Spektralzerlegung

Autor und Herausgeber

Alexander Jung*, Department of Computer Science, Aalto University, Espoo, Finland (ORCID 0000-0001-7538-0990)
Editors: Konstantina Olioumtsevits (ORCID 0000-0003-2697-8659), Ekkehard Schnoor (ORCID 0000-0001-8252-1499), Aalto University, Espoo, Finland
*Corresponding author: alex.jung@aalto.fi

Zusammenfassung

Eine Spektralzerlegung stellt eine symmetrische Matrix als Summe von orthogonale Projektionen auf zueinander senkrechte Geraden dar, von denen jede mit einem reellen Eigenwert skaliert ist. Es handelt sich um eine Eigenwertzerlegung, deren Eigenvektoren eine orthonormale Basis bilden, sodass die Matrix in einem gedrehten Koordinatensystem diagonal ist. Eine reelle Matrix besitzt genau dann eine solche Zerlegung, wenn sie symmetrisch ist, und über den komplexen Zahlen bilden die normale Matrizen die entsprechende Klasse. Die Zerlegung folgt aus der Maximierung der quadratischen Form $\mathbf{x}^\top \mathbf{A}\mathbf{x}$ über der Einheitssphäre: Ein Maximierer existiert, weil die Sphäre kompakt ist, die Symmetrie macht ihn zu einem Eigenvektor, und die Symmetrie macht wiederum die zu ihm orthogonalen Vektoren zu einem invarianten Unterraum, sodass sich das Argument um eine Dimension reduziert wiederholt. Die Eigenschaft, positiv semidefinit zu sein, kommt erst am Ende ins Spiel, wo sie gleichbedeutend damit ist, dass jeder Eigenwert nichtnegativ ist. Die Zerlegung liefert die Spektralnorm, den Rang und die Quadratwurzel einer Matrix, und sie liegt der Hauptkomponentenanalyse (PCA), den Kernel-Methoden sowie der Konvergenzanalyse des Gradientenabstieg zugrunde.

Definition

Streckt man die Ebene um einen Faktor 3 entlang einer Geraden und lässt eine dazu senkrechte Gerade unverändert, so wird ein Kreis zu einer Ellipse, deren Achsen entlang dieser beiden Geraden verlaufen. Eine Matrix, die auf diese Weise wirkt, ist

$$\mathbf{A} = \begin{pmatrix} 5/2 & \sqrt{3}/2 \\ \sqrt{3}/2 & 3/2 \end{pmatrix},$$

deren beide Geraden bei 30° und 120° liegen (siehe Abb. 1). Eine Spektralzerlegung besagt, dass jede symmetrische Matrix auf diese Weise wirkt, indem sie jede von d zueinander senkrechten Richtungen um ihren eigenen reellen Faktor skaliert.

Eine Spektralzerlegung ist eine Eigenwertzerlegung, deren Eigenvektoren or-

thonormal sind. Für eine reelle symmetrische Matrix $\mathbf{A} \in \mathbb{R}^{d\times d}$ lautet sie

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top = \sum_{j=1}^{d} \lambda_j \mathbf{u}^{(j)} \big(\mathbf{u}^{(j)}\big)^\top, \tag{1}$$

mit einer orthogonalen Matrix $\mathbf{U} = \big(\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(d)}\big)$ aus Eigenvektoren und reellen Eigenwerte $\lambda_1 \geq \cdots \geq \lambda_d$ auf der Diagonalen von $\mathbf{\Lambda}$. Die Umkehrung gilt ebenfalls, da $\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$ mit ihrer eigenen Transponierte übereinstimmt: Eine reelle Matrix mit einer orthonormalen Basis aus Eigenvektoren ist symmetrisch. Diese Äquivalenz ist als Spektralsatz bekannt (Axler, 2024, 7.29). Über den komplexen Zahlen gilt dieselbe Aussage für eine normale Matrix, mit der konjugierte Transponierte anstelle der Transponierte und einer unitäre Matrixen anstelle einer orthogonalen Matrix, und die Eigenwerte sind genau dann reell, wenn $\mathbf{A}$ hermitesch ist.

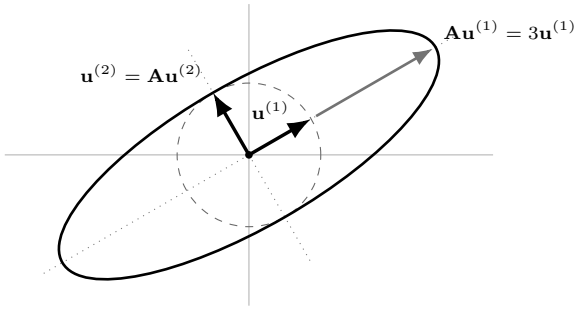


Abbildung 1: Die Matrix $\mathbf{A}$ aus der obigen abgesetzten Formel bildet den gestrichelten Einheitskreis auf die durchgezogene Ellipse ab. Die gepunkteten Geraden tragen die beiden Eigenvektoren $\mathbf{u}^{(1)}$ und $\mathbf{u}^{(2)}$ bei 30° und 120°, und die Achsen der Ellipse liegen entlang dieser Geraden mit den Halbachsenlängen $\lambda_1 = 3$ und $\lambda_2 = 1$. Keine der beiden Richtungen wird in eine andere gedreht: $\mathbf{A}$ streckt entlang $\mathbf{u}^{(1)}$ und belässt $\mathbf{u}^{(2)}$ an seinem Platz

Eine Herleitung des symmetrischen Falls stützt sich auf eine einzige analytische Tatsache und konstruiert die Eigenvektoren, anstatt ihre Existenz vorauszusetzen. Sie geht von der quadratischen Form $\mathbf{x}^\top\mathbf{A}\mathbf{x}$ eingeschränkt auf die Einheitssphäre $\mathbb{S}^{(d-1)} = \big\{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| = 1\big\}$ aus, wo sie mit dem Rayleigh-Quotienten $\mathbf{x}^\top\mathbf{A}\mathbf{x}/\mathbf{x}^\top\mathbf{x}$ übereinstimmt (Meyer, 2000, Sect. 7.5, Eq. (7.5.4)). Die Sphäre ist abgeschlossen und beschränkt, also kompakt, und die quadratische Form ist stetig, sodass sie auf der Sphäre in einem Punkt $\mathbf{u}^{(1)}$ ein Maximum annimmt (Rudin, 1976, Thm. 4.16). Man schreibe λ_1 für dieses Maximum. Dies ist der gesamte analytische Teil, und die verbleibenden Schritte sind Algebra.

Der Maximierer ist ein Eigenvektor. Man wähle einen zu $\mathbf{u}^{(1)}$ orthogonalen Vektor $\mathbf{v}$ mit Norm eins, sodass die Kurve $\mathbf{x}(t) = \big(\mathbf{u}^{(1)} + t\mathbf{v}\big)/\sqrt{1+t^2}$ auf der Sphäre verläuft, und kürze $c = \big(\mathbf{u}^{(1)}\big)^\top\mathbf{A}\mathbf{v}$ sowie $e = \mathbf{v}^\top\mathbf{A}\mathbf{v}$ ab. Entlang

dieser Kurve ist die quadratische Form $f(t) = \big(\lambda_1 + 2tc + t^2 e\big)/\big(1 + t^2\big)$, eine differenzierbare Funktion mit einem Maximum bei $t = 0$, also gilt $f'(0) = 2c = 0$. Hier kommt die Symmetrie ins Spiel: Es ist $c = \big(\mathbf{A}\mathbf{u}^{(1)}\big)^\top \mathbf{v}$, sodass der Vektor $\mathbf{A}\mathbf{u}^{(1)}$ orthogonal zu jedem zu $\mathbf{u}^{(1)}$ orthogonalen Vektor ist und folglich ein Vielfaches von $\mathbf{u}^{(1)}$ darstellt. Multiplikation von $\mathbf{A}\mathbf{u}^{(1)} = \lambda \mathbf{u}^{(1)}$ mit $\big(\mathbf{u}^{(1)}\big)^\top$ liefert $\lambda = \lambda_1$.

Das Argument wiederholt sich anschließend auf einer kleineren Matrix. Der Unterraum $\mathcal{W}$ der zu $\mathbf{u}^{(1)}$ orthogonalen Vektoren ist invariant unter $\mathbf{A}$, da $\big(\mathbf{u}^{(1)}\big)^\top \mathbf{A}\mathbf{w} = \big(\mathbf{A}\mathbf{u}^{(1)}\big)^\top \mathbf{w} = \lambda_1 \big(\mathbf{u}^{(1)}\big)^\top \mathbf{w} = 0$ für jedes $\mathbf{w} \in \mathcal{W}$ gilt. Die Symmetrie wird ein zweites Mal genutzt, und genau hier scheitert eine allgemeine Matrix: Für $\mathbf{N} = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$ spannt der Eigenvektor $(1, 0)^\top$ zwar eine invariante Gerade auf, doch $\mathbf{N}(0, 1)^\top = (1, 0)^\top$ verlässt das orthogonale Komplement, weshalb eine allgemeine Matrix überhaupt nicht diagonalisierbar sein muss. Fasst man eine orthonormale Basis von $\mathcal{W}$ zu einer Matrix $\mathbf{Q} \in \mathbb{R}^{d \times (d-1)}$ zusammen und setzt $\mathbf{M} = \mathbf{Q}^\top \mathbf{A}\mathbf{Q}$, so erhält man eine um eine Dimension kleinere symmetrische Matrix; per Induktion ergibt sich eine orthonormale Basis aus deren Eigenvektoren, und $\mathbf{Q}$ hebt diese wieder zu Eigenvektoren von $\mathbf{A}$ hoch, die orthogonal zu $\mathbf{u}^{(1)}$ sind. Jedes nachfolgende Maximum wird über einer kleineren Menge gebildet, sodass die Eigenwerte in absteigender Reihenfolge auftreten.

Die Nichtnegativität der Eigenwerte ist ein gesonderter Punkt, und hier kommt die Eigenschaft, positiv semidefinit zu sein, ins Spiel. Symmetrie ist eine Voraussetzung und keine Folgerung aus $\mathbf{x}^\top \mathbf{A}\mathbf{x} \geq 0$: Eine schiefsymmetrische Matrix $\mathbf{S}$ mit $\mathbf{S}^\top = -\mathbf{S}$ erfüllt $\mathbf{x}^\top \mathbf{S}\mathbf{x} = 0$ für jedes $\mathbf{x}$, sodass die quadratische Form keinen schiefsymmetrischen Anteil erkennen kann. Entwickelt man für eine symmetrische Matrix einen Vektor in der Basis der Eigenvektoren, so ergibt sich $\mathbf{x}^\top \mathbf{A}\mathbf{x} = \sum_{j=1}^{d} \lambda_j \big(\big(\mathbf{u}^{(j)}\big)^\top \mathbf{x}\big)^2$, was genau dann für jedes $\mathbf{x}$ nichtnegativ ist, wenn jeder Eigenwert nichtnegativ ist. Jedes $\mathbf{u}^{(j)}\big(\mathbf{u}^{(j)}\big)^\top$ in (1) ist die orthogonale Projektion auf eine Gerade, und diese Projektionen annullieren sich gegenseitig und summieren sich zur Identität, sodass eine symmetrische Matrix einen Vektor entlang d zueinander senkrechter Achsen zerlegt, jede Komponente skaliert und die Teile wieder aufaddiert. Die Eigenwerte sind durch $\mathbf{A}$ eindeutig bestimmt, die einzelnen Eigenvektoren hingegen nicht, wenn ein Eigenwert mehrfach auftritt: Jede orthonormale Basis des zugehörigen Unterraums leistet dasselbe, und eindeutig bestimmt ist lediglich die orthogonale Projektion auf diesen Unterraum.

Mehrere Größen lassen sich unmittelbar aus (1) ablesen. Die Spektralnorm ist λ_1, aufgrund derselben Maximierung, mit der die Herleitung begann, während $\operatorname{tr}(\mathbf{A})$ und $\det \mathbf{A}$ die Summe bzw. das Produkt der Eigenwerte sind. Der Rang von $\mathbf{A}$ zählt die von null verschiedenen Eigenwerte, und die zu den Null-Eigenwerten gehörenden Eigenvektoren spannen den Nullraum auf. Eine positiv semidefinite Matrix besitzt die Quadratwurzel $\mathbf{A}^{1/2} = \sum_{j=1}^{d} \sqrt{\lambda_j}\, \mathbf{u}^{(j)} \big(\mathbf{u}^{(j)}\big)^\top$, welche die einzige positiv semidefinite Matrix ist, die quadriert $\mathbf{A}$ ergibt, sodass $\mathbf{A} = \mathbf{B}^\top \mathbf{B}$ für $\mathbf{B} = \mathbf{A}^{1/2}$ gilt (Axler, 2024, 7.39). Führt man die Maximierung direkt über Unterräume statt schritt-

weise über einzelne Deflationen durch, so wird jeder Eigenwert auf dieselbe variationale Weise charakterisiert (Meyer, 2000, Sect. 7.5, Eq. (7.5.5)).

Das maschinelles Lernen (ML) begegnet symmetrischen Matrizen überall dort, wo eine quadratische Form auftritt. Die Stichproben-Kovarianzmatrix eines Datensatzes ist symmetrisch und positiv semidefinit, und die Hauptkomponentenanalyse (PCA) behält die Eigenvektoren ihrer größten Eigenwerte bei, sodass der obige variationale Schritt genau das ist, was die erste Hauptrichtung zur Richtung der größten Varianz macht. Eine Kernel-Matrix ist nach Konstruktion positiv semidefinit, und ihre Quadratwurzel stellt sie als Gram-Matrix von Merkmalsvektoren dar, was es einer Kernel-Methode überhaupt erst ermöglicht, ohne diese Merkmalsvektoren zu arbeiten. Die Trainingszielfunktion der lineare Regression besitzt die konstante Hesse-Matrix $(2/m)\mathbf{X}^\top\mathbf{X}$, deren Eigenwerte darüber entscheiden, wie schnell der Gradientenabstieg konvergiert, und die Laplace-Matrix eines ungerichteter Graphen ist symmetrisch und positiv semidefinit, was beim Spektrales Clustering zerlegt wird.

Zwei Bemerkungen zum Gültigkeitsbereich. Ersetzt man die Transponierte durch die konjugierte Transponierte, so überträgt sich die Herleitung wörtlich auf eine hermitesche Matrix, für welche die quadratische Form $\mathbf{x}^H\mathbf{A}\mathbf{x}$ reell ist. Im Unendlichdimensionalen bricht das Argument an genau einer Stelle zusammen, nämlich bei der Existenz des Maximierers, da die Einheitssphäre eines unendlichdimensionalen Hilbertraums nicht kompakt ist; setzt man stattdessen voraus, dass der Operator kompakt und selbstadjungiert ist, so ist dieser Schritt wiederhergestellt und der Rest der Herleitung verläuft unverändert (Young, 1988).

Siehe auch: Eigenwertzerlegung, normale Matrix, symmetrische Matrix, Eigenvektor, Eigenwert, positiv semidefinit, orthonormal, orthogonal, orthogonale Projektion, Spektrum, Spektralnorm, PCA, Stichproben-Kovarianzmatrix, Spektrales Clustering.

Querverweise

Literatur

1. Axler (2024). Linear Algebra Done Right. 4th ed., Springer, Cham, Switzerland.

2. Meyer (2000). Matrix Analysis and Applied Linear Algebra. SIAM, Philadelphia, PA, USA.

3. Rudin (1976). Principles of Mathematical Analysis. 3rd ed., McGraw-Hill, New York, NY, USA.

4. Young (1988). An Introduction to Hilbert Space. Cambridge Univ. Press, New York, NY, USA.