A Crash Course on Linear Algebra II

Ye Wang
University of North Carolina at Chapel Hill

Mathematics and Statistics For Political Research
POLI783

Eigenvalues

- For a $N \times N$ square matrix $\mathbf{A}$, $\mathbf{Ax}$ does two things to $\mathbf{x}$: rotating and scaling.
- How do we distinguish these two operations?
- We try to find real numbers λ and vectors $\mathbf{v}$ such that

$$\mathbf{Av} = \lambda\mathbf{v} = \lambda\mathbf{Iv}.$$

- Along the direction of $\mathbf{v}$, $\mathbf{A}$ scales an vector by a factor of λ.
- To ensure that a non-zero $\mathbf{v}$ exists, we need

$$det(\mathbf{A} - \lambda\mathbf{I}) = 0,$$

  i.e., the dimensionality of $(\mathbf{A} - \lambda\mathbf{I})$'s kernel space is not zero.
- As $\mathbf{A}$ is a $N \times N$ matrix, $det(\mathbf{A} - \lambda\mathbf{I})$ is a Nth-degree polynomial and has N roots.
- This is due to the fundamental theorem of algebra.
- These roots are known as $\mathbf{A}$'s eigenvalues.

Eigenvectors

- We can show that $det(\mathbf{A}) = \prod_{i=1}^{N} \lambda_i$.
- Therefore, if $\mathbf{A}$ is invertible/full rank or has a non-zero determinant, none of its eigenvalues can be zero.
- One problem: not all the eigenvalues are real numbers.
- When $\mathbf{A}$ is symmetric, all its eigenvalues are real.
- Moreover, $rank(\mathbf{A})$ equals the number of non-zero eigenvalues.
- Next, we solve $(\mathbf{A} - \lambda_i \mathbf{I})\mathbf{v} = \mathbf{0}$ for each eigenvalue.
- The solution is known as eigenvectors for λ_i.
- Eigenvectors for any eigenvalue is non-unique, but eigenvectors for different eigenvalues are orthogonal.
- Choosing one eigenvector for each eigenvalue, we obtain a $N \times N$ orthogonal matrix $\mathbf{Q}$.

Eigenvalue decomposition

- Consider $\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$, then

$$det(\mathbf{A} - \lambda \mathbf{I}) = det \begin{pmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{pmatrix} = (2-\lambda)^2 - 1.$$

- Letting it be zero, we obtain $\lambda_1 = 1$ and $\lambda_1 = 3$.
- We derive eigenvectors for λ_1 by solving

$$(\mathbf{A} - \mathbf{I})\mathbf{v}_1 = \begin{pmatrix} v_{11} + v_{12} \\ v_{11} + v_{12} \end{pmatrix} = \mathbf{0}.$$

- A natural choice is $\mathbf{v}_1 = (1, -1)'$.
- Similarly, we can derive an eigenvector for λ_2: $\mathbf{v}_2 = (1, 1)'$.
- Defining $\mathbf{Q} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$, we can verify that $\mathbf{Q}$ is an orthogonal matrix.

Eigenvalue decomposition

- Let's define the diagonal matrix of eigenvalues:
  $\Lambda = diag\{\lambda_1, \lambda_2, \dots, \lambda_N\}$, then

  $$\mathbf{A} = \mathbf{Q}\Lambda\mathbf{Q}',$$

  which is known as the eigenvalue decomposition of $\mathbf{A}$.
- $det(\mathbf{A}) = det(\mathbf{Q})det(\Lambda)det(\mathbf{Q}') = det(\Lambda) = \prod_{i=1}^{N} \lambda_i$.
- $\mathbf{Q}$ and its transpose capture the rotation created by the matrix, and Λ captures scaling from it.
- Note that $\frac{1}{N}\mathbf{A}'\mathbf{A}$ and $\frac{1}{M}\mathbf{A}\mathbf{A}'$ are always symmetric thus admit an eigenvalue decomposition.
- This is known as the principal component analysis (PCA) and widely used in social science.

Principal component analysis

- A dataset with N legislators and their votes on M bills can be seen as a $N \times M$ matrix $\mathbf{A}$.
- $\frac{1}{N}\mathbf{A}'\mathbf{A}$ is an $M \times M$ symmetric matrix.
- Its eigenvectors tell us along which dimensions the bills differ.
- The corresponding eigenvalues measure the importance (variance explained) of each dimension.
- The eigenvector associated with the largest eigenvalue often captures the primary ideological dimension of the bills (e.g., liberal vs. conservative).
- The eigenvector associated with the second-largest eigenvalue captures the next main dimension (e.g., establishment vs. populism).
- Similarly, we can perform eigenvalue decomposition on $\frac{1}{M}\mathbf{A}\mathbf{A}'$.
- The eigenvectors indicate dimensions along which the legislators differ.
- The eigenvector associated with the largest eigenvalue provides each legislator's position on the main ideological spectrum.

Singular value decomposition

- For any $M \times N$ matrix $\mathbf{A}$, we can conduct the following eigenvalue decompositions:

$$\mathbf{A}'\mathbf{A} = \mathbf{V}\Lambda_N\mathbf{V}', \mathbf{A}\mathbf{A}' = \mathbf{U}\Lambda_M\mathbf{U}',$$

  where $\mathbf{V}$ and $\mathbf{U}$ are orthogonal matrices.
- We can show that $\mathbf{A}'\mathbf{A}$ and $\mathbf{A}\mathbf{A}'$ have the same rank R and the same non-zero eigenvalues.
- Therefore, $\Lambda_N = diag\{\lambda_1, \lambda_2, \ldots, \lambda_R, 0, \ldots, 0\}$ and $\Lambda_M = diag\{\lambda_1, \lambda_2, \ldots, \lambda_R, 0, \ldots, 0\}$, where $\lambda_r \geq 0$ for any r.
- Then, we obtain the singular value decomposion (SVD) of $\mathbf{A}$:

$$\mathbf{A} = \mathbf{U}_R\Sigma_R\mathbf{V}'_R,$$

  where $\mathbf{U}_R$ is the first R columns of $\mathbf{U}$, $\mathbf{V}_R$ is the first R columns of $\mathbf{V}$, and $\Sigma_R = diag\{\sqrt{\lambda}_1, \sqrt{\lambda}_2, \ldots, \sqrt{\lambda}_R\}$.
- SVD allows us to learn the ideal point of each legislator and each bill directly.

Multivariate calculus: Jacobian

- Consider a multivariate function with N arguments:

$$y_1 = f_1(x_1, x_2, \ldots, x_N).$$

- Its gradient is a vector, $\nabla f_1(\mathbf{x})$.
- Now, with M different functions, $f_m(x_1, x_2, \ldots, x_N)$, with $1 \leq m \leq M$, we define their Jacobian matrix at $\mathbf{x}$ as

$$\mathbf{J_f}(\mathbf{x}) = \begin{pmatrix} \nabla' f_1(\mathbf{x}) \\ \nabla' f_2(\mathbf{x}) \\ \vdots \\ \nabla' f_M(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_N} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_M}{\partial x_1} & \frac{\partial f_M}{\partial x_2} & \cdots & \frac{\partial f_M}{\partial x_N} \end{pmatrix}.$$

- With the Jacobian, we can write the first-order approximation for the M-dimensional function $\mathbf{f}(\cdot)$ as

$$\mathbf{f}(\mathbf{x}) \approx \mathbf{f}(\mathbf{x}_0) + \mathbf{J_f}(\mathbf{x})(\mathbf{x} - \mathbf{x}_0).$$

Multivariate calculus: Hessian

- We define the Hessian matrix of a function $f(\mathbf{x})$ at $\mathbf{x}$ as

$$\mathbf{H_f}(\mathbf{x}) = \mathbf{J}_{\nabla' f}(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 x_N} \\ \frac{\partial^2 f}{\partial x_2 x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 x_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_N x_1} & \frac{\partial^2 f}{\partial x_N x_2} & \cdots & \frac{\partial^2 f}{\partial x_N^2} \end{pmatrix}.$$

- The Hessian matrix is symmetric and generalizes the second-order derivative for univariate functions.
- We can express the second-order Taylor expansion of a multivariate function as

$$\begin{aligned} f(\mathbf{x}) =& f(\mathbf{x}_0) + \nabla' f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + \frac{(\mathbf{x} - \mathbf{x}_0)' \mathbf{H}_f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)}{2} \\ &+ o\left(||\mathbf{x} - \mathbf{x}_0)||^2\right). \end{aligned}$$

Quadratic form

- Remember that we need to examine both the first- and second-order derivative for extreme values of univariate functions.
- For a multivariate function, we should consider its Hessian.
- For a square matrix $\mathbf{A}$ and a vector $\mathbf{v}$, $\mathbf{v}'\mathbf{A}\mathbf{v}$ is known as a quadratic form.
- It generalizes the quadratic term of scalars.
- Consider $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}$ and $\mathbf{v} = (v_1, v_2)'$, then

$$\mathbf{v}'\mathbf{A}\mathbf{v} = a_{11}v_1^2 + 2a_{12}v_1v_2 + a_{22}v_2^2.$$

- When $\mathbf{A}$ is symmetric, it admits an eigenvalue decomposition, thus

$$\mathbf{v}'\mathbf{A}\mathbf{v} = \mathbf{v}'\mathbf{Q}\Lambda\mathbf{Q}'\mathbf{v} = \tilde{\mathbf{v}}'\Lambda\tilde{\mathbf{v}} = \sum_{i=1}^{N} \lambda_i \tilde{v}_i^2.$$

Semi positive-definite

- If $\lambda_i > 0$ for any i, then $\mathbf{v}'\mathbf{A}\mathbf{v} > 0$ for any $\mathbf{v}$.
- We say the matrix $\mathbf{A}$ is positive-definite.
- A positive-definite matrix is invertible.
- If If $\lambda_i \geq 0$ for any i, then $\mathbf{v}'\mathbf{A}\mathbf{v} \geq 0$ for any $\mathbf{v}$, and we say $\mathbf{A}$ is semi positive-definite.
- Matrices with the form of $\mathbf{A}'\mathbf{A}$ are always semi positive-definite.
- For any vector $\mathbf{v}$, let $\mathbf{u} = \mathbf{A}\mathbf{v}$

$$\mathbf{v}'\mathbf{A}'\mathbf{A}\mathbf{v} = (\mathbf{A}\mathbf{v})'\mathbf{A}\mathbf{v} = \mathbf{u}'\mathbf{u} = ||\mathbf{u}||^2 \geq 0.$$

- That's why all eigenvalues of $\mathbf{A}'\mathbf{A}$ ($\mathbf{A}\mathbf{A}'$) are non-negative.
- We can similarly define negative-definite and semi negative-definite matrices.

Optimization

- To determine the extrema of $f(\mathbf{x})$, we first find all the stationary points $\mathbf{x}^*$, where $\nabla f(\mathbf{x}^*) = \mathbf{0}$.
- Next, we examine the Hessian's value at $\mathbf{x}^*$.
- We have a local minimum (maximum) if $\mathbf{H}_f(\mathbf{x}^*)$ is positive (negative) definite.
- E.g., $f(x, y) = x^2 + y^2 - 4x - 6y$.
- $\nabla f(x, y) = (2x - 4, 2y - 6)'$, thus $x^* = 2$ and $y^* = 3$.
- $\mathbf{H}_f(x, y) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ is positive definite.
- Therefore, $(2, 3)'$ is a local minimization point for $f(x, y)$.

Matrix calculus

- Matrix calculus extends standard differentiation to mappings we have seen in linear algebra.
- For the linear function $\mathbf{y} = f(\mathbf{x}) = \mathbf{a}'\mathbf{x}$, we can verify that $\nabla f = \mathbf{a}$.
- For the linear mapping $\mathbf{y} = \mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x}$, we have

$$\mathbf{J}_{\mathbf{f}}(\mathbf{x}) = \mathbf{A}.$$

- For the quadratic form $f(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{i=1}^{N} \sum_{j=1}^{N} x_i a_{ij} x_j$, we have

$$\nabla f = (\mathbf{A} + \mathbf{A}')\mathbf{x}.$$

- If $\mathbf{A}$ is symmetric, then $\nabla f = 2\mathbf{A}\mathbf{x}$.
- We can also see the quadratic form as a function of $\mathbf{A}$, then

$$\frac{\partial f}{\partial \mathbf{A}} = \mathbf{x}\mathbf{x}',$$

a $N \times N$ matrix.

Multivariate random variables

- All P-dimensional vectors of r.v.s comprise a Hilbert space.
- The inner product of $\mathbf{X}$ and $\mathbf{Y}$ is defined as

$$\text{Cov}[\mathbf{X}, \mathbf{Y}] = \mathbb{E}\left[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{Y} - \mathbb{E}[\mathbf{Y}])'\right],$$

  a $P \times P$ matrix.
- $\text{Var}[\mathbf{X}] = \text{Cov}[\mathbf{X}, \mathbf{X}]$ is a positive semi-definite matrix and admits an eigenvalue decomposition $\mathbf{Q}\Lambda\mathbf{Q}$.
- Define $\sqrt{\mathbf{V}} = diag\{\sqrt{\text{Var}[X_1]}, \sqrt{\text{Var}[X_2]}, \ldots, \sqrt{\text{Var}[X_P]}\}$.
- It is common to report the correlation matrix $\left(\sqrt{\mathbf{V}}\right)^{-1} \text{Var}[\mathbf{X}] \left(\sqrt{\mathbf{V}}\right)^{-1}$, where the (i,j)th entry is $\rho(X_i, X_j)$.
- Consider $\mathbf{X}_i = (X_{i1}, X_{i2}, \ldots, X_{iP})'$ and $\bar{\mathbf{X}}_N = \frac{1}{N}\sum_{i=1}^{N} \mathbf{X}_i$.
- If $(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N)$ are i.i.d., and $\text{Var}[X_{ip}] < \infty$ for any $1 \leq p \leq P$, then

$$\sqrt{N}\left(\bar{\mathbf{X}}_N - \mathbb{E}[\mathbf{X}_i]\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, N\,\text{Var}\left[\bar{\mathbf{X}}_N\right]\right).$$

Multivariate delta method

- Let's define $\mu = \mathbb{E}[\mathbf{X}_i]$ and $\Sigma = N \operatorname{Var}\left[\bar{\mathbf{X}}_N\right]$.
- For a continuously differentiable function $g(\cdot)$ with the gradient $\nabla g(\mu)$ at μ, we have

$$\sqrt{N}\left(g\left(\bar{\mathbf{X}}_N\right) - g(\mu)\right) \xrightarrow{d} \mathcal{N}\Big(0,\ \nabla g(\mu)' \Sigma \nabla g(\mu)\Big).$$

- For the ratio estimator $\hat{\tau}_N = \frac{\bar{X}_N}{\bar{Y}_N}$, we know that $g(X, Y) = \frac{X}{Y}$ and $\nabla g(\mu) = \left(\frac{1}{\mu_Y}, -\frac{\mu_X}{\mu_Y^2}\right)'$, thus

$$\sqrt{N}\left(\hat{\tau}_N - \frac{\mu_X}{\mu_Y}\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma_X^2}{\mu_Y^2} + \frac{\sigma_Y^2 \mu_X^2}{\mu_Y^4} - \frac{2\sigma_{XY}\mu_X}{\mu_Y^3}\right).$$