VORLESUNG

ANALYTISCHE CHEMIE III

Chemometrie

Lineare Modelle, Optimierung und Kalibration

Prof. Dr. E. Pretsch

Das vorliegende Skript entstand unter Mitarbeit von

Dr. M. Badertscher

Dr. A. Bezegh

T. Brodmeier

A. Gloor

Zürich, Oktober 2002

Inhaltsverzeichnis

1. Einleitung, Literatur

Der Begriff "Chemometrie" ist nicht scharf definiert. Sie ist eine *chemische* Disziplin, die verschiedene mathematische und statistische Verfahren einsetzt, um optimale Messverfahren auszuwählen und um die Messdaten zu interpretierbarer Information weiterzuverarbeiten. Alle formellen Methoden der analytischen Chemie für die Optimierung einer Methode oder einer Messung, sowie für die Auswertung von Messdaten werden üblicherweise zur Chemometrie gezählt.

Viele der chemometrischen Techniken sind längst bekannte statistische Verfahren. Sie sind jedoch aus verschiedenen Gründen erst in letzter Zeit in der analytisch-chemischen Praxis relevant geworden. Die rasche Verbreitung chemometrischer Techniken hat mit der Allgegenwart des Computers im Laboratorium zu tun. Computerisierte Instrumente liefern häufig eine so grosse Datenmenge, dass die Resultate ohne mathematisch-statistische Verfahren gar nicht analysierbar sind. Vermehrt werden an Instrumentencomputern chemometrische Programme mitgeliefert. Ihre Anwendung ohne Grundkenntnisse der Methoden kann irreführende Resultate produzieren.

Da ein Charakteristikum vieler chemometrischer Techniken die Mehrdimensionalität ist, sind die Grundlagen der Methoden mit der Matrixalgebra einfach und kompakt darstellbar. Im Rahmen der Vorlesung werden deshalb zunächst die für die Chemometrie wichtigen Aspekte der Matrixalgebra rekapituliert. Dabei wird viel Gewicht auf die anschauliche, intuitiv zugängliche Interpretation gelegt. Übungen mit Tabellenkalkulation sollen helfen, das Gefühl für die grundlegenden Zusammenhänge zu entwickeln.

Anschliessend wird die Anpassung linearer Modelle an Daten behandelt. Eine der wichtigsten Anwendungen der Modellbildung ist die Optimierung und die damit verbundene Versuchsplanung. Deshalb wird die Anwendung linearer Modelle in diesem Zusammenhang ausführlich geübt. Schliesslich werden multivariate Kalibrationstechniken vorgestellt.

1.1. LITERATUR

Alexander Basilevsky
Applied matrix algebra in the statistical sciences
North-Holland, New York, 1983.

Ch.K. Bayne, I.B. Rubin
Practical experimental design and optimization methods for chemists
VCH, Weinheim, 1986.

L.M.C. Buydens, P.J. Schoenmakers
Intelligent Software for Chemical Analysis
Elsevier, Amsterdam, 1993.

H.M. Cartwright
Application of Artificial Intelligence in Chemistry
Oxford University Press, Oxford, 1993.

S.N. Deming, S.L. Morgan
Experimental design: A chemometric approach
Elsevier, Amsterdam, 2. Auflage 1993.

N.R. Draper, H. Smith
Applied regression analysis
Wiley, New York, 2. Auflage 1981.

H. Freiser
Concepts and Calculations in Analytical Chemistry: A Spredsheet Approach
CRC Press, London, 1992.

P. Gans
Data Fitting in the Chemical Sciences
Wiley, Chichester 1992.

D.E. Goldberg
Genetic algorithms
Addison-Wesley, Reading, 1989.

P.E. Green
Mathematical tools for applied multivariate analysis
Academic Press, 1976.

R. Henrion, G. Henrion
Multivariate Datenanalyse
Springer-Verlag, Berlin, 1995.

D. Huff
How to lie with statistics
Gollancz, London, 1954.

J.H. Kalivas, P.M. Lang
Mathematical analysis of spectral orthogonality
Marcel Dekker, New York, 1994.

E.R. Malinowski, D.G. Howery
Factor analysis in chemistry
Wiley, New York, 1980.

H. Martens, T. Naes
Multivariate calibration
Wiley, New York, 1989.

D.L. Massart, B.G.M. Vanderginste, L.M.C. Buydens, S. de Jong, P.J. Lewi, J. Smeyers-Verbeke
Handbook of Chemometrics and Qualimetrics: Parts A and B
Elsevier, Amsterdam, 1997, 1998.

P.C. Meier, R.E. Zünd
Statistical Methods in Analytical Chemistry
Wiley, New York, 1993.

D.C. Montgomary, E.A. Peck
Introduction to Linear Regression Models
Wiley: New York, 1991.

M. Otto
Chemometrie, Statistik und Computereinsatz in der Analytik
VCH, Weinheim, 1997

John Allen Paulos
Innumeracy: mathematical illiteracy and its consequences
(Third printing), Hill and Wang, New York, 1989.

F.P. Pitard
Pierry Gy's Sampling Theory and Sampling Practice
CRC Press, Boca Raton, 1989.

Shayle R. Searle
Matrix algebra useful for statistics
Wiley, New York,1982.

M. A. Sharaf, D.L. Illman, B. Kowalski,
Chemometrics
Wiley, New York, 1986.

K. Varmuza
Pattern recognition in chemistry
Lecture Notes in Chemistry Nr. 21, Springer, Berlin, 1980.

F. Walters, R. Parker, Jr., S.L. Morgan, S.N. Deming,
Sequential Simplex Optimization: A Technique for Improving Quality and Productivity in Research, Development an Manufacturing
CRC Press, London, 1991.

M. Wheeler
Lies, damn lies and statistics: The manipulation of public opinion in America
Liveright, 1976.

J. Zupan, J. Gasteiger
Neural networks for chemists
VCH, Weinheim, 1993.

2. Vektoren und Matrizen, Excel

Viele Teilgebiete der Chemometrie können als Verallgemeinerungen der Statistik eindimensionaler Daten in die Statistik vieldimensionaler Daten verstanden werden. Die Algebra wird durch eine Matrixalgebra ersetzt. Demenstprechend ist ein Denken in Matrizen für das Verständnis der Chemometrie vor zentraler Bedeutung. Da die Matrixalgebra eine kompakte Schreibweise komplexer Zusammenhänge erlaubt, sind die Zusammenhänge nach einer Gewöhnung an die Matrixschreibweise sogar einfacher verständlich.

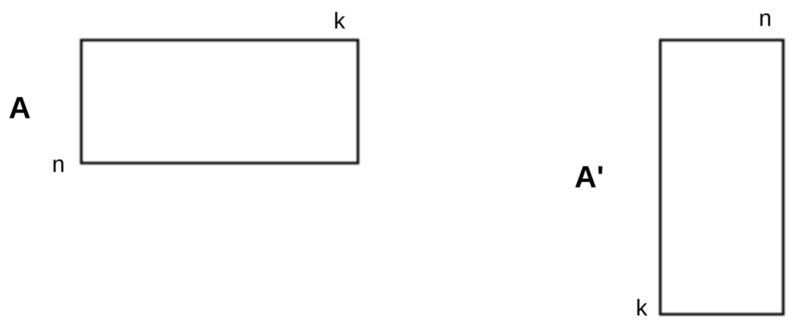
Im Rahmen dieser Vorlesung werden nur die notwendigsten Grundlagen der Matrixalgebra rekapituliert. Eine ausführliche Behandlung findet sich in den Werken von Basilevsky (umfassend aber etwas unübersichtlich, viele Druckfehler), Searle und Green (sehr einfache und anschauliche Behandlung der notwendigsten Grundlagen, vgl. Literaturliste). Geometrische Analogien sind für das Verständis der Matrixoperationen oft hilfreich, und werden deshalb hier besonders berücksichtigt.

2.1. DEFINITIONEN

Eine *Matrix* ist ein rechteckiges Zahlenschema. Sie wird üblicherweise mit einem fettgedruckten Grossbuchstaben symbolisiert. Das Element der i-ten Reihe und j-ten Kolonne einer Matrix $\mathbf{A}_{nk}$ ist a_{ij}.

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & . & . & a_{1k} \\ a_{21} & a_{22} & . & . & a_{2k} \\ . & . & . & . & . \\ . & . & . & . & . \\ a_{n1} & a_{n2} & . & . & a_{nk} \end{bmatrix}$$

Durch Vertauschen der Zeilen und Kolonnen erhält man die Transponierte einer Matrix. Die Transponierte der Matrix **A** wird als $\mathbf{A}'$ oder als $\mathbf{A}^T$ bezeichnet.



Eine Matrix, mit einer einzigen Kolonne ist ein *Kolonnenvektor* (mit fettgedrucktem Kleinbuchstaben symbolisiert), eine mit einer einzigen Zeile ist ein Zeilenvektor (die Transponierte eines Kolonnenvektors, mit fettgedrucktem Kleinbuchstaben mit Apostroph oder mit hochgestelltem Index T symbolisiert).

Eine *quadratische Matrix* hat gleich viele Zeilen wie Kolonnen. Eine quadratische Matrix ist *symmetrisch*, wenn für alle Werte von i und j $a_{ij}=a_{ji}$ ($\mathbf{A}=\mathbf{A}^T$). Wenn alle Elemente ausser den Diagonalelementen (a_{ii}) gleich null sind, ist die quadratische Matrix *diagonal*.

Die *Einheitsmatrix* **I** (auch $\mathbf{I_n}$ bezeichnet, wenn man die Ordnung n explizite mitteilen will) ist eine diagonale Matrix mit den Diagonalelementen von 1.

Die *Nullmatrix* **0** ist eine Matrix mit allen Elementen $a_{ij}=0$. $\mathbf{1_n}$ ($\mathbf{1_n^T}$) ist ein Kolonnenvektor (Zeilenvektor) der Länge n mit den Elementen 1.

2.2. EINFACHE OPERATIONEN

Addition

Bei der Matrixaddition ist jedes Element der Summenmatrix die Summe der entsprechenden Elemente der Summanden: **C** = **A** + **B** bedeutet, dass $c_{ij}=a_{ij}+b_{ij}$ für alle i und j. Für die Transponierte gilt: $\mathbf{C}^T = \mathbf{A}^T + \mathbf{B}^T$.

Multiplikation mit einem Skalar

Bei einer Multiplikation mit einem Skalar wird jedes Element der Matrix mit dem Skalar multipliziert: $\mathbf{E} = k\mathbf{A}$ bedeutet, dass $e_{ij}=ka_{ij}$ für alle i und j. Die Multiplikation mit einem Skalar ist assoziativ ($k_1(k_2\mathbf{A})=(k_1k_2)\mathbf{A}$) und distributiv ($(k_1+k_2)\mathbf{A}=k_1\mathbf{A}+k_2\mathbf{A}$). Im weiteren gilt: $\mathbf{A}k\mathbf{B}=k\mathbf{AB}$.

Matrixmultiplikation

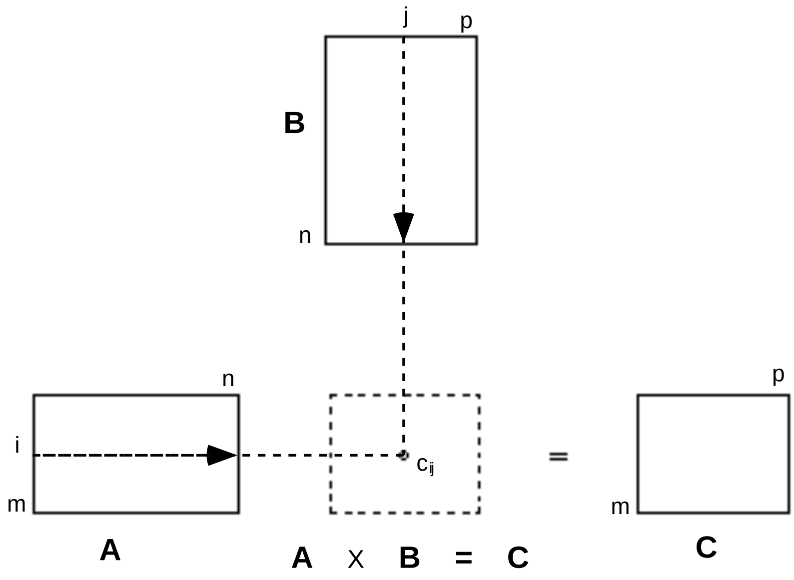
Eine Matrix mit m Zeilen und n Kolonnen $\mathbf{A_{mn}}$ kann mit einer Matrix mit n Zeilen (und beliebiger Anzahl Kolonnen) rechtsmultipliziert und mit einer Matrix mit m Kolonnen linksmultipliziert werden. Das Produkt einer Multiplikation $\mathbf{A_{mn}B_{np}}=\mathbf{C_{mp}}$ wird Matrixelemente haben, die folgendermassen bestimmt sind:

$$c_{ij}=\sum_{k=1}^{n} a_{ik}b_{kj}.$$

Auch wenn die Dimensionen der Matrizen die Bildung beider Produkte **AB** und **BA** erlauben, ist im allgemeinen $\mathbf{AB}\neq\mathbf{BA}$ (die Matrixmultiplikation ist nicht kommutativ).

Diesen Regeln entsprechend resultiert aus der Prämultiplikation einer Matrix mit einem Zeilenvektor ein Zeilenvektor ($\mathbf{a}^T$) und aus der Postmultiplikation einer Matrix mit einem Kolonnenvektor (**a**) ein Kolonnenvektor. Das Produkt $\mathbf{ab}^T$ (Kolonnenvektor rechtsmultipliziert mit einem Zeilenvektor) ist eine Matrix (äusseres Produkt), und $\mathbf{a}^T\mathbf{b}$ ein Skalar (inneres Produkt). $\mathbf{a}^T\mathbf{a}$ ist die Summe der Quadrate der einzelnen Komponenten von **a**.

Die Multiplikation mit einer Einheitsmatrix lässt die Matrix unverändert: **AI**=**A** und **IA**=**A**.



Die Matrixmultiplikation ist assoziativ: ((**AB**)**C**=**A**(**BC**)) und distributiv (**A**(**B**+**C**)=**AB**+**AC** sowie (**B**+**C**)**A**=**BA**+**CA**). Die Transponierte eines Produkts ist gleich dem Produkt der transponierten Matrizen in umgekehrter Reihenfolge: $(\mathbf{ABC})^T=\mathbf{C}^T\mathbf{B}^T\mathbf{A}^T$.

Im Gegensatz zur Multiplikation von skalaren Grössen bedeutet **AB**=**AC** (bei **A**≠0) nicht dass **B**=**C**. Analogerweise bedeutet **AB**=0 nicht, dass entweder **A** oder **B** gleich null ist. Das Produkt einer Matrix mit ihrer Transponierten ($\mathbf{AA}^T$ und auch $\mathbf{A}^T\mathbf{A}$) ist eine symmetrische Matrix.

Die Prämultiplikation einer Matrix mit einer diagonalen Matrix **D** führt zu einer Matrix, in der jedes Element einer Zeile durch das ursprüngliche Element × das Element in der entsprechenden Zeile der Diagonalmatrix gegeben ist:

$$\underset{\mathbf{D}}{\begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}} \underset{\mathbf{A}}{\begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \\ 3 & 2 & 1 & 1 \end{bmatrix}} = \underset{\mathbf{DA}}{\begin{bmatrix} 3 & 6 & 9 & 12 \\ 4 & 2 & 8 & 6 \\ 3 & 2 & 1 & 1 \end{bmatrix}}$$

Analoges gilt für die Kolonnen einer Matrix nach Postmultiplikation mit einer diagonalen Matrix.

Excel

Bei der Matrixmultiplikation mit dem Tabellenkalkulationsprogramm Excel muss die richtige Dimension der resultierenden Matrix aktiviert werden. Anschliessend ist der Befehl "=MMULT(ai:cj,dk:el)" (ai:cj und dk:el definieren dabei die beiden zu multiplizierenden Matrizen, sie entsprechen der oberen linken und unteren rechten Ecke). Der Befehl ist eine "Arrayoperation" und muss mit "Command-Enter" eingegeben werden.

Übung

Berechnen Sie $\mathbf{M} = \frac{1}{n} \mathbf{1_n} \mathbf{1_n^T} \mathbf{X}$ und $\mathbf{Y}=\mathbf{X}-\mathbf{M}$

Determinanten

Die Determinante einer Matrix **A** (bezeichnet durch "det **A**" oder durch "|**A**|") ist eine nur für quadratische Matrizen definierte Grösse. Sie ist eine Zahl, die eine Funktion der Matrixelemente ist. Da die allgemeine Definition einer Determinante recht unanschaulich ist, soll hier eine Prozedur zur Ermittlung der Determinante kurz skizziert werden:

1. Man soll alle Produkte ermitteln, die aus den Elementen der m×m Matrix gebildet werden können, so dass aus jeder Zeile und jeder Kolonne nur ein Faktor genommen wird (es gibt m! Produkte, d.h. 6 für eine 3×3 Matrix):

$$a_{11}\,a_{22}\,a_{33} \qquad a_{12}\,a_{23}\,a_{31} \qquad a_{13}\,a_{21}\,a_{32}$$
$$a_{13}\,a_{22}\,a_{31} \qquad a_{11}\,a_{23}\,a_{32} \qquad a_{12}\,a_{21}\,a_{33}$$

2. Die Produkte sind so geordnet, dass die Subskripte der Zeilen von links nach rechts zunehmen. Nun soll jeweils gezählt werden, wievielmal die Elemente vertauscht werden müssen, damit die Kolonnensubskripte von links nach rechts zunehmen (es sind 0,2,2 bzw 3,1, und 1). Jedes Produkt soll nun mit $(-1)^t$ multipliziert werden, wobei t die Anzahl der Tauschoperationen ist.

3. Die einzelnen Produkte sollen aufaddiert werden. Für die 3×3 Matrix ist:
$|\mathbf{A}|=a_{11}a_{22}a_{33}+a_{12}a_{23}a_{31}+a_{13}a_{21}a_{32}-a_{13}a_{22}a_{31}-a_{11}a_{23}a_{32}-a_{12}a_{21}a_{33}$

Die Determinanten haben einige wichtige Eigenschaften:

- Wenn zwei Zeilen oder zwei Kolonnen einer Matrix **A** vertauscht werden, so gilt für die Determinante der so erhaltenen Matrix **B**, dass |**B**|= -|**A**|.

- Die Determinante einer Matrix ist gleich null, wenn alle Elemente einer Zeile oder einer Kolonne gleich null sind.

- Die Determinante einer Matrix ist gleich null, wenn zwei Zeilen oder zwei Kolonnen gleich gross sind. Noch allgemeiner: sie ist gleich null, wenn irgendeine Zeile (oder Kolonne) als Linearkombination von anderen Zeilen (Kolonnen) ausgedrückt werden kann.

- Die Determinante einer Matrix und ihrer Transponierten sind gleich gross ($|\mathbf{A}|=|\mathbf{A}^T|$).

- Die Determinante des Produkts zweier Matrizen ist das Produkt der Determinanten der beiden Matrizen: $|\mathbf{AB}|=|\mathbf{A}||\mathbf{B}|$.

- Wenn jedes Element einer Zeile (Kolonne) mit einem Skalar multipliziert wird, und das Produkt zu den entsprechenden Elementen einer anderen Zeile (Kolonne) addiert wird, bleibt die Determinante unverändert.

- Eine Matrix heisst singulär, wenn ihre Determinante gleich null ist.

- Der Rang einer Matrix entspricht der Ordnung der grössten (quadratischen) Submatrix, deren Determinante ungleich null ist. Eine Submatrix kann aus einer Matrix durch Weglassen einer beliebigen Anzahl von Zeilen und/oder Kolonnen gebildet werden.

Excel

Die Determinante einer Matrix kann mit dem Tabellenkalkulationsprogramm Excel mit dem Befehl "=MDETERM(ai:ck)" berechnet werden.

Inverse einer Matrix

Eine nichtsinguläre quadratische Matrix **A** hat eine Inverse $\mathbf{A}^{-1}$, die der folgenden Relationen genügt:

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I} \qquad (\mathbf{I} \text{ ist die Einheitsmatrix})$$

Für eine 2×2 Matrix ist die Berechnung der Inversen sehr einfach:

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|}\begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \qquad |\mathbf{A}|=a_{11}a_{22}-a_{12}a_{21}$$

Die Inversion von nicht allzugrossen Matrizen kann mit Computerprogrammen (inklusive einiger Tabellenkalkulationsprogramme, z.B. Excel) leicht durchgeführt werden. (Excel Befehl: "=MINVERSE(ai:ck)").

Die Inverse eines Produktes ist das Produkt der Inversen in umgekehrter Reihenfolge:

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}.$$

Die Inverse der Inversen ist die Matrix selbst:

$$(\mathbf{A}^{-1})^{-1} = \mathbf{A}.$$

Die Inverse einer Matrix, multipliziert mit einem Skalar, ist die Inverse multipliziert mit dem reziproken Skalar:

$$(k\mathbf{A})^{-1} = (1/k)\mathbf{A}^{-1}.$$

Die Inverse der Transponierten ist gleich der Transponierten der Inversen:

$$(\mathbf{A}^{T})^{-1} = (\mathbf{A}^{-1})^{T}.$$

Die Inverse einer diagonalen Matrix ist eine diagonale Matrix, deren Elemente gleich dem Reziproken der ursprünglichen Diagonalelemente sind:

$$\mathbf{A}^{-1} = \mathrm{diag}(d_{ii}^{-1}).$$

Orthogonale Matrizen

Eine quadratische Matrix **C** ist orthogonal, wenn $\mathbf{C}^{T}\mathbf{C}=\mathbf{I}=\mathbf{C}\mathbf{C}^{T}$ (d.h. die Transponierte entspricht der Inversen, $\mathbf{C}^{-1}=\mathbf{C}^{T}$). Die Zeilenvektoren und die Kolonnenvektoren einer orthogonalen Matrix sind zueinander orthonormal (d.h. $\mathbf{u}^{T}\mathbf{u}=1$ und $\mathbf{u}^{T}\mathbf{v}=0$, vgl. unten). Nicht jede Matrix mit orthonormalen Zeilenvektoren oder orthonormalen Kolonnenvektoren ist orthogonal:

Bsp.: $\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ enthält zwar orthonormale Zeilenvektoren, sie ist aber keine orthogonale Matrix.

Idempotente Matrizen

Eine Quadratische Matrix **M** ist idempotent wenn **MM**=**M**. Idempotente Matrizen sind (ausser einem Spezialfall: der Einheitsmatrix) immer singulär. Sie sind Projektionsmatrizen (vgl. unten). Wenn eine Matrix **M** symmetrisch und idempotent ist, dann entspricht die Multiplikation mit dieser Matrix einer orthogonalen Projektion. Im weiteren gilt dann: $(\mathbf{I}-2\mathbf{M})^T(\mathbf{I}-2\mathbf{M}) = \mathbf{I}$ d.h. die Matrix $(\mathbf{I}-2\mathbf{M})$ ist orthogonal und nicht singulär. Symmetrische, idempotente Matrizen sind in der multivariaten Statistik von besonderer Bedeutung.

Bsp.: eine Zentrierungsmatrix $\mathbf{H_n} = \mathbf{I} - \frac{1}{n}\mathbf{1_n}\,\mathbf{1_n^T}$ ist idempotent so dass:

$$(\mathbf{H}_n\mathbf{X})^T\mathbf{H}_n\mathbf{X} = \mathbf{X}^T\mathbf{H_n^T}\mathbf{H}_n\mathbf{X} = \mathbf{X}^T\mathbf{H}_n\mathbf{X}.$$

Spur einer Matrix

Die Spur (engl. trace) einer quadratischen Matrix ist die Summe der Diagonalelemente. Spur (**AB**) = Spur (**BA**) (damit beide Produkte definiert sind, müssen die Dimensionen $\mathbf{A_{pq}}$ und $\mathbf{B_{qp}}$ sein). Dementsprechend gilt für die Spur der aus einer n×p Matrix **X** gebildeten Projektionsmatrix: $\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ (vgl. später):

$\mathrm{Spur}(\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T) = \mathrm{Spur}(\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}) = \mathrm{Spur}(\mathbf{I_p}) = p$ (Man soll beachten, dass die Projektionsmatrix eine n×n Matrix ist).

Die Spur einer symmetrischen, idempotenten Matrix ist gleich ihrem Rang. Die Dimension des Raumes, in die man durch eine aus $\mathbf{X_{np}}$ gebildete Projektionsmatrix projiziert, ist also p.

Die Quadratsumme aller Elemente einer Matrix $\mathbf{A_{nm}}$ kann mit der Spur einfach ausgedrückt werden:

$$\mathrm{Spur}(\mathbf{A}^T\mathbf{A}) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}^2$$

Eigenwerte und Eigenvektoren von Matrizen

Gewisse quadratische Matrizen **A** können in ein Produkt folgender Art aufgespalten werden:

$$\mathbf{A} = \mathbf{P}\,\boldsymbol{\Lambda}\,\mathbf{P}^{-1}$$

$\boldsymbol{\Lambda}$ ist dabei eine diagonale Matrix. Die Diagonalelemente λ_i sind die Eigenwerte der Matrix **A**. Die Kolonnenvektoren von **P** sind ihre Eigenvektoren.

Man erhält die Eigenwerte durch Lösung des Gleichungssystems

$$\mathbf{A}\,\mathbf{p}_i = \lambda_i\,\mathbf{p}_i$$

Durch Umformung:

$$\mathbf{A}\,\mathbf{p}_i - \mathbf{E}\,\lambda_i\,\mathbf{p}_i = (\mathbf{A} - \mathbf{E}\,\lambda_i)\mathbf{p}_i = 0$$

Ausser der Triviallösung $p_i = 0$ ist diese Gleichung für $|\mathbf{A} - \mathbf{E}\,\boldsymbol{\lambda}_i| = 0$ erfüllt. Der Wert der Determinante ist für eine n×n-Matrix **A** durch ein Polynom n-ten Grades gegeben und hat im allgemeinen Fall n Lösungen. Es ist jedoch möglich, dass Lösungen komplex sind, dass Lösungen mehrfach vorkommen oder dass eine oder mehrere Lösungen gleich null sind.

Durch Substitution der Lösungen λ_i in der Gleichung erhält man die entsprechenden Eigenvektoren $\mathbf{p}_i$. Diese sind aber nur bis auf einen multiplikativen Faktor definiert, d. h. wenn $\mathbf{p}_i$ ein Eigenvektor ist, so ist auch k $\mathbf{p}_i$ ein Eigenvektor. üblicherweise werden die Eigenvektoren auf einheitliche Länge normiert.

Aus dem Zusammenhang

$$\mathbf{A} = \mathbf{P}\,\boldsymbol{\Lambda}\,\mathbf{P}^{-1}$$

folgen:

$$\mathbf{A}\,\mathbf{P} = \mathbf{P}\,\boldsymbol{\Lambda}$$

und

$$\boldsymbol{\Lambda} = \mathbf{P}^{-1}\,\mathbf{A}\,\mathbf{P}$$

Die Eigenwerte und Eigenvekoren *symmetrischer* Matrizen mit realen Matrixelementen sind real. Zudem sind die Eigenvektoren orthogonal, wenn die Eigenwerte paarweise verschieden sind. Die Matrix der Eigenvektoren wird dann eine orthogonale Matrix. Die Inversen können in den oberen Gleichungen durch Transponierte ersetzt werden:

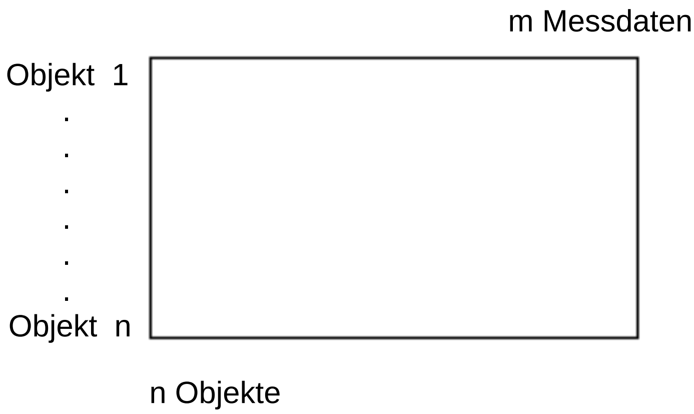
$$\mathbf{A} = \mathbf{P}\,\boldsymbol{\Lambda}\,\mathbf{P}^T \text{ und } \boldsymbol{\Lambda} = \mathbf{P}^T\,\mathbf{A}\,\mathbf{P}$$

Die Eigenwerte von Matrizen, die als ein Produkt $\mathbf{A}^T\mathbf{A}$ angegeben werden können, sind nichtnegativ. Die Anzahl der von null verschiedenen Eigenwerte gibt in diesem Fall den Rang der Matrix **A** (und von $\mathbf{A}^T\mathbf{A}$) an.

Die Gleichung $\mathbf{A}\,\mathbf{p}_i = \lambda_i\,\mathbf{p}_i$ kann *geometrisch* als Streckung verstanden werden, die die Matrix **A** auf ihre Eigenvektoren ausübt.

Einfache Statistik mit Matrixoperation

Irgendeine Serie von Zahlenwerten (z.B. eine Reihe von Messwerten) kann als ein n-dimensionaler Vektor angeschaut werden. Mehrdimensionale Messwerte entsprechen einer Matrix (Datenmatrix). Sie werden üblicherweise folgendermassen angeordnet:



Die m Messdaten können z.B. m Wellenlängen eines Diodenarraydetektors sein oder auch m verschiedene Messgrössen (verschiedene Analysenmethoden) die an jedem der n Objekten ermittelt worden sind.

Eindimensionale Daten (Datenvektor $\mathbf{y}$)

Mittelwert $$\bar{y} = \frac{1}{n}\sum y_i = \frac{1}{n}\mathbf{1}^T\mathbf{y}$$

Summe der Fehlerquadrate:

$$\sum (y_i-\bar{y})^2 = \sum y_i^2 - \frac{(\sum y_i)^2}{n} = \mathbf{y}^T\mathbf{y} - \frac{1}{n}(\mathbf{y}^T\mathbf{1})(\mathbf{1}^T\mathbf{y}) = \mathbf{y}^T\mathbf{H}_n\mathbf{y} \text{ mit } \mathbf{H}_n=\mathbf{I} - (\frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T)$$

Varianz:

$$s^2 = \frac{1}{n-1}\mathbf{y}^T\mathbf{H}_n\mathbf{y}$$

Für mehrdimensionale Daten erhält man mit der analogen Matrixoperationen die Varianz-Kovarianzmatrix **S:**

$$\mathbf{S} = \frac{1}{n-1}\mathbf{A}^T\mathbf{A} - \frac{1}{n}(\mathbf{A}^T\mathbf{1})(\mathbf{1}^T\mathbf{A}) = \mathbf{A}^T\mathbf{H}_n\mathbf{A}$$

Die Diagonalelemente sind die Varianzen, die Ausserdiagonalelemente die Kovarianzen der entsprechenden Variablen.

$$\mathbf{H}_n\mathbf{A} = \begin{bmatrix} x_1-\bar{x} & y_1-\bar{y} & z_1-\bar{z} \\ x_2-\bar{x} & y_2-\bar{y} & z_2-\bar{z} \\ x_3-\bar{x} & y_3-\bar{y} & z_3-\bar{z} \end{bmatrix}$$

$$\mathbf{S} = \frac{1}{n-1}\mathbf{A}^T\mathbf{H}_n\mathbf{A} = \frac{1}{n-1}\begin{bmatrix} \Sigma(x_i-\bar{x})^2 & \Sigma(x_i-\bar{x})(y_i-\bar{y}) & \Sigma(x_i-\bar{x})(z_i-\bar{z}) \\ \Sigma(x_i-\bar{x})(y_i-\bar{y}) & \Sigma(y_i-\bar{y})^2 & \Sigma(y_i-\bar{y})(z_i-\bar{z}) \\ \Sigma(x_i-\bar{x})(z_i-\bar{z}) & \Sigma(y_i-\bar{y})(z_i-\bar{z}) & \Sigma(z_i-\bar{z})^2 \end{bmatrix} =$$

$$\begin{bmatrix} \mathrm{var}(x) & \mathrm{cov}(x,y) & \mathrm{cov}(x,z) \\ \mathrm{cov}(x,y) & \mathrm{var}(y) & \mathrm{cov}(y,z) \\ \mathrm{cov}(x,z) & \mathrm{cov}(y,z) & \mathrm{var}(z) \end{bmatrix}$$

Die Bedeutung von Varianz und Kovarianz ist im Folgenden kurz rekapituliert. Mittelwert μ und Varianz σ^2 sind Eigenschaften der Verteilung einer Zufallsvariablen ξ und als solche theoretische Werte. (**E**[.] ist der Erwartungswert einer Grösse.)

$\mu = \mathbf{E}[\xi]$

$\sigma^2 = \mathbf{E}[(\xi-\mu)^2]$

Auf Grund von N Beobachtungen dieser Zufallsvariablen $x_1, x_2, ..., x_N$ lassen sich Mittelwert und Varianz abschätzen

$$\mu \approx m = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\sigma^2 \approx s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - m)^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - m\sum_{i=1}^{n} x_i\right)$$

Kovarianz

Die gemeinsame Verteilung zweier zusammengehöriger Zufallsvariablen (x,h) sind durch eine zweidimensionale Verteilung charakterisiert. Ein charakteristischer Parameter dieser zweidimensionalen Verteilung ist neben den Mittelwerten (μ_ξ und μ_η) und Varianzen (σ_ξ^2 und σ_η^2) die Kovarianz $\sigma_{\xi\eta}$. Die Kovarianz ist ein Mass für den Zusammenhang zwischen ξ und η.

$$\sigma_{\xi\eta} = \mathbf{E}[(\xi-\mu_\xi)(\eta-\mu_\eta)]$$

Die Abschätzung aufgrund von N Beobachtungspaaren $(x_1\ y_1),(x_2\ y_2), ..., (x_N\ y_N)$ ist:

$$\sigma_{\xi\eta} \approx s_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - m_x)(y_i - m_y) = \frac{1}{N-1}\left(\sum_{i=1}^{N} x_i y_i - m_y\sum_{i=1}^{N} x_i\right)$$

Für die Beschreibung der Wahrscheinlichkeitsdichtefunktion einer mehrdimensionalen Normalverteilung wird die Varianz-Kovarianzmatrix anstelle der Varianz bei der eindimensionalen Normalverteilung eingesetzt.

Eindimensionale Normalverteilung:

$$f(x) = (2\pi\sigma^2)^{-1/2} \exp\{-\frac{1}{2} (\frac{x-\mu}{\sigma})^2\} = (2\pi\sigma^2)^{-1/2} \exp\{-\frac{1}{2} (x-\mu) (\sigma^2)^{-1}(x-\mu)\}$$

Vieldimensionale Normalverteilung:

$$f(x) = (2\pi\mathbf{\Sigma})^{-1/2} \exp\{-\frac{1}{2} (\mathbf{x}-\boldsymbol{\mu})^T \mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\}$$

$\mathbf{\Sigma}$ ist dabei die Varianz-Kovarianz-Matrix (die Schätzung dafür ist **S**). Produkte der Form: $\mathbf{a}^T\mathbf{X}\mathbf{a}$ nennt man quadratische Formen. Im allgemeinen entsprechen sie einer (Hyper)Elliplse. Ist die Varianz-Kovarianz Matrix diagonal (keine Kovarianz) dann sind die Hauptachsen der Ellipse parallel der Koordinatenachsen. Sind die beiden Diagonalelemente zudem gleich gross, so erhält man einen Kreis (Kugel, Hyperkugel) statt einer (Hyper)Ellipse.

Aus der Varianz-Kovarianzmatrix kann man die Korrelationsmatrix wie folgt ableiten. Zuerst erzeugt man eine Diagonalmatrix **D**, die die jeweiligen reziproken Standardabweichungen enthält (vgl. Anhang 2). Die Korrelationsmatrix ist dann:

$$\mathbf{R} = \mathbf{D}\ \mathbf{S}\ \mathbf{D}$$

Die Diagonalelemente der Korrelationsmatrix haben den Wert 1, die Ausserdiagonalelmente sind die Korrelationskoeffizienten der entsprechenden Paare von Variablen.

Korrelationskoeffizient

Ein normiertes Mass für den Zusammenhang ist der Korrelationskoeffizient ρ mit Werten zwischen -1 und +1.

$$\rho = \frac{\sigma\xi\eta}{\sigma\xi\sigma\eta}$$

Für eine Abschätzung des Korrelationskoeffizienten gilt:

$$\rho \approx r = \frac{s_{xy}}{s_x s_y}$$

2.3. GEOMETRISCHE INTERPRETATIONEN

2.3.1. Geometrische Interpretationen von Vektoren

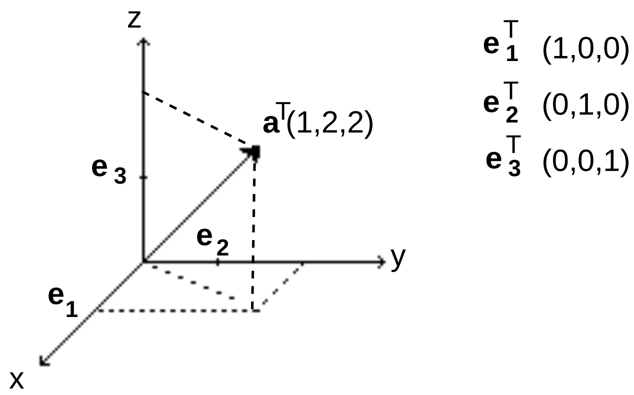
Vektor

Ein n-dimensionaler Vektor: ($\mathbf{a}^T = [a_1, a_2, ... , a_n]$) entspricht einem Punkt in einem n-dimensionalen Raum (oder kann als ein Vektor der vom Ursprung des Koordinatensystems zu diesem Punkt zeigt, aufgefasst werden). Die Beträge der einzelnen Komponenten ($|a_i|$) sind die Längen, die man nach einer senkrechten Projektion auf die entsprechenden Koordinaten erhält. Das Vorzeichen zeigt an, ob die Projektion auf die positive oder auf die negative Achse fällt.

Eine k×n-Matrix $\mathbf{A}_{kn}$ kann als eine Sammlung von k n-dimensionalen Zeilenvektoren oder von n k-dimensionalen Kolonnenvektoren vorgestellt werden. Sie entsprechen k Punkten in einem n-dimensionalen Raum oder n Punkten in einem k-dimensionalen Raum. Entsprechend dieser Interpretation kann eine (eindimensionale) Reihe von n Messungen als ein Punkt (Vektor) in einem n-dimensionalen Raum vorgestellt werden. k Messreihen der Länge n ergeben eine k×n-Matrix.

Basisvektoren

Richtung und Skalierung der Koordinatenachsen können mit Hilfe der Basisvektoren angegeben werden. In einem karthesischen Koordinatensystem haben sie eine einheitliche Länge und zeigen in Richtung der einzelnen Koordinatenachsen. Dementsprechend hat der i-te Basisvektor eine 1 an der i-ten Stelle und sonst Nullen.



Alle Vektoren der entsprechenden Dimension können als Linearkombination der Basisvektoren angegeben werden:

$$\mathbf{a} = \sum_{i=1}^{n} a_i e_i = 1\, e_1 + 2\, e_2 + 2\, e_3 = 1\begin{bmatrix}1\\0\\0\end{bmatrix} + 2\begin{bmatrix}0\\1\\0\end{bmatrix} + 2\begin{bmatrix}0\\0\\1\end{bmatrix} = \begin{bmatrix}1\\2\\2\end{bmatrix}$$

Dies gilt auch dann, wenn die Basisvektoren nicht orthogonal (senkrecht zueinander) und nicht normiert (Einheitslänge) sind.

Der Euklidische Raum

Ein n-dimensionaler Euklidischer Raum ist die Sammlung aller n-dimensionalen Vektoren, für die die folgenden Operationen erlaubt sind: Addition, Substraktion und Multiplikation mit einem Skalar.

Für jedes Paar von Vektoren ist eine nichtnegative Zahl, die Euklidische Distanz zwischen den Vektoren definiert (vgl. Gesetz von Pythagoras):

$$\|\mathbf{a}-\mathbf{b}\| = \sqrt{(a_1-b_1)^2+(a_2-b_2)^2+\ldots+(a_n-b_n)^2} = \sqrt{(\mathbf{a}-\mathbf{b})^T(\mathbf{a}-\mathbf{b})} = \sqrt{\mathbf{a}^T\mathbf{a}+\mathbf{b}^T\mathbf{b}-2\mathbf{a}^T\mathbf{b}}$$

Die Länge eines Vektors

Die Länge $\|a\|$ eines n-dimensionalen Vektors $\mathbf{a}^T=(a_1,a_2,...,a_n)$ ist sein Abstand vom Koordinatenursprung:

$$\|a\| = \sqrt{a_1^2 + a_2^2 + ... + a_n^2} = \sqrt{\mathbf{a}^T\mathbf{a}}$$

$\mathbf{a}^T\mathbf{a}$ ist das "innere Produkt" des Vektors mit sich selbst.

Richtungscosinus

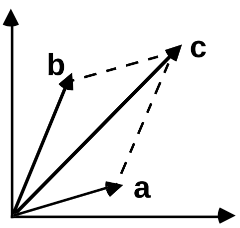
Ein Vektor ist normalisiert (auf die Länge 1), wenn jede seiner Komponenten durch seine Länge dividiert wird. Die so erhaltenen Komponenten entsprechen cos α_i, α_i ist der Winkel des Vektors relativ zur i-ten Achse:

$$\cos\alpha_i = \frac{a_i}{\|a\|}$$

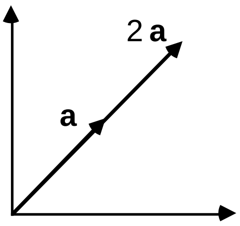
Da $\sum a_i^2 = \|a\|^2$ ist, wird $\sum \cos^2\alpha_i = 1$. Wenn cos $\alpha_i = 0$ ist, liegt der Vektor orthogonal (senkrecht) zur entsprechenden Achse.

Geometrische Interpretation der Vektoraddition und der Multiplikation mit einem Skalaren

Geometrisch ist das Resultat einer Vektoraddition **c** die Diagonale des Parallelograms, welches durch die beiden Summanden **a** und **b** definiert ist. Im Fall der Addition mehrerer Vektoren ist das Resultat die Diagonale eines Parallelepipeds (n-dimensionales Parallelogramm):



Das Resultat einer Multiplikation mit einem Skalar ist ein Vektor der in die gleiche Richtung zeigt (ist kolinear):



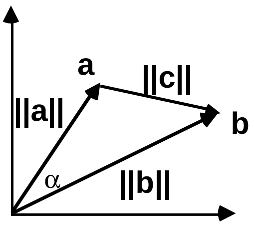
$$k\mathbf{a} = \begin{bmatrix} ka_1 \\ ka_2 \\ ka_3 \end{bmatrix}$$

Linearkombination

Die Kombination der beiden Operationen (Addition und Multiplikation mit einem Skalar) ist eine Linearkombination. Durch Linearkombination kann von n linear unabhängigen (vgl. unten) n-dimensionalen Vektoren jeder andere n-dimensionale Vektor erzeugt werden.

Winkel zwischen zwei Vektoren

Der Winkel zwischen zwei Vektoren folgt aus der Trigonometrie (Kosinussatz):



$$c^2 = a^2 + b^2 - 2ab\cos\alpha$$

$$\|\mathbf{c}\|^2 = \|\mathbf{a} - \mathbf{b}\|^2 = \|\mathbf{a}\|^2 + \|\mathbf{b}\|^2 - 2\|\mathbf{a}\|\|\mathbf{b}\|\cos\alpha_{ab}$$

$$\text{oder: } \cos\alpha_{ab} = \frac{\|\mathbf{a}\|^2 + \|\mathbf{b}\|^2 - \|\mathbf{a} - \mathbf{b}\|^2}{2\|\mathbf{a}\|\|\mathbf{b}\|}$$

Für orthogonale Vektoren ($\alpha=90^o$) ist $\|\mathbf{a} - \mathbf{b}\|^2 = \|\mathbf{a}\|^2 + \|\mathbf{b}\|^2$

Geometrische Bedeutung des Skalarprodukts

Das Skalarprodukt zweier Vektoren ist das Produkt ihrer Längen mal der Kosinus des von ihnen eingeschlossenen Winkels:

$$\mathbf{a}^T\mathbf{b} = \|\mathbf{a}\|\|\mathbf{b}\|\cos\alpha_{ab} \qquad \cos\alpha_{ab} = \frac{\mathbf{a}^T\mathbf{b}}{\|\mathbf{a}\|\|\mathbf{b}\|}$$

In einem orthogonalen Koordinatensystem (orthogonaler Basis) ist diese Definition gleichbedeutend mit der oben eingeführten Definition des inneren Produkts ($\mathbf{a}^T\mathbf{b} = \sum a_i b_i$), da nach dem Kosinussatz:

$$\|\mathbf{a}\|\|\mathbf{b}\|\cos\alpha_{ab} = \frac{1}{2}\left(\|\mathbf{a}\|^2 + \|\mathbf{b}\|^2 - \|\mathbf{a} - \mathbf{b}\|^2\right) = \sum a_i b_i$$

Lineare Abhängigkeit

Der Satz der Vektoren $\mathbf{a_i}$ ist linear abhängig, wenn die Gleichung

$$k_1a_1 + k_2a_2 + ... + k_pa_p = 0$$

eine nichttriviale Lösung aufweist (d.h. eine andere Lösung als $k_1=k_2= ... =k_p=0$). Dies bedeutet, dass mindestens einer der Vektoren als Linearkombination der anderen ausgedrückt werden kann, z.B.

$$a_k = -\frac{k_1a_1+k_2a_2+...+k_{k-1}a_{k-1}+k_{k+1}a_{k+1}+...+k_pa_p}{k_k}$$

d.h. er "bringt" keine neue Dimension. Irgendein Satz von p n-dimensionalen Vektoren ist linear abhängig, wenn p>n ist. Die Dimension eines Raumes entspricht der maximalen Anzahl von linear unabhängigen Vektoren des Raumes.

Ein Satz von p linear unabhängigen Vektoren "spannt" einen p-dimensionalen Euklidischen Raum auf, d.h. alle Vektoren dieses Raumes können als Linearkombination dieser p Vektoren beschrieben werden. Dieser Satz ist dann eine Basis des n-dimensionalen Raumes. Eine orthonormale Basis besteht aus orthogonalen Vektoren (Skalarprodukt =0) mit der Einheitslänge.

2.3.2. Geometrische Interpretation von Matrixoperationen und von Determinanten

Geometrische Interpretation der Matrixmultiplikation

Ein n-dimensionaler Vektor definiert einen Punkt in einem n-dimensionalen Raum. Eine $k\times n$-dimensionale Matrix entspricht dann einer Punkteschar von k Punkten im Raum. Durch Multiplikation des n-dimensionalen Vektors mit einer $n\times n$ Matrix erhält man einen anderen n-dimensionalen Vektor, d.h. einen anderen Punkt. Sämtliche Multiplikationen eines Vektors mit einer Matrix kann man sich also geometrisch vorstellen als eine Transformation eines Punktes in einen anderen Punkt.

Eine alternative Vorstellung wäre, dass man durch die Matrixmultiplikation den Punkt an Ort und Stelle belässt, aber die Koordinaten des Koordinatensystems so transformiert, dass der gleiche Punkt nun im neuen Koordinatensystem andere Koordinaten hat. Wir werden zuerst die erste Vorstellung anhand einer Reihe von Beispielen erläutern und dann die Koordinatentransformation behandeln.

Schliesslich wird sich die Frage stellen, wie eine Transformation in einem Koordinatensystem (in einer Basis) in einem anderen Koordinatensystem (Basis) ausgedrückt werden kann. Diese Vorstellungen werden eine anschauliche Sicht verschiedener multivariater Techniken erlauben. Die Beispiele sind der Einfachheit halber alle zweidimensional, die Aussagen gelten aber für beliebig viele Dimensionen. Umkehrbare Transformationen entsprechen Multiplikationen mit invertierbaren Matrizen $\mathbf{T}$ (durch Multiplikation mit der Inversen der Matrix, wird die Transformation rückgängig gemacht, da $\mathbf{T}^{-1}\mathbf{TM}=\mathbf{IM}=\mathbf{M}$).

Neben reversiblen Transformationen gibt es auch solche, die nicht umkehrbar sind, weil durch die Transformation Information verlorengeht. In diesen Fällen ist der Rang der Transformationsmatrix niedriger als die Dimension des ursprünglichen Vektors; der Punkt wird in einen Raum von niedgrigerer Dimension projiziert. Für eine senkrechte Projektion von Punkten auf eine Gerade ist beispielsweise die umgekehrte Transformation nicht eindeutig, da unendlich viele Punkte auf den gleichen Punkt der Gerade projiziert werden. Der Rang einer Matrix entspricht der Anzahl der linear unabhängigen Vektoren, d.h. der Dimension des durch die Matrix definierten Unterraumes, und dadurch der durch die Multiplikation mit der Matrix beibehaltenen Dimensionen.

Geometrische Interpretation der Determinanten

Geometrisch können Determinanten wie folgt vorgestellt werden: Die n Kolonnenvektoren einer quadratischen n x n Matrix spannen ein Parallelepiped im n-dimensionalen Raum auf (Parallelogramm im zweidimensionalen Raum). Der Beitrag der Determinanten ist das n-dimensionale Volumen des Parallelepipeds. Das Volumen wird dann gleich null, wenn das Parallelepiped ein weniger als n-dimensionaler Körper (d.h. "plattgedrückt") ist.

Die Wirkung der Multiplikation mit einer Matrix auf das (n-dimensionale) Volumen eines Körpers hängt direkt mit der Determinanten der Matrix zusammen. Wird durch eine Reihe von Vektoren (die in eine Matrix zusammengefasst werden können) ein Objekt definiert, und dieses Objekt durch Multiplikation mit einer Matrix (Transformationsmatrix) transformiert, so ist das Verhältnis der n-dimensionalen Volumina des transformierten Objektes und des ursprünglichen Objektes gleich der Determinanten der Transformationsmatrix.

Nicht umkehrbare Operationen sind solche, die die Dimensionalität des n-dimensionalen Körpers reduzieren (den Körper "plattdrücken"). Die entsprechende $n \times n$ Transformationsmatrix ist singulär, ihre Determinante ist gleich null und ihr Rang, der besagt, auf wieviele Dimensionen die Operation den ursprünglichen Körper reduziert, ist niedriger als die Dimension der Matrix.

Orthogonale Transformationen

Orthogonale Matrizen sind solche, deren Transponierte der Inversen entspricht ($\mathbf{C}\mathbf{C}^T=\mathbf{C}^T\mathbf{C}=\mathbf{I}$). Dies bedeutet, dass das Produkt der Multiplikation eines Kolonnenvektors mit einem Zeilenvektor immer gleich null ist, ausser bei der Multiplikation mit sich selbst, was 1 ergibt (d.h. die Länge des Vektors ist gleich 1, da $\|\mathbf{a}\| = \sqrt{\mathbf{a}^T\mathbf{a}}$).

Die Multiplikation mit einer orthogonalen Matrix ist eine orthogonale Transformation. Da die Determinante einer orthogonalen Matrix entweder +1 oder -1 ist, werden durch orthogonale Transformationen die Körper nicht verzerrt und die n-dimensionalen Volumina (Fläche, wenn 2 Dimensionen) bleiben unverändert.

Die orthogonale Transformation entspricht einer Rotation (wenn die Determinante der orthogonalen Matrix =1 ist) oder einer Rotation plus Inversion (engl. improper rotation) wenn die Determinante = -1 ist. Die einzelnen Elemente der orthogonalen Matrix entsprechen Winkelfunktionen.

Im zweidimensionalen Fall, kann eine Rotation mit einem einzigen Winkel beschrieben werden. Die Rotationsmatrix ist dann für eine Rotation um den Winkel α im Gegenuhrzeigersinn:

$$\begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}$$

Im zweidimensionalen Fall ist leicht einzusehen, dass die Rotation eines Punktes um α Grad einer Rotation des Koordinatensystems um $-\alpha$ Grad entspricht.

Für den mehrdimensionalen Fall gilt für die *Koordinatentransformation* :

$$\begin{bmatrix} \cos\theta_{11} & \cos\theta_{21} & \dots & \cos\theta_{n1} \\ \cos\theta_{12} & \cos\theta_{22} & \dots & \cos\theta_{n2} \\ \cos\theta_{13} & \cos\theta_{23} & \dots & \cos\theta_{nn} \end{bmatrix}$$

θ_{ij} ist dabei der Winkel zwischen der i-ten Koordinate im alten Koordinatensystem und der j-ten Koordinate im neuen Koordinatensystem. (Auch der zweidimensionale Fall entspricht genau dieser Formel, da $\cos(\alpha+90^o)=-\sin\alpha$ und $\cos(\alpha-90^o)=\sin\alpha$).

Geometrische Interpretation der Multiplikation mit speziellen Transformationsmatrizen

Man kann zeigen, dass die Multiplikation mit einer beliebigen Matrix **T** auf die Multiplikation mit drei einfachen Matrizen zurückgeführt werden kann. Es gilt immer, dass $\mathbf{T}=\mathbf{P\Delta Q}^T$, wobei **P** und **Q** Rotationsmatrizen sind und $\mathbf{\Delta}$ eine diagonale Matrix (Streckungsmatrix) ist.

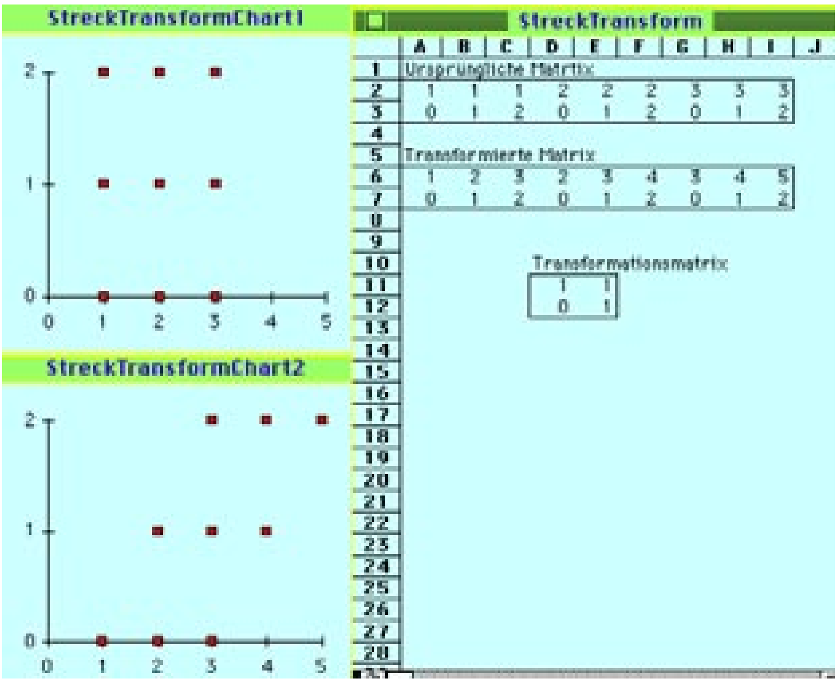
Hier werden zunächst Transformationsmatrizen für einfache geometrische Operationen behandelt. Um die geometrische Wirkung nachvollziehen zu können, ist es sinnvoll, die Transformation auf eine Punkteschar (Matrix **X**, 9 Punkte im zweidimensionalen Raum) wirken zu lassen:

$$\mathbf{X}=\begin{bmatrix} 1 & 1 & 1 & 2 & 2 & 2 & 3 & 3 & 3 \\ 0 & 1 & 2 & 0 & 1 & 2 & 0 & 1 & 2 \end{bmatrix}$$

Bemerkung: Man könnte mit der Transponierten der Matrix **X** arbeiten, und dann müsste man jeweils mit der Transponierten der Transformationsmatrix postmultiplizieren, da $\mathbf{A}=\mathbf{TX}=(\mathbf{X}^T\mathbf{T}^T)^T$.

Übung

Um die geometrische Wirkung der Matrixmultiplikation zu erfahren, ist es dringend empfohlen die folgenden Übungen mit dem Tabellenkalkulationsprogramm Excel selbständig auszuführen. Man soll dabei die oben angegebene Matrix in einem "Chart" abbilden, und die transformierten Punkte daneben in einem anderen. Dies kann z.B. folgendermassen aussehen:



1. Reflexion: **T**= $\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$

2. Permutation der Achsen: **T**= $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

3. Zentrale Dehnung: **T**= $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$

4. Rotation um 45^{o}: **T**= $\begin{bmatrix} 0.707 & 0.707 \\ -0.707 & 0.707 \end{bmatrix}$

(Andere Werte können für cosα und sinα eingesetzt werden)

5. "Shear" **T**= $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$; $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a+c & b+d \\ c & d \end{bmatrix}$

6. Translation

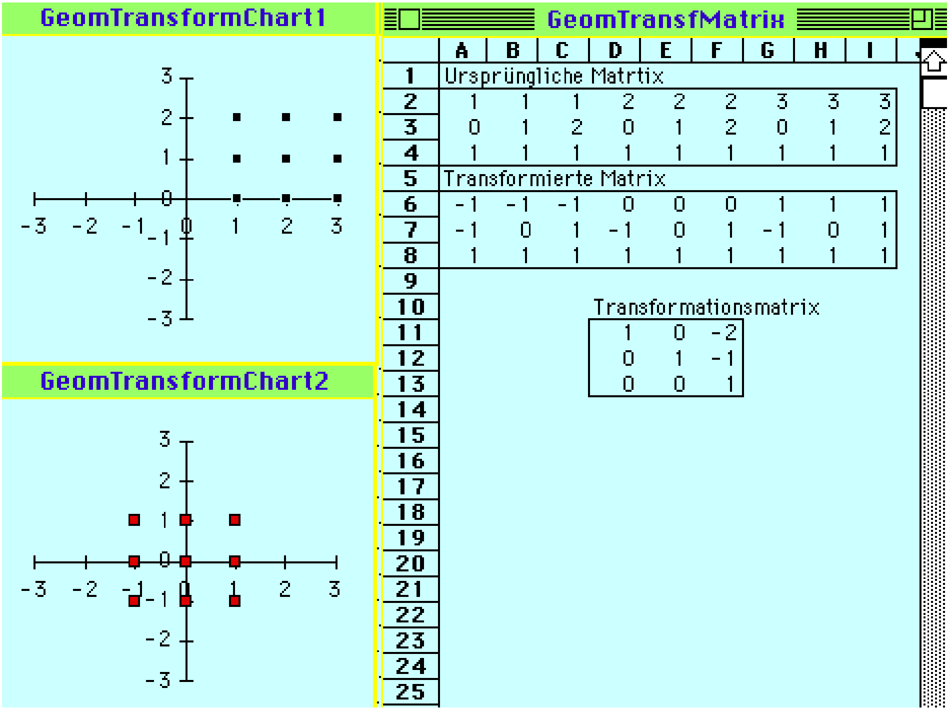
Alle Transformationen können mit Matrixmultiplikationen vom Typ **X***=**TX** durchgeführt werden ausser der Translation (**X***=**T**+**X**). Mit einer Erweiterung der ursprünglichen Matrix **X** mit einer Zeile ($\mathbf{1}^T$) und der Transformationsmatrix mit einer Zeile und einer Kolonne kann aber auch die Translation mit einer Matrixmultiplikation durchgeführt werden:

Die Transformation: $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ --> $\begin{bmatrix} a+h & b+h \\ c+k & d+k \end{bmatrix}$ kann mit der folgenden Matrixmultiplikation

erzielt werden:

$$\begin{bmatrix} 1 & 0 & h \\ 0 & 1 & k \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} a+h & b+h \\ c+k & d+k \\ 1 & 1 \end{bmatrix}$$

Wenn man die ursprüngliche Matrix zentrieren will, dann wird man die Mittelwerte der einzelnen Zeilen von den Werten der ursprünglichen Matrix abziehen (vgl. Zentrierungsmatrix der Übung).



7. Kombinierte Transformationen.

Wird eine Matrix X nacheinander mit der Transformationsmatrizen **T** und **Q** transformiert ($\mathbf{X}^*=\mathbf{QTX}$), ist das Resultat gleich, wie wenn man die Operation $\mathbf{X}^*=\mathbf{SX}$ durchführt, wobei $\mathbf{S}=\mathbf{QT}$ (die Matrixmultiplikation ist assoziativ). Im allgemeinen wird das Resultat der Transformationen in umgekehrter Reihenfolge ($\mathbf{X}^{**}=\mathbf{TQX}$) verschieden vom ursprünglichen Resultat ($\mathbf{X}^*\neq\mathbf{X}^{**}$ da im allgemeinen $\mathbf{QT}\neq\mathbf{TQ}$). Man soll im Rahmen der Übung diese Fälle ausprobieren.

Man kann zeigen, dass jede beliebige Transformation auf eine Kombination von drei Operationen: Rotation (oder Rotation und Inversion) + Strecken + Rotation zurückgeführt werden kann.

Transformation durch Änderung der Basis

Matrixmultiplikationen wurden bisher geometrisch so interpretiert, dass die m Koordinaten eines Punktes (definiert in einem m-dimensionalen Vektor) oder n Punkten (definiert in einer m×n-Matrix) in neue Koordinaten transformiert werden. Im Falle von reversiblen Transformationen gibt es eine andere geometrische Interpretation. Danach bleibt der Punkt (bzw. die Punkte) an Ort und Stelle, aber das Koordinatensystem wird verändert. Im Falle einer Rotation ist es anschaulich einsehbar, dass eine Rotation eines Punktes um $+\alpha^{o}$ der Rotation des Koordinatensystems um $-\alpha^{o}$ entspricht. Im Falle nicht orthogonaler Transformationen ist das geometrische Bild etwas komplizierter, da nichtorthogonale Transformationen auch die Winkel zwischen den Koordinatenachsen ändern.

Eine Transformationsmatrix **T** erzeugt aus der ursprünglichen orthogonalen Basis **E** eine neue Basis **F**, die durch $\mathbf{F}=\mathbf{E}\mathbf{T}^{T}$ gegeben ist. Die Koordinaten eines Punktes x sind in der ursprünglichen Basis durch den Vektor **x** gegeben. Die Koordinaten $\mathbf{x}^{\mathbf{f}}$ in der neuen Basis **F** können wegen $\mathbf{F}\mathbf{x}^{\mathbf{f}}=\mathbf{E}\mathbf{x}$ durch

$$\mathbf{x}^{f} = \mathbf{F}^{-1}\mathbf{x} = (\mathbf{T}^{T})^{-1}\mathbf{x}$$

berechnet werden. (Eine nicht reversible Transformation eines Punktes hat keine äquivalente Koordinatentransformation, die Transformationsmatrix ist nicht invertierbar).

Übungsbeispiel

Transformation des Punktes $\mathbf{x}^{T}$ [1,2] mit der Transformationsmatrix

$$\mathbf{T}=\begin{bmatrix} 0.9 & 0.44 \\ 0.6 & 0.8 \end{bmatrix}$$

1. Berechnung der transformierten Koordinaten **x*** in der ursprünglichen Basis

2. Berechnung der Koordinaten $\mathbf{x}^{\mathbf{f}}$ in der neuen Basis.

Zusammenfassung:

1. Transformation eines Punktes in der ursprünglichen Basis:
   Transformation $\mathbf{x}^{*}=\mathbf{T}\mathbf{x}$, Rücktransformation: $\mathbf{x}=\mathbf{T}^{-1}\mathbf{x}^{*}$
   Für orthogonale Transformationsmatrizen **A**:
   Transformation: $\mathbf{x}^{*}=\mathbf{A}\mathbf{x}$, Rücktransformation: $\mathbf{x}=\mathbf{A}^{T}\mathbf{x}^{*}$

2. Transformation durch Änderung der Basisvektoren.
   Die neue Basis **F** ist durch $\mathbf{F}=\mathbf{E}\mathbf{T}^{T}$ gegeben.
   Transformation: $\mathbf{x}^{\mathbf{f}}=(\mathbf{T}^{T})^{-1}\mathbf{x}$, Rücktransformation: $\mathbf{x}=\mathbf{T}^{T}\mathbf{x}^{\mathbf{f}}$
   Für orthogonale Transformationsmatrizen **A**:

Transformation: $\mathbf{x^f}=\mathbf{Ax}$ (für eine orthogonale Matrix ist $(\mathbf{A}^T)^{-1} = \mathbf{A}$),
Rücktransformation: $\mathbf{x}=\mathbf{A}^T\mathbf{x^f}$

Transformationsmatrix für die ursprüngliche Basis

In der Praxis trifft man immer wieder Matrixoperation der Form $\mathbf{U}^{-1}\mathbf{AU}$ an. Diese können geometrisch folgendermassen interpretiert werden. Man stelle sich eine Transformationsmatrix **A** in der ursprünglichen Basis vor. Eine zweite Matrix **L** soll eine Basistransformation bewirken ($\mathbf{F}=\mathbf{EL}^T$). Dann wird die Transformationsmatrix **A** in der neuen Basis **F** ($\mathbf{A^f}$) folgendermassen gegeben:

$$\mathbf{A}^f= (\mathbf{L}^T)^{-1}\mathbf{AL}^T.$$

Die Operation kann schrittweise veranschaulicht werden: $\mathbf{L}^T$ transformiert einen Punkt von der **F**-Basis in die ursprüngliche Basis, **A** führt die Transformation in der ursprünglichen Basis aus und schliesslich transformiert $(\mathbf{L}^T)^{-1}$ von der ursprünglichen Basis in die **F**-Basis zurück. Wenn **L** symmetrisch ist ($\mathbf{L}=\mathbf{L}^T$), vereinfacht sich die obere Gleichung zu:

$$\mathbf{A}^f= \mathbf{L}^{-1}\mathbf{AL}$$

Irreversible Transformationen

Die Multiplikation mit einer Matrix, deren Rang niedriger ist als die Dimension des Vektors, entspricht einer nicht umkehrbaren Transformation (mit Informationsverlust verbunden), die entsprechende Matrix ist singulär. Von besonderer Bedeutung sind Projektionsmatrizen. Projektionsmatrizen sind idempotent ($\mathbf{M}=\mathbf{MM}$).

Eine Matrix **D**, die durch die folgende Operation aus zwei Matrizen erzeugt wird, ist immer idempotent:

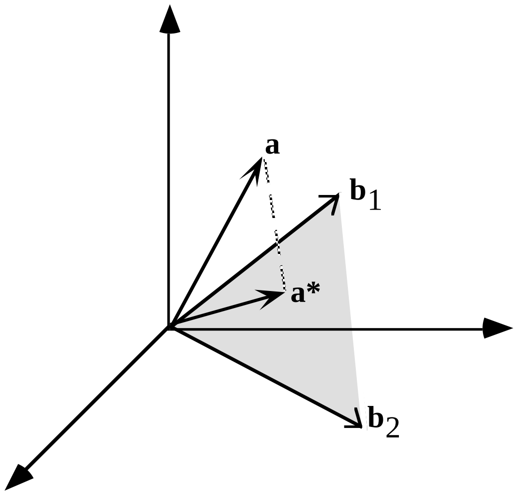
$$\mathbf{D} = \mathbf{B}(\mathbf{A}^T\mathbf{B})^{-1}\mathbf{A}^T$$

da $\mathbf{DD}=\mathbf{B}(\mathbf{A}^T\mathbf{B})^{-1}\mathbf{A}^T\mathbf{B}(\mathbf{A}^T\mathbf{B})^{-1}\mathbf{A}^T$ und $\mathbf{A}^T\mathbf{B}(\mathbf{A}^T\mathbf{B})^{-1}=\mathbf{E}$.

Orthogonale Projektion

Ein Spezialfall ist die orthogonale Projektion bei der die Projektionsmatrix symmetrisch ist.

Beispiel: Projektion eines Vektors **a** (im 3D-Raum) auf eine Ebene, die durch zwei Vektoren $\mathbf{b_1}$ und $\mathbf{b_2}$ definiert ist (**a** wird zu **a***)



1. Orthoganilitätsbedingung

$(\mathbf{a^*}-\mathbf{a})^T\mathbf{b_1} = 0$; $(\mathbf{a^*}-\mathbf{a})^T\mathbf{b_2} = 0$
d.h. $\mathbf{a^*}^T\mathbf{b_1} = \mathbf{a}^T\,\mathbf{b_1}$ und $\mathbf{a^*}^T\mathbf{b_2} = \mathbf{a}^T\,\mathbf{b_2}$ oder: $\mathbf{a^*}^T\mathbf{B} = \mathbf{a}^T\,\mathbf{B}$

2. **a*** ist in der Ebene, die durch **B** definiert ist. Man kann deshalb **a*** als Linearkombination von zwei unabhängigen Vektoren, die diese diese Ebene definieren, ausdrücken::

$$\mathbf{a^*} = \sum p_i b_i = p_1\mathbf{b_1} + p_2\mathbf{b_2} = \mathbf{Bp}$$

daraus $\mathbf{a^*}^T = \mathbf{p}^T\mathbf{B}^T$

Oben eingesetzt: $\mathbf{p}^T\mathbf{B}^T\mathbf{B} = \mathbf{a}^T\,\mathbf{B}$

Wenn $\mathbf{B}^T\mathbf{B}$ invertierbar ist (d.h. die Matrix **B** hat den vollen Rang, da die Vektoren $\mathbf{b_i}$ linear unabhängige Basisvektoren sind) gilt: $\mathbf{p}^T = \mathbf{a}^T\,\mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}$

und damit: $\mathbf{a^*} = \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{a}$

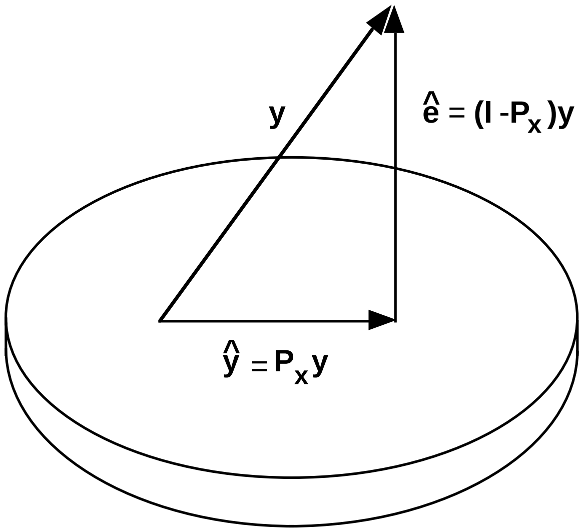
Beispiel mit $\mathbf{B} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$

$$\mathbf{B}^T\mathbf{B} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

$$\left(\mathbf{B}^T\mathbf{B}\right)^{-1} = \frac{1}{3}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

$$\mathbf{B}\left(\mathbf{B}^T\mathbf{B}\right)^{-1}\mathbf{B}^T = \frac{1}{3}\begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2/3 & 1/3 & 1/3 \\ 1/3 & 2/3 & -1/3 \\ 1/3 & -1/3 & 2/3 \end{bmatrix}$$

Die orthogonale Projektion ist bei der Regressionsanalyse von besonderer Bedeutung (vgl. auch die Notation in der folgenden Abbildung):



Die Quadratsummen entsprechen den quadratierten Längen der Vektoren (vgl oben). Im Falle der projizierten Vektoren vereinfachen sich die Resultate, weil die Projektionsmatrix symmetrisch und idempotent ist:

$$\sum \hat{y}_i^2 = (\mathbf{P}_x\mathbf{y})^T(\mathbf{P}_x\mathbf{y}) = \mathbf{y}^T\mathbf{P}_x{}^T\mathbf{P}_x\mathbf{y} = \mathbf{y}^T\mathbf{P}_x\mathbf{P}_x\,\mathbf{y} = \mathbf{y}^T\mathbf{P}_x\mathbf{y} = \mathbf{y}^T\hat{\mathbf{y}}$$

Es ist auch klar, dass die beiden Vektoren, die durch orthogonale Projektionen entstanden sind, als Summenvektor den ursprünglichen Vektor ergeben:

$$\mathbf{y} = \mathbf{P}_x\mathbf{y} + (\mathbf{I}\text{-}\mathbf{P}_x)\mathbf{y}$$
$$\mathbf{y} = \hat{\mathbf{y}} + \hat{\mathbf{e}}$$

Der Rang der n×n Projektionsmatrix $\mathbf{P_x}$=$\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ ist gleich der Spur dieser Matrix und somit gleich der Spur der p×p Matrix $\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}$ (Spur(**AB**) = Spur(**BA**)) und ist damit gleich $\mathbf{I_p}$=p, wenn die Matrix **X** die Dimension n×p hat. Dieser Zusammenhang ist deshalb nützlich, weil die Freiheitsgrade, die mit den durch $\mathbf{y}^T\mathbf{P_x}\mathbf{y}$ erhaltenen Quadratsummen verbundenen sind, durch den Rang der Projektionsmatrix gegeben sind.

Übung

Berechnen Sie die Projektionsmatrix $\mathbf{P_x}$ für die Projektion auf eine Ebene, die durch die Vektoren $\mathbf{x_1}$=$(0,1,1)^T$ und $\mathbf{x_2}$=$(1,1,0)^T$ definiert ist. Berechnen Sie die Spur von $\mathbf{P_x}$. Projizieren Sie den Vektor $(1,1,1)^T$auf diese Ebene und berechnen Sie den Abstand dieses Vektors von der Ebene. Skizzieren Sie die geometrischen Zusammenhänge.