Eigendecomposition, SVD, and Power Iterations

Objectives

- Define eigenvectors and eigenvalues
- Relate the eigendecomposition to SVD
- Power iterations for computing eigenvector with largest eigenvalue

Eigendecomposition applies to square matrices

k×1

Eigenvector $\underline{e}_i$, eigenvalue λ_i, $\underline{B}$ $(k \times k)$

$$\underline{B}\,\underline{e}_i = \lambda_i\,\underline{e}_i$$ matrix mult $\Leftrightarrow$ scalar mult

$$\underline{e}_i \longrightarrow \boxed{\underline{B}} \longrightarrow \lambda_i\,\underline{e}_i \qquad i = 1, 2, \cdots, k$$

- K eigenvalues, possibly complex valued
- Distinct $\lambda_i \Rightarrow$ linearly independent $\underline{e}_i$
- Symmetric $\underline{B} \Rightarrow K$ orthonormal $\underline{e}_i$ $\quad \underline{E}\,\underline{E}^T = \underline{E}^T\underline{E} = \underline{I}$

orthonormal

$$\underline{B}\,\underline{e}_i = \lambda_i\,\underline{e}_i \Rightarrow \underline{B}\,[\underline{e}_1\ \underline{e}_2 \cdots \underline{e}_k] = [\underline{e}_1\ \underline{e}_2 \cdots \underline{e}_k]\begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \ \lambda_k \end{bmatrix}$$

diagonal entries are λ_i

$$\underline{B}\,\underline{E} = \underline{E}\,\underline{\Lambda} \Rightarrow \underline{B} = \underline{E}\,\underline{\Lambda}\,\underline{E}^T = \sum_{i=1}^{k} \lambda_i\,\underline{e}_i\,\underline{e}_i^T$$

Symmetric PSD matrices and SVD

PSD → positive definite

$$\underline{A} = [\underline{a}_1\ \underline{a}_2 \cdots \underline{a}_M]_{N\times M} = \begin{bmatrix} \underline{x}_1^T \\ \underline{x}_2^T \\ \vdots \\ \underline{x}_N^T \end{bmatrix}_{N\times n} = \underline{U}\,\underline{\Sigma}\,\underline{V}^T \quad (N\times M,\ N>M)$$

→ A is rows form

full SVD

1) $\underline{B}^{n\times n} = \underline{A}\,\underline{A}^T = \sum_{i=1}^{M} \underline{a}_i \underline{a}_i^T$ → symmetric because $(AA^T)^T = (AA^T)$

$$\underline{B} = \underline{U}\,\underline{\Sigma}\,\underline{V}^T\,\underline{V}\,\underline{\Sigma}^T\underline{U}^T = \underline{U}\,\underline{\Sigma}\,\underline{\Sigma}^T\underline{U}^T = \underline{U}\begin{bmatrix} \sigma_1^2 & & & \\ & \ddots & & \underline{0} \\ & & \sigma_M^2 & \\ \underline{0} & & & 0 \ddots 0 \end{bmatrix}\underline{U}^T$$

left SV of $\underline{A} \Leftrightarrow$ eigenvectors $\underline{B}$

$$\lambda_i = \begin{cases} \sigma_i^2 & i = 1, 2, \cdots, M \\ 0 & i = M+1, \cdots, N \end{cases}$$

2) $\underline{B}^{M\times M} = \underline{A}^T\underline{A} = \sum_{i=1}^{N} \underline{x}_i \underline{x}_i^T$

$$= \underline{V}\,\underline{\Sigma}^T\underline{U}^T\underline{U}\,\underline{\Sigma}\,\underline{V}^T = \underline{V}\,\underline{\Sigma}^T\underline{\Sigma}\,\underline{V}^T = \underline{V}\begin{bmatrix} \sigma_1^2 & & \underline{0} \\ & \ddots & \\ \underline{0} & & \sigma_M^2 \end{bmatrix}\underline{V}^T$$

right SV of $\underline{A} \Leftrightarrow$ eigenvectors $\underline{B}$, $\lambda_i = \sigma_i^2$, $i = 1, 2, \cdots M$

(eigenvalue: λ_i; square of singular values: σ_i^2)

Power iteration for computing 1^{st} principal component

$\underline{A} : N \times M$, $N \gg M$ want $\underline{v}_1$ 1^{st} principal component

right SV of $\underline{A}$, eigenvector of $\underline{B} = \underline{A}^T \underline{A}$ $(M \times M)$

Power Iteration

Algorithm for finding the eigenvector of B (the max one)

pick $\underline{c}_{0\,M\times 1}$ (random)

for $k = 1, 2, \ldots$ to converge

$$\underline{c}_k = \underline{B}\,\underline{c}_{k-1} / \|\underline{B}\,\underline{c}_{k-1}\|_2$$

Normalize

end

This is what we want

$$\underline{v}_1 = \underline{c}_{end}$$

(☆ Idea: multiplying c by B many times and normalizing it)

$$\underline{B}\,\underline{c}_{k-1} = \underbrace{\underline{B} \cdot \underline{B} \cdots \underline{B}}_{k \text{ times}}\,\underline{c}_0 = \underline{B}^k \underline{c}_0$$

Times c_0 by k times of B

$$\underline{B}^k = \underbrace{\underline{V}\,\underline{\Lambda}\,\cancel{\underline{V}^T \underline{V}}\,\underline{\Lambda}\,\cancel{\underline{V}^T \cdots \underline{V}}\,\underline{\Lambda}\,\underline{V}^T}_{k \text{ times}}$$

$$= \underline{V}\,\underline{\Lambda}^k \underline{V}^T$$

Let $\underline{c}_0 = \underline{V}\,\underline{g} = \underline{V} \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_M \end{bmatrix}$

Just coefficients

c_0 can be represented by B's feature vectors V and g as linear combinations

$$\underline{B}^k \underline{c}_0 = \underline{V}\,\underline{\Lambda}^k \cancel{\underline{V}^T \underline{V}}\,\underline{g}$$

$$= \underline{V}\,\underline{\Lambda}^k \underline{g}$$

Power iteration...

$$\underline{c}_k = \underline{B}\underline{c}_{k-1} / \|\underline{B}\underline{c}_{k-1}\|_2 = \overset{\underline{B}^k \underline{c}_0 \\ \|}{\underline{V}\underline{\Lambda}^k \underline{g}} / \|\underline{V}\underline{\Lambda}^k \underline{g}\|_2$$

λ_1 is the largest

$$\underline{V}\underline{\Lambda}^k \underline{g} = [\underline{v}_1 \; \underline{v}_2 \; \cdots \; \underline{v}_m] \begin{bmatrix} \lambda_1^k & & \underline{0} \\ & \lambda_2^k & \\ \underline{0} & & \ddots \; \lambda_m^k \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_m \end{bmatrix}$$

but $\frac{\lambda_i}{\lambda_1} < 1$

Factor out the λ_1 and g_1 from matrix

$$= \lambda_1^k g_1 \underline{V} \begin{bmatrix} 1 & & \underline{0} \\ & \left(\frac{\lambda_2}{\lambda_1}\right)^k & \\ & & \ddots \\ \underline{0} & & \left(\frac{\lambda_m}{\lambda_1}\right)^k \end{bmatrix} \begin{bmatrix} 1 \\ g_2/g_1 \\ \vdots \\ g_m/g_1 \end{bmatrix}$$

so $\left(\frac{\lambda_i}{\lambda_1}\right)^k \to 0$ as k goes large

$$\underline{V}\underline{\Lambda}^k \underline{g} \to \lambda_1^k g_1 \underline{V} \begin{bmatrix} 1 & & 0 \\ & 0 \ddots & \\ 0 & & 0 \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \end{bmatrix} = \lambda_1^k g_1 \underline{v}_1$$

$$\underline{c}_k \to \frac{\lambda_1^k g_1 \underline{v}_1}{\|\lambda_1^k g_1 \underline{v}_1\|_2} = \frac{\underline{v}_1}{\|\underline{v}_1\|_2} = \underline{v}_1$$

Guaranteed to converge