Eigenwertzerlegung

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 Eigenwertzerlegung (EVD) einer quadratischen Matrix zerlegt die durch sie dargestellte lineare Abbildung in elementare Operationen: das Entwickeln eines Vektors nach den Eigenvektoren der Matrix, das Multiplizieren jedes Entwicklungskoeffizienten mit dem entsprechenden Eigenwert und das anschließende Aufsummieren der skalierten Terme. Eine Matrix, die eine EVD besitzt, heißt diagonalisierbar; symmetrische Matrizen und Matrizen mit paarweise verschiedenen Eigenwerte sind diagonalisierbar. Ein Datensatz mit numerischen Merkmalsvektoren lässt sich durch die EVD seiner Stichproben-Kovarianzmatrix charakterisieren. Die EVD kann verwendet werden, um Berechnungen wie den Gradientenschritt von Gradientenabstieg-Verfahren zu beschleunigen. Bei der lineare Regression reduziert sich jede Gradientenabstieg-Iteration dann auf elementweise Operationen. Die EVD der Laplace-Matrix eines ungerichteter Graphen wird beim Spektrales Clustering genutzt.

Definition

Jeder Tag des Jahres 2024 an der Wetterstation Krems an der Donau (Österreich) ist ein Datenpunkt, der mit zwei Merkmale erfasst wurde, seiner Lufttemperatur um 06:00 Uhr (UTC) und um 15:00 Uhr (UTC), zusammengefasst zu einem Merkmalsvektor $\mathbf{x} = \big(x_1, x_2\big)^\top \in \mathbb{R}^2$. Abb. 1 zeigt 20 der $m = 365$ Tage $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m)}$ als Streudiagramm zusammen mit einer Ellipse, die aus ihrem Stichprobenmittel und ihrer Stichproben-Kovarianzmatrix berechnet wird. Der Stichprobenmittel der Tage ist

$$\widehat{\boldsymbol{\mu}} = \frac{1}{m} \sum_{r=1}^{m} \mathbf{x}^{(r)},$$

und ihre Stichproben-Kovarianzmatrix ist

$$\widehat{\mathbf{C}} = \frac{1}{m} \sum_{r=1}^{m} \big(\mathbf{x}^{(r)} - \widehat{\boldsymbol{\mu}}\big)\big(\mathbf{x}^{(r)} - \widehat{\boldsymbol{\mu}}\big)^\top \in \mathbb{R}^{2\times 2}.$$

Die Form der Ellipse, die den Datensatz umreißt, wird durch eine EVD der Stichproben-Kovarianzmatrix $\widehat{\mathbf{C}}$ bestimmt, deren Eigenvektoren die Achsen der Ellipse bilden.

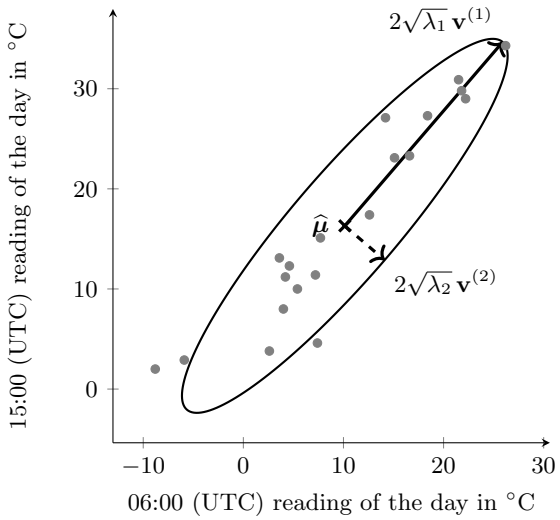


Abbildung 1: Jeder 19. der 365 Tage des Jahres 2024 in Krems an der Donau, insgesamt 20 Tage, dargestellt als graue Punkte. Jeder Tag ist ein Datenpunkt mit zwei Merkmale: dem Messwert um 06:00 Uhr (UTC) und dem um 15:00 Uhr (UTC) desselben Tages. Die Ellipse wird aus dem Stichprobenmittel (Kreuz) und der Stichproben-Kovarianzmatrix aller 365 Tage berechnet. Der durchgezogene Pfeil und der gestrichelte Pfeil sind die beiden Eigenvektoren dieser Stichproben-Kovarianzmatrix, skaliert auf die Halbachsen der Ellipse, $2\sqrt{\lambda_j}\,\mathbf{v}^{(j)}$. Daten erzeugt mit `pythondemos/evd.py`

Eine EVD für eine quadratische Matrix $\mathbf{A} \in \mathbb{R}^{d\times d}$ ist eine Faktorisierung der Form

$$\mathbf{A} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1}.$$

Die Spalten der Matrix $\mathbf{V} = \big(\mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(d)}\big)$ sind Eigenvektoren der Matrix $\mathbf{A}$. Die Diagonal-Matrix $\boldsymbol{\Lambda} = \operatorname{diag}\big\{\lambda_1, \ldots, \lambda_d\big\}$ enthält die zu den Eigenvektoren $\mathbf{v}^{(j)}$ gehörenden Eigenwerte λ_j. Multipliziert man die Faktorisierung mit $\mathbf{v}^{(j)}$, so erhält man $\mathbf{A}\mathbf{v}^{(j)} = \lambda_j\mathbf{v}^{(j)}$: Jeder Eigenvektor gibt eine Richtung an, die von $\mathbf{A}$ lediglich um den Faktor λ_j skaliert wird; ein einzelnes solches Paar erfüllt $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$, und die EVD fasst d davon, indiziert durch j, zusammen.

Die EVD zerlegt die lineare Abbildung $\mathbf{u} \mapsto \mathbf{A}\mathbf{u}$ in drei elementare Operationen. Ein Vektor $\mathbf{u}$ wird nach den Eigenvektoren entwickelt, $\mathbf{u} = \sum_{j=1}^{d} \beta_j\mathbf{v}^{(j)}$ mit den Entwicklungskoeffizienten $\boldsymbol{\beta} = \mathbf{V}^{-1}\mathbf{u}$; jeder Koeffizient wird mit seinem Eigenwert multipliziert; und die skalierten Terme werden wieder aufsummiert, $\mathbf{A}\mathbf{u} = \sum_{j=1}^{d} \lambda_j\beta_j\mathbf{v}^{(j)}$.

Im Gegensatz zur Singulärwertzerlegung (SVD), die für jede Matrix existiert, ob quadratisch oder nicht (Horn and Johnson, 2013, Thm. 2.6.3), ist die EVD nur für quadratische Matrizen definiert und muss nicht existieren. Zudem kann dieselbe Matrix verschiedene EVDs zulassen: Jede vom Nullvektor

verschiedene Linearkombination von Eigenvektoren zum selben Eigenwert ist wieder ein Eigenvektor zu diesem Eigenwert. Die geometrische Vielfachheit eines Eigenwerts ist die maximale Anzahl linear unabhängiger Eigenvektoren zu diesem Eigenwert (Horn and Johnson, 2013, Sect. 1.4).

Matrizen, die eine EVD zulassen, heißen diagonalisierbar. Ein Beispiel für eine diagonalisierbare Matrix ist

$$\mathbf{A} = \begin{pmatrix} 1.3 & 0.8 \\ 0.4 & 0.9 \end{pmatrix}.$$

Abb. 2 veranschaulicht ihre EVD.

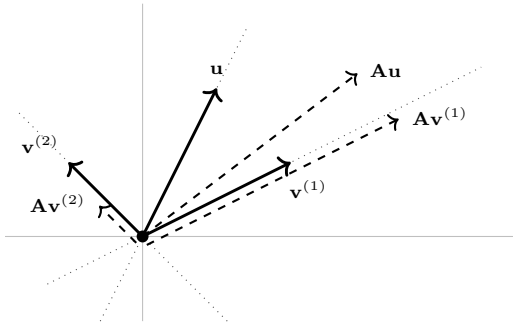


Abbildung 2: Die Matrix $\mathbf{A} \in \mathbb{R}^{2\times 2}$ aus der obigen Gleichung besitzt die Eigenwerte $\lambda_1 = 1.7$ und $\lambda_2 = 0.5$, mit den Eigenvektorenen $\mathbf{v}^{(1)}$ entlang $(2, 1)^\top$ und $\mathbf{v}^{(2)}$ entlang $(-1, 1)^\top$ (durchgezogene Pfeile). Die gestrichelten Pfeile sind die Bilder unter $\mathbf{A}$, und jede gepunktete Gerade trägt einen durchgezogenen Pfeil. Der Eigenvektor $\mathbf{v}^{(1)}$ bleibt auf seiner Geraden, gestreckt auf den längeren gestrichelten Pfeil $\mathbf{A}\mathbf{v}^{(1)} = \lambda_1\mathbf{v}^{(1)}$; der Eigenvektor $\mathbf{v}^{(2)}$ bleibt auf seiner Geraden, gestaucht auf den kürzeren gestrichelten Pfeil $\mathbf{A}\mathbf{v}^{(2)} = \lambda_2\mathbf{v}^{(2)}$. Der Vektor $\mathbf{u} = \mathbf{v}^{(1)} + \mathbf{v}^{(2)}$ verlässt seine gepunktete Gerade: Sein Bild $\mathbf{A}\mathbf{u} = \lambda_1\mathbf{v}^{(1)} + \lambda_2\mathbf{v}^{(2)}$ zeigt in eine andere Richtung, da seine beiden Entwicklungskoeffizienten mit unterschiedlichen Faktoren skaliert werden

Zwei hinreichende Bedingungen dafür, dass eine reellwertige quadratische Matrix diagonalisierbar ist, bestehen darin, dass sie eine symmetrische Matrix ist oder dass ihre d Eigenwerte paarweise verschieden sind (Golub and Van Loan, 2013; Horn and Johnson, 2013, Thm. 1.3.9, Thm. 4.1.5). Für eine reellwertige symmetrische Matrix weist die EVD eine zusätzliche Struktur auf: Die Eigenvektoren können so gewählt werden, dass sie eine orthonormale Basis bilden, d. h., die Matrix $\mathbf{V}$ in der Faktorisierung kann orthogonal gewählt werden (Horn and Johnson, 2013, Thm. 4.1.5). Die zugleich notwendige und hinreichende Bedingung lautet, dass jede Nullstelle des Minimalpolynoms der Matrix einfach ist (Horn and Johnson, 2013, Cor. 3.3.10). Für eine komplexwertige Matrix lautet die erste Bedingung, dass sie hermitesch ist; eine komplexe symmetrische

Matrix muss nicht diagonalisierbar sein (Horn and Johnson, 2013, Thm. 4.1.5, Problem 4.4.P10).

Die Matrix $\mathbf{A}$ aus Abb. 2 ist nicht symmetrisch, da ihre Außerdiagonaleinträge 0.8 und 0.4 voneinander abweichen, aber ihre beiden Eigenwerte 1.7 und 0.5 sind verschieden, sodass sie eine EVD besitzt. Ist keine der beiden Bedingungen erfüllt, existiert möglicherweise keine EVD: Die Matrix

$$\mathbf{N} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

ist nicht symmetrisch und besitzt den Eigenwert 0 doppelt, da die Eigenwerte einer Dreiecks-Matrix ihre Diagonaleinträge sind (Horn and Johnson, 2013, Sect. 1.2, Übung zu oberen Dreiecksmatrizen nach Def. 1.2.9). Die Eigenvektoren von $\mathbf{N}$ liegen alle auf der Geraden $\{(a, 0)^\top : a \in \mathbb{R}\}$, sodass es keine invertierbare Matrix $\mathbf{V}$ gibt, deren Spalten Eigenvektoren von $\mathbf{N}$ sind.

Die Ellipse in Abb. 1 ist die Menge der Merkmalsvektoren $\mathbf{x}$, für die die quadratische Form $(\mathbf{x} - \widehat{\boldsymbol{\mu}})^\top \widehat{\mathbf{C}}^{-1}(\mathbf{x} - \widehat{\boldsymbol{\mu}})$ gleich 4 ist. Diese Menge lässt sich mithilfe einer EVD der Stichproben-Kovarianzmatrix $\widehat{\mathbf{C}}$ bequemer charakterisieren. Die Matrix $\widehat{\mathbf{C}}$ ist symmetrisch und besitzt daher eine EVD $\widehat{\mathbf{C}} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^\top$ mit einer orthogonalen Matrix $\mathbf{V}$, die in ihren Spalten die Eigenvektoren enthält. Transformiert man den Merkmalsvektor $\mathbf{x}$ eines Tages gemäß $\widetilde{\mathbf{x}} := \mathbf{V}^\top(\mathbf{x} - \widehat{\boldsymbol{\mu}})$, so geht die quadratische Form über in

$$(\mathbf{x} - \widehat{\boldsymbol{\mu}})^\top \widehat{\mathbf{C}}^{-1}(\mathbf{x} - \widehat{\boldsymbol{\mu}}) = \widetilde{\mathbf{x}}^\top \boldsymbol{\Lambda}^{-1} \widetilde{\mathbf{x}} = \frac{\widetilde{x}_1^2}{\lambda_1} + \frac{\widetilde{x}_2^2}{\lambda_2},$$

da $\widehat{\mathbf{C}}^{-1} = \mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{V}^\top$ gilt. Die Ellipse ist somit die Menge, für die $\widetilde{x}_1^2/\lambda_1 + \widetilde{x}_2^2/\lambda_2 = 4$ gilt, mit den Halbachsen $2\sqrt{\lambda_1}$ und $2\sqrt{\lambda_2}$ entlang der Koordinatenachsen von $\widetilde{\mathbf{x}}$, also entlang der Eigenvektoren. Die transformierten Tage haben die Stichproben-Kovarianzmatrix $\mathbf{V}^\top \widehat{\mathbf{C}} \mathbf{V} = \boldsymbol{\Lambda}$: Die beiden neuen Koordinaten haben die empirische Kovarianz null, und die Eigenwerte sind ihre Stichprobenvarianzen. Ersetzt man jeden Tag allein durch seine erste Koordinate $\widetilde{x}_1$, so bleibt die Richtung der größten Varianz erhalten, was der Grundgedanke der Hauptkomponentenanalyse (PCA) ist (Bishop, 2006, Sect. 12.1.1).

Neben der Beschreibung der Form eines Datensatzes kann die EVD verwendet werden, um die Implementierung einer maschinelles Lernen (ML)-Methode zu beschleunigen. Betrachtet sei das Gradientenabstieg für die lineare Regression mit der Aktualisierung $\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} - (2\eta/m)\mathbf{X}^\top\big(\mathbf{X}\mathbf{w}^{(t)} - \mathbf{y}\big)$ für $t = 0, 1, \ldots$, mit der Merkmalsmatrix $\mathbf{X}$ und dem Labelvektor $\mathbf{y}$ des Trainingsdatensatzes (siehe lineare Regression). Die Matrix $\mathbf{X}^\top\mathbf{X}$ ist symmetrisch und besitzt daher eine EVD $\mathbf{X}^\top\mathbf{X} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^\top$ mit einer orthogonalen Matrix $\mathbf{V}$. In den transformierten Koordinaten $\widetilde{\mathbf{w}}^{(t)} := \mathbf{V}^\top\mathbf{w}^{(t)}$ entkoppelt sich die Aktualisierung zu

$$\widetilde{w}_j^{(t+1)} = \big(1 - (2\eta/m)\lambda_j\big)\widetilde{w}_j^{(t)} + (2\eta/m)\big(\mathbf{V}^\top\mathbf{X}^\top\mathbf{y}\big)_j.$$

Nach der einmaligen Berechnung der EVD besteht jede Iteration aus einer elementweisen Multiplikation und Addition von Vektoren der Länge d

(`pythondemos/evd.py`).
Eine weitere Anwendung der EVD im ML ist das Clustering von graphstrukturierten Daten (von Luxburg, 2007). Betrachtet seien Datenpunkte, die als die Knoten $i = 1, \ldots, n$ eines ungerichteter Graphen $\mathcal{G}$ dargestellt sind, z. B. die Nutzer eines sozialen Netzwerks, dessen Kanten durch Freundschaften gegeben sind. Die Laplace-Matrix $\mathbf{L}^{(\mathcal{G})}$ des Graphen $\mathcal{G}$ ist symmetrisch und positiv semidefinit (von Luxburg, 2007, Prop. 1). Sie besitzt eine EVD $\mathbf{L}^{(\mathcal{G})} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\top$ mit einer orthogonalen Matrix $\mathbf{V}$ und nichtnegativen Eigenwerte.

Ist der Graph $\mathcal{G}$ zusammenhängend, so weist die EVD eine zusätzliche Struktur auf: Der Eigenwert 0 hat die Vielfachheit eins, mit dem Eins-Vektor als zugehörigem Eigenvektor (von Luxburg, 2007, Prop. 2). Darüber hinaus gibt es dann einen ausgezeichneten Eigenvektor, den sogenannten Fiedler-Vektor: einen Eigenvektor zum kleinsten von null verschiedenen Eigenwert, der algebraische Konnektivität von $\mathcal{G}$ (Fiedler, 1973). Der Fiedler-Vektor kann zum Clustern der Knoten des Graphen verwendet werden, was die Idee des Spektrales Clusterings ist (von Luxburg, 2007). Abb. 3 zeigt den Fiedler-Vektor eines zusammenhängenden Graphen mit sechs Knoten: zwei Cluster von je drei Knoten und eine einzelne Kante, die die beiden Cluster verbindet.

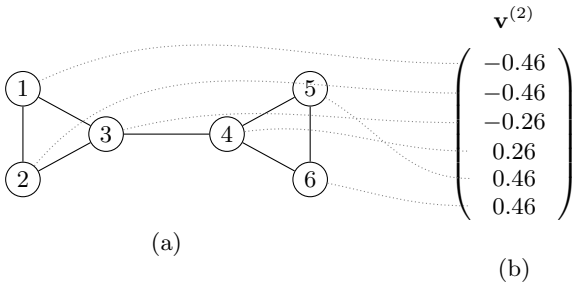


Abbildung 3: (a) Ein zusammenhängender ungerichteter Graph mit sechs Knoten: das Cluster der Knoten 1, 2, 3 und das Cluster der Knoten 4, 5, 6, wobei jeder Knoten mit den beiden anderen seines Clusters verbunden ist. Die beiden Cluster sind über die einzelne Kante zwischen den Knoten 3 und 4 verbunden. (b) Der Fiedler-Vektor $\mathbf{v}^{(2)}$ der Laplace-Matrix des Graphen in (a), als Spalten-Vektor; eine dünne gepunktete Linie verbindet jeden Knoten i mit seinem Eintrag $v_i^{(2)}$. Die Vorzeichen der Einträge stellen die beiden Cluster wieder her. Einträge berechnet mit `pythondemos/evd.py`

Siehe auch: Matrix, Eigenvektor, Eigenwert, diagonalisierbar, invertierbar, orthogonal, Laplace-Matrix, algebraische Konnektivität, Fiedler-Vektor, Spektrales Clustering, Gradientenabstieg, lineare Regression, Stichproben-Kovarianzmatrix, PCA, Spektralzerlegung, SVD, Minimalpolynom.

Querverweise

Literatur

1. Bishop (2006). Pattern Recognition and Machine Learning. Springer Science+Business Media, New York, NY, USA.

2. Fiedler (1973). Algebraic connectivity of graphs. Czechoslovak Mathematical Journal, 23(2), pp. 298–305.

3. Golub and Van Loan (2013). Matrix Computations. 4th ed., The Johns Hopkins Univ. Press, Baltimore, MD, USA.

4. Horn and Johnson (2013). Matrix Analysis. 2nd ed., Cambridge Univ. Press, New York, NY, USA.

5. von Luxburg (2007). A tutorial on spectral clustering. Statist. Comput., 17(4), pp. 395–416.