The Spectral Theory Basis of PCA

Ng Yen Kaow

Eigenvectors

- Only concerned with **square** matrices
- A **eigenvector** for a square matrix M is vector u where $\mathrm{Mu} = \lambda\mathrm{u}$
  - u is **invariant** under transformation M
  - The corresponding scaling factor is called a **eigenvalue**
- The **eigenspace** of a matrix M is the set of all the vectors u that fulfills $\mathrm{Mu} = \lambda\mathrm{u}$
- A **eigenbasis** of a $n \times n$ matrix M is a set of n **orthogonal** eigenvectors of M

Underlined = will not be defined here

Eigendecomposition

- A eigendecomposition of matrix M is
$$\mathrm{M} = \mathrm{Q}\Lambda\mathrm{Q}^{-1}$$
where Λ is **diagonal**, and Q contains (not necessarily orthogonal) **eigenvectors** of M
- Any **normal** M can be eigendecomposed
- Furthermore, for **real symmetric** M,
  - Eigenvectors that correspond to distinct eigenvalues are orthogonal (Spectral Theorem)
  - For an orthogonal matrix Q, $\mathrm{Q}^{-1} = \mathrm{Q}^{\mathrm{T}}$

Hence for real symmetric M, $\mathrm{M} = \mathrm{Q}\Lambda\mathrm{Q}^{\mathrm{T}}$

Rayleigh Quotient

- Consider an $n \times n$ real symmetric M
- $\mathrm{M} = \mathrm{Q}\Lambda\mathrm{Q}^\mathrm{T}$, where Λ is diagonal, and Q is the eigenbasis of M

- Denote the eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$.
  - Then, for all unit vector u

Maximum of Rayleigh Quotient

$$\max_{\|\mathrm{u}\|=1} \frac{\mathrm{u}^\mathrm{T}\mathrm{M}\mathrm{u}}{\mathrm{u}^\mathrm{T}\mathrm{u}} = \lambda_1$$

Similarly, λ_n is the minimum of the Rayleigh Quotient

  - And for all orthogonal matrix P and $k \leq n$

Minimax Principle

$$\max_{\mathrm{P}\in\mathbb{R}^{k\times n}, \mathrm{P}^\mathrm{T}\mathrm{P}=\mathrm{I}} \mathrm{tr}\left(\mathrm{P}^\mathrm{T}\mathrm{M}\mathrm{P}\right) = \lambda_1 + \cdots + \lambda_k$$

Similarly, $\min_{\mathrm{P}\in\mathbb{R}^{k\times n}, \mathrm{P}^\mathrm{T}\mathrm{P}=\mathrm{I}} \mathrm{tr}\left(\mathrm{P}^\mathrm{T}\mathrm{M}\mathrm{P}\right) = \lambda_{n-k+1} + \cdots + \lambda_n$

Eigendecomposition applications

- Matrix inverse
- Matrix approximation
- Minimization or maximization through the Rayleigh Quotient
  - PCA
    - Covariance matrix
    - Find maximum
  - Spectral clustering
    - Graph Laplacian
    - Find minimum

Singular Value Decomposition

- Any matrix can be singular value decomposed
- $\mathrm{M} = \mathrm{U\Sigma V}^*$
  - M is $m \times n$ matrix
  - U is an $m \times m$ **unitary** matrix

    Orthogonal
  - Σ is an $m \times n$ diagonal matrix
  - V is an $n \times n$ **unitary** matrix

- For a real M, $\mathrm{V}^* = \mathrm{V}^{\mathrm{T}}$ (and $\mathrm{U} = \mathrm{U}^{\mathrm{T}}$) hence

$$\mathrm{M} = \mathrm{U\Sigma V}^{\mathrm{T}}$$

SVD applications

- Solving linear equations
- Linear regression
- Pseudoinverse
- Kabsch algorithm
- Matrix approximation
- As a eigendecomposition (see next slide)

SVD and eigendecomposition

- SVD is a eigendecomposition but not of M
  - Given an SVD of $\mathrm{M} = \mathrm{U\Sigma V^*}$
  - Then, clearly
    - $\mathrm{M^*M} = \mathrm{V\Sigma^*U^*U\Sigma V^*} = \mathrm{V(\Sigma^*\Sigma)V^*}$
    - $\mathrm{MM^*} = \mathrm{U\Sigma V^*V\Sigma^*U^*} = \mathrm{U(\Sigma^*\Sigma)U^*}$
  - Hence V is the eigenbasis of $\mathrm{M^*M}$ and U is the eigenbasis of $\mathrm{MM^*}$ respectively
  - That is, U and V are **eigenbases of the squared matrices of** M
    - However the eigenbasis of $\mathrm{M^*M}$ and $\mathrm{MM^*}$ are in general not the eigenbasis of M

Principal Component Analysis

- Let X be an $m \times n$ matrix where each row represents a vector in an n-D space
  - That is, X represents the input data
- What do we ideally expect to be the "principal components" of X
  1. The components **form a basis**
  2. The components are **orthogonal**
  3. The **first component accounts for the most variation**, the second component accounts for the most variation after removing the first component, and so on

Principal Component Analysis

- Let X be an $m \times n$ matrix where each row represents a vector in an n-D space
  - X represents m datapoints in n-D
- Assume that the rows in X are generated by a random vector $\mathbf{X} \in \mathbb{R}^n$
  - Note the difference between $\mathbf{X}$ and X
  - The theory of PCA is based on $\mathbf{X}$ (and its $n \times n$ covariance matrix M)
- For the first component, we want to find unit vector $\mathrm{u} \in \mathbb{R}^n$ such that $\mathrm{var}\left(\mathrm{u}^{\mathrm{T}}\mathbf{X}\right)$ is maximized

Principal Component Analysis

- The eigenvector u of the covariance matrix of $\mathbf{X}$ with the largest eigenvalue maximizes $\mathrm{var}(\mathrm{u}^{\mathrm{T}}\mathbf{X})$

Let $\mathbf{X} \in \mathbb{R}^n$ be a random vector with

- mean $\mu \in \mathbb{R}^n$ and
- covariance matrix $\mathrm{M} = \mathbb{E}[(\mathbf{X}-\mu)(\mathbf{X}-\mu)^{\mathrm{T}}]$

For any $\mathrm{u} \in \mathbb{R}^n$, the projection of $\mathrm{u}^{\mathrm{T}}\mathbf{X}$ has

- $\mathbb{E}[\mathrm{u}^{\mathrm{T}}\mathbf{X}] = \mathrm{u}^{\mathrm{T}}\mu$ and
- $\mathrm{var}(\mathrm{u}^{\mathrm{T}}\mathbf{X}) = \mathbb{E}\left[\left(\mathrm{u}^{\mathrm{T}}\mathbf{X} - \mathrm{u}^{\mathrm{T}}\mu\right)^2\right]$
  $= \mathbb{E}[\mathrm{u}^{\mathrm{T}}(\mathbf{X}-\mu)(\mathbf{X}-\mu)^{\mathrm{T}}\mathrm{u}] = \mathrm{u}^{\mathrm{T}}\mathrm{M}\mathrm{u}$

From earlier slide, $\mathrm{u}^{\mathrm{T}}\mathrm{M}\mathrm{u}$ is maximized when u is the eigenvector of M with the largest eigenvalue

Principal Component Analysis

- Extend to k principal components, we want
  - k-D subspace of $\mathbf{X}$ that is defined by orthogonal basis $\mathrm{p}_1, \ldots, \mathrm{p}_k \in \mathbb{R}^d$ and displacement $\mathrm{p}_0 \in \mathbb{R}^d$
  - Distance from $\mathbf{X}$ to this subspace is minimized

- Projection of $\mathbf{X}$ onto subspace is $\mathrm{P}^{\mathrm{T}}\mathbf{X} + \mathrm{p}_{\mathbf{0}}$, where P is matrix whose rows are $\mathrm{p}_1, \ldots, \mathrm{p}_k$
- Squared distance to subspace is $\mathbb{E}\left\|\mathbf{X} - \left(\mathrm{P}^{\mathrm{T}}\mathbf{X} + \mathrm{p}_{\mathbf{0}}\right)\right\|^2$
- By calculus, $\mathrm{p}_{\mathbf{0}} = \mathbb{E}\left\|\mathbf{X} - \mathrm{P}^{\mathrm{T}}\mathbf{X}\right\| = \left(1 - \mathrm{P}^{\mathrm{T}}\right)\mu$, hence
  $$\mathbb{E}\left\|\mathbf{X} - \left(\mathrm{P}^{\mathrm{T}}\mathbf{X} + \mathrm{p}_{\mathbf{0}}\right)\right\|^2 = \mathbb{E}\|\mathbf{X} - \mu\|^2 - \mathbb{E}\left\|\mathrm{P}^{\mathrm{T}}(\mathbf{X} - \mu)\right\|^2$$
- To maximize that, need to maximize $\mathbb{E}\left\|\mathrm{P}^{\mathrm{T}}(\mathbf{X} - \mu)\right\|^2 = \mathrm{var}\left(\mathrm{P}^{\mathrm{T}}\mathbf{X}\right)$
- Finally, same as in previous slide, $\mathrm{p}_1, \ldots, \mathrm{p}_k$ are eigenvectors of M

Principal Component Analysis

- If X is normalized such that each column has zero mean, an unbiased estimator of M can be obtained as

$$\mathrm{M} = \frac{1}{n-1}\mathrm{X}^{\mathrm{T}}\mathrm{X}$$

- Since SVD of X eigendecomposes $\mathrm{X}^{\mathrm{T}}\mathrm{X}$
  - This allows us to solve PCA through either
    1. Eigendecompose M, or
    2. Solve SVD for X