Value Decomposition (SVD)



Singular Value Decomposition (SVD)

Dimensionality Reduction using SVD

- Only the first few singular values are large.
- Terms except the first few, *k*, can be ignored without losing much of the information.



Singular Value Decomposition (SVD)

Dimensionality Reduction using SVD

Given a data matrix X that each row represents a data point while each column is a variable.

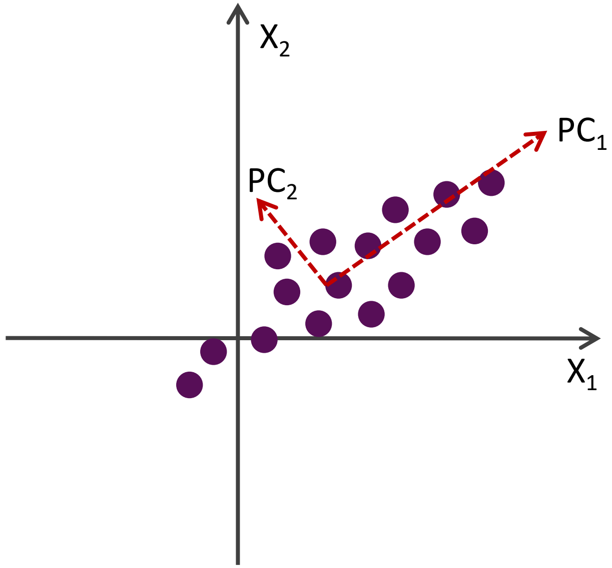
STEP 1: represent the matrix X by 3 smaller matrices U, Σ and V.

STEP 2: select the first *k* singular values and truncate the 3 matrices accordingly.

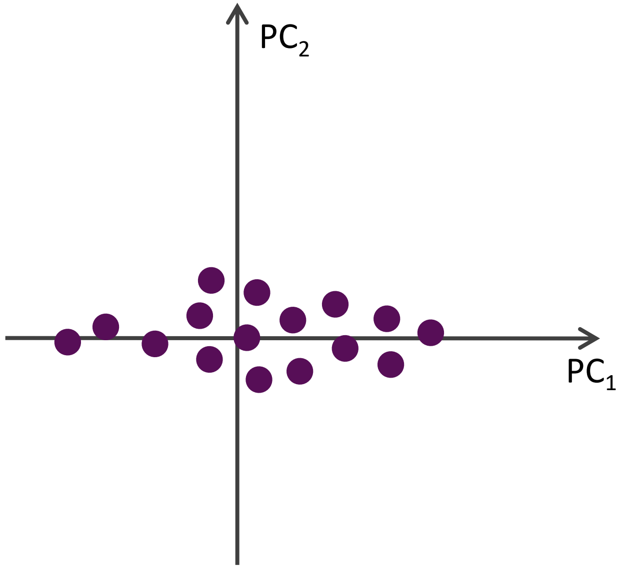
STEP 3: approximate X' by using the truncated matrixes.

Principal Component Analysis (PCA)

- Explain most of the variability observed in the original data.
  - The first PC of the data is a vector along which the observations vary the most, or in other words, a linear combination of the variables in the dataset that maximizes the variance.
- PCA aims to find the directions of maximum variance in high-dimensional data and projects it onto a new subspace with equal or fewer dimensions than the original one.



Original space



PC space

Principal Component Analysis (PCA)

- New feature vector, $z = [z_1, z_2, \dots, z_k]$, (in PC space) is linear combination of original features:

$$z = x^T W$$

  where $x = [x_1, x_2, \dots, x_d]$ is an original feature vector and W is a $d \times k$ dimensional transformation matrix.

- We want to find basic vectors which points in the direction of maximum variance.

$$\max_w w^T \Sigma w$$

$$\text{subject to } w^T w = 1$$

Principal Component Analysis (PCA)

Dimensionality Reduction using PCA

STEP 1: make the mean of data to be zero by subtracting by the mean vector: $X = X - \mu$.

STEP 2: calculate the covariance matrix $\Sigma = XX^T$

STEP 3: calculate the eigen values and eigen vectors of Σ. We obtain the eigen values $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$ and eigen vectors $v_1, v_2, \ldots, v_d$

STEP 4: select *k* eigen vectors with *k* largest eigen values .

STEP 5: construct a matrix $P = [v_1; v_2; \ldots; v_k]$

STEP 6: calculating the new features $x_i' = P^T(x_i - \mu)$

References & Study Resources

- Alice Zheng and Amanda Casari. (2018). *Feature Engineering for Machine Learning*. O'Reilly Media, Inc.
- Pablo Duboue. (2020). *The Art of Feature Engineering: Essentials for Machine Learning*. Cambridge University Press.
- Soledad Galli. (2020). *Python Feature Engineering Cookbook*. Packt Publishing.