Principal Component Analysis and Singular Value Decomposition

Pritha Banerjee

University of Calcutta

banerjee.pritha74@gmail.com

March 28, 2025

Overview

1. PCA
2. Singular Value Decomposition

Principal Component Analysis (PCA):

Steps of Computing PCA:

- Consider a d-dimensional data set $A_{n \times d}$ with n entries
- Compute mean for every column and subtract the mean from each element in that column $\bar{A} = [\bar{x}_{i1} \ \ \bar{x}_{i2} \cdots \bar{x}_{id}]$,ie. **mean centered data matrix**. Let the new matrix be $X_{n \times d} = A_{n \times d} - \bar{A}_{n \times d}$, ie.
- Compute Covariance matrix $V_{d \times d}$, which is a symmetric matrix (to capture covariance between different attributes)
- Compute Eigenvales and eigenvectors of V, Eigenvectors are linearly independent as V is symmetric.
- Sort the eigenvalues in decreasing order. Highest eogenvalue corresponds to the maximum variance.
- Create a matrix, called **projection matrix** $W_{d \times k}$, $k < d$, by choosing k eigenvectors corresponding to decreasing eigenvalues.
- Use $W_{d \times k}$ to transform X to new subspace of dimension k. $X' = X_{n \times d}.W_{d \times k}$

Principal Component Analysis: Merits/ Demerits

Can we get back the original A from X'? if all d eigenvectors are available, $\hat{A} = X'W^T + \bar{A}$

Merits:

- PCA works best for dimensionality reduction when attributes are highly correlated
- PCA helps in data compression
- Eliminates components with low variance (assumed to be noise), enhancing data clarity.
- Helps identifying outlier, by showing which point significant deviates in the reduces space.

Demerits:

- PCA needs proper scaling (mean centered data, standardization)
- If number of principal components chosen are small there may be information loss
- PCA works well if relationship between attributes are linear
- PCA may be slow for very large dataset
- If too many components are chosen or data set is too small PCA may not generalize well

Diagonalization of a Matrix

- Given a square matrix A, find a matrix P and a diagonal matrix D such that $D = P^{-1}AP$ or $A = PDP^{-1}$
- Multiplying on both sides by P, we get $PD = AP$,
- For P^{-1} to exist, columns of P must be linearly independent,
- Eigenvectors corresponding to **distinct** non-zero eigenvalues are linearly independent. Thus, we use eigenvalue and eigenvectors to obtain D.
- k^{th} column of AP is $AP_{.k}$ and k^{th} column of PD is $PD_{.k}$; $P_{.k}$ and $D_{.k}$ are k^{th} column of P and D
- Since D is diagonal, the only non-zero entry of $D_{.k}$ is D_{kk}. Therefore, $PD_{.k} = D_{kk}.P_{.k} \implies AP_{.k} = D_{kk}.P_{.k}$
- From above, $P_{.k}$ is an eigenvector of A corresponding to eigenvalue D_{kk}. Thus, non-zero diagonal entries in D are eigenvalues of A and P are the eigenvectors corresponding to them.
- Note that, depending on the arrangements of eigenvalues λ, P matrix may differ.

Singular Value Decomposition

- helps to produce approximate representation of a matrix of any desired number of dimension by eliminating the least important "content"
- **Definition:** Let $A_{m \times n}$ be a matrix with rank r, then we can find matrices, U, V, Σ such that $A = U_{m \times r}\Sigma_{r \times r}V^T r \times n$, where
  - U is column orthonormal matrix (dot product of two columns is zero)
  - Σ is a digonal matrix and values of the diagonals are called singular values of M (square root of the positive eigen values of AA^T or A^TA
  - **Note:** AA^T or A^TA are symmetric matrices and have same eigenvalues
  - V column orthonormal matrix, thus, rows of V^T are orthonormal
- We will not diagonalize A by $D = P^{-1}AP$; sice A may not be square matrix, eigenvectors may not be orthogonal
- **Singular vectors** of A solve the above problems.

Singular Value Decomposition: Intuition (1)

- There are two sets of singular vectors $u's$ and $v's$. $u's$ are in $\mathbb{R}^m$ and $v's$ are in $\mathbb{R}^n$
- singular vectors will be the columns of an $m \times m$ matrix U and $n \times n$ matrix V.
- SVD in terms of those basis vectors or SVD in terms of orthonormal matrices U and V
- $u's$ and $v's$ give bases for four fundamental subspaces:
  - $u_1, \cdots u_r$ is an orthonormal basis for the **column space**
  - $u_{r+1}, \cdots, u_m$ is an orthonormal basis for the **left nullspace** $N(A^T)$
  - $v_1, \cdots v_r$ is an orthonormal basis for the **row space**
  - $u_{r+1}, \cdots, u_m$ is an orthonormal basis for the **left nullspace** $N(A)$
  - These basis vectors diagonalize the matrix A
- A **is diagonalized:** $Av_1 = \sigma_1 u_1, Av_2 = \sigma_2 u_2 \cdots Av_r = \sigma_r u_r$, $\sigma_1, \cdots \sigma_r$ are positive and σ_i are lengths of Av_i, σ's go into diagonal matrix Σ

Singular Value Decomposition: Intuition (2)

- Since $u's$ are orthonormal, the matrix U with those r columns has $U^T U = I$, similarly $V^T V = I$
- $Av_i = \sigma_i u_i$ gives $AV_r = U_r \Sigma_r$, where A is $m \times n$, V_r is $n \times r$, U_r is $m \times r$ and Σ_r is $r \times r$

$$A \begin{bmatrix} v_1 & \cdots & v_r \end{bmatrix} = \begin{bmatrix} u_1 & \cdots & u_r \end{bmatrix} \begin{bmatrix} \sigma_1 & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \sigma_r \end{bmatrix}$$

- Above $u's$ and $v's$ define rowspace and columnspace of A
- We have $m - r$ more $u's$ and $n - r$ more $v's$ from the nullspace $N(A)$ and left nullspace $N(A^T)$
- Since they are in nullspace they are orthogonal. also they are orthogonal to first $u's$ and $v's$
- Considering all u and v in U and V, they become square matrix. $A_{m \times n} V_{n \times n} = U_{m \times m} \Sigma_{m \times m}$
- Σ has $m - r$ extra zero rows and $n - r$ new zero columns.

Singular Value Decomposition: Intuition (3)

- Now $AV = U\Sigma$ becomes $A = U\Sigma V^T$, multiplying both sides by V^T
- **SVD:** r matrices, each $u_i \sigma_i v_i^T$ is a column of A and has rank 1 for

  $$A = U\Sigma V^T = u_1\sigma_1 v_1^T + \cdots + u_r\sigma_r v_r^T$$

  Equation (1)
- To show that each σ_i^2 is an eigenvalue of A^TA and also AA^T. When **singular values** are put in descending order, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, Equation 1 gives the r, rank 1 pieces of A in order of importance.

When is $\Lambda = U\Sigma V^T$ (singular values) the same as $X\Lambda X^{-1}$ (eigenvalues)?

- A needs orthonormal eigenvectors to allow $X = U = V$
- A also needs eigenvalues $\lambda \geq 0$ if $\Lambda = \Sigma$
- Thus, A must be a positive semidefinite or definite symmetric matrix. Then only $A = X\Lambda X^{-1} = U\Sigma V^T$

Singular Value Decomposition: Proof

Proof of SVD: To show how $u's$ and $v's$ are constructed

- $v's$ are orthonormal eigenvectors of A^TA. Since $A^TA = (U\Sigma V^T)^T(U\Sigma V^T) = V\Sigma^T U^T U\Sigma V^T = V\Sigma^T\Sigma V^T$
- On the right, Eigenvector matrix V for symmetric positive (semi) definite matrix A^TA; Thus, $(\Sigma^T\Sigma)$ must be eigenvale matrix of of A^TA, each σ^2 is λ of A^TA
- In $Av_i = \sigma_i u_i$, u_i's are orthonormal because of v
- $u_i^T u_j = (\frac{Av_i}{\sigma_i})^T(\frac{Av_i}{\sigma_i}) = \frac{v_i^T A^T AV_j}{\sigma_i\sigma_j} = 0$, since v_i's are orthogonal.
- Note that u are the eigenvectors of AA^T