Lineare Abbildungen, Matrizen & Determinante

Irene Winkler & Nico Görnitz

Arbeitsgruppe Maschinelles Lernen

Bsp: Lineare Regression

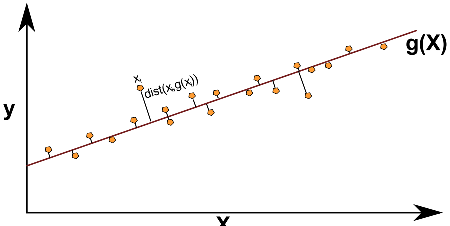
- Gegeben sind n Datenpunkte $X = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ und dazugehörige Labels $\mathbf{y} = (y_1, y_2, \ldots, y_n)$
- Lineares Model: $y = f_{\mathbf{w}}(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle$
- Einstellen des Parameters $\mathbf{w}$ dr. Methode der kleinsten Quadrate:

$$\min_{\mathbf{w}} \frac{1}{2} \sum_i \text{dist}(y_i, \langle \mathbf{w}, \mathbf{x}_i \rangle) = \min_{\mathbf{x}} \frac{1}{2} \sum_i ||y_i - \mathbf{w}^T \mathbf{x}_i||_2^2 = \frac{1}{2}(\mathbf{y}^T \mathbf{y} - 2\mathbf{w} X \mathbf{y} + \mathbf{w}^T X X^T \mathbf{w})$$

- Auflösen nach $\mathbf{w}$ (Ableitung auf 0 setzen=Minima suchen):

$$0 = \frac{\partial}{\partial \mathbf{w}} \{ \frac{1}{2}(\mathbf{y}^T \mathbf{y} - 2\mathbf{w}^T X \mathbf{y} + \mathbf{w}^T X X^T \mathbf{w}) \} = -X\mathbf{y} + X X^T \mathbf{w}$$

$$X\mathbf{y} = X X^T \mathbf{w} \quad \Rightarrow \quad (X X^T)^{-1} X \mathbf{y} = \mathbf{w}$$



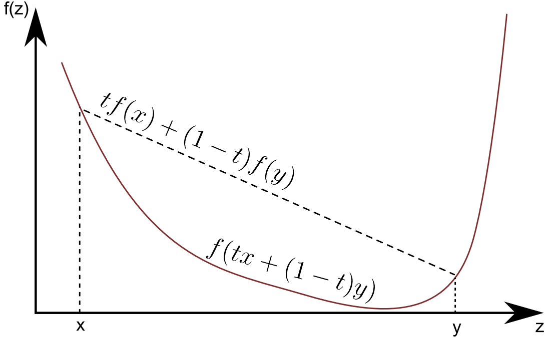
Allgemeine Abbildungen

Eine Funktion $f : X \to Y, \quad x \mapsto f(x)$ heisst

- **injektiv**
  - wenn fuer alle $x, y \in X$ mit $f(x) = f(y)$ folgt, dass $x = y$ ist.
  - Zu jedem x gibt es genau 1 $f(x)$.
- **surjektiv**
  - wenn $f(X) = Y$ ist.
  - Also wenn alle Elemente in Y von X aus "angesprochen" werden koennen.
- **bijektiv***
  - wenn f injektiv und surjektiv ist.
  - D.h. alle Elemente von $x \in X$ zeigen auf genau 1 $y \in Y$ und es gibt keine "unbenutzten" Elemente in Y.

Konvexe Funktionen

Eine reellwertige Funktion $f : C \to \mathbb{R}$, die auf einer konvexen Teilmenge C eines reellen Vektorraums definiert ist, heißt *konvex*, wenn für alle $\mathbf{x}, \mathbf{y}$ aus C und für alle $0 \leq t \leq 1$ gilt, dass $f(t\mathbf{x} + (1-t)\mathbf{y}) \leq tf(\mathbf{x}) + (1-t)f(\mathbf{y})$.



Lineare Abbildungen

Definition. Seien V und W reelle Vektorräume. Eine Abbildung $f : V \mapsto W$ heisst *linear* wenn

1. $f(\mathbf{v} + \mathbf{w}) = f(\mathbf{v}) + f(\mathbf{w})$ und
2. $f(\lambda \mathbf{v}) = \lambda f(\mathbf{v})$

für alle $\mathbf{v}, \mathbf{w} \in V$ und $\lambda \in \mathbb{R}$

Arten von Linearen Abbildungen

Eine Funktion $f : V \to W, \quad x \mapsto f(x)$ heisst

- **Homomorphismus**
  - wenn es eine lineare Abbildung zwischen zwei Vektorräumen V und W ist.
- **Isomorphismus**
  - Wenn die lineare Abbildung *bijektiv* ist.
  - D.h. wenn es eine Umkehrabbildung geben kann.
- **Endomorphismus**
  - wenn $W = V$ ist.

Basis & Lineare Abbildung

Seien V und W reelle Vektorräume, $\mathbf{v}_1, \ldots, \mathbf{v}_n$ eine Basis von V und $f : V \mapsto W$ eine linear Abbildung.

Linearität $\Rightarrow$ für jedes $\mathbf{v} \in V$ gilt:

$$f(\mathbf{v}) = f(\lambda_1 \mathbf{v}_1 + \ldots + \lambda_n \mathbf{v}_n) = \lambda_1 f(\mathbf{v}_1) + \ldots + \lambda_n f(\mathbf{v}_n)$$

$\Rightarrow$ Es genügt die Bilder der Basisvektoren zu kennen, um die ganze Funktion zu charakterisieren!

$\Rightarrow$ repräsentierbar als Matrix, in der die Koordinaten von $\mathbf{v}$ auf Koordinaten von $f(\mathbf{v})$ abgebildet werden.

Bild, Faser & Kern

Ist Funktion $F : V \to W$ eine lineare Abbildung, dann ist

- $\mathrm{Im}F := F(V)$ das Bild von F
- $F^{-1}(\mathbf{w}) := \{\mathbf{v} \in V | F(\mathbf{v}) = \mathbf{w}\}$ die Faser über W
- $\mathrm{Ker}F := F^{-1}(0)$ der Kern von F

Existenz & Eindeutigkeit

Satz. Seien V und W reele Vektorräume und $(v_1, \dots, v_n)$ eine Basis von V. Dann gibt es zu *jedem* n-Tupel $(w_1, \dots, w_n)$ von Vektoren in W genau eine lineare Abbildung $f : V \mapsto W$ mit $f(v_i) = w_i, i = 1, \dots n$.

Beweis.

1. Existenz: Fuer $x = \lambda_1 v_1 + \dots + \lambda_n v_n$ definiere $f(x) := \lambda_1 w_1 + \dots \lambda_n w_n$
2. Eindeutigkeit: Seien $f, f' : V \mapsto W$ lineare Abbildungen mit $f(v_i) = f'(v_i) = w_i, i = 1, \dots n$. Dann $f(x) = f(\lambda_1 v_1 + \dots + \lambda_n v_n) = \lambda_1 w_1 + \dots \lambda_n w_n = f'(\lambda_1 v_1 + \dots + \lambda_n v_n) = f'(x)$

Matrizen

Sei $g : \mathbb{R}^n \mapsto \mathbb{R}^m$ eine lineare Abbildung. Wir wollen sie charakterisieren durch $g(\mathbf{x}) = A \cdot \mathbf{x}$ für $\mathbf{x} \in \mathbb{R}^n$ und A eine $m \times n$ Matrix.

$$g(\mathbf{x}) = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + \ldots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \ldots + a_{mn}x_n \end{pmatrix}$$

Wie muss A aussehen?

Es genügt, die Bilder der Basisvektoren $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ zu kennen. Die Spalten der Matrix A sind die Bilder der Einheitsvektoren!

Beispiele

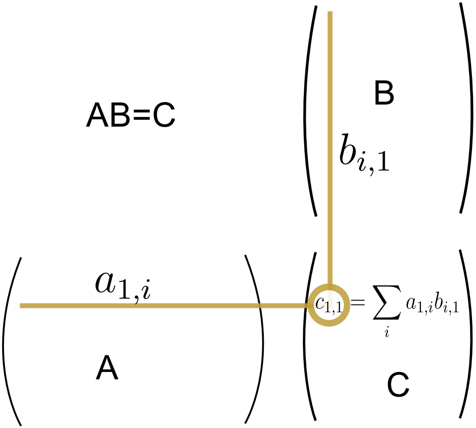
Lineare Abbildungen lassen sich also als Lineares Gleichungssystem $\mathbf{y} = g(\mathbf{x}) = A\mathbf{x}$ darstellen:

- Identitätsmatrix I
- $g : \mathbb{R}^2 \mapsto \mathbb{R}^2$ Drehung um $90°$
- $g : \mathbb{R}^2 \mapsto \mathbb{R}^2$ Spiegelung an der x-Achse
- $g : \mathbb{R}^2 \mapsto \mathbb{R}^2$ Projektion auf die Gerade durch $(0, 0)$ und $(1, 1)$

Ps: Praktische Lösung von $A\mathbf{x} = \mathbf{y}$ (wenn $\mathbf{x}$ gesucht ist) durch z.B. QR-Zerlegung, Cholesky-Zerlegung, Conjugate Gradient oder dem allseits bekannten Gauss-Verfahren.

Matrixrechnung

- Addition
- Skalarmultiplikation
- Multiplikation



Weiteres

- Spur einer quadratischen Matrix = Summe der Hauptdiagonalelemente.
  Für $A \in \mathbb{R}^{n \times m}$ und $B \in \mathbb{R}^{m \times n}$ gilt

$$\mathrm{Spur}(AB) = \mathrm{Spur}(BA)$$

  .

- Sind $A^T, B^T \in \mathbb{R}^{n \times m}$ die transponierten Matrizen zu $A, B \in \mathbb{R}^{m \times n}$ und $\lambda \in \mathbb{R}$ dann

$$(A + B)^T = A^T + B^T$$

$$(\lambda \cdot A)^T = \lambda \cdot A^T,$$

$$(A \cdot B)^T = B^T \cdot A^T$$

Komposition & Transformation

Seien $F : V \mapsto W$ und $G : U \mapsto V$ lineare Abbildungen, dann ist die Komposition (Hintereinanderausfuehrung) $F \circ G$ wieder linear. Dies entspricht der Multiplikation der zu F und G gehoerenden Matrix.

Sei $F : V \mapsto W$ eine lineare Abbildung, und seien $\mathcal{A}$ und $\mathcal{B}$ Basen von V und W. Dann hat man ein Diagramm linearer Abbildungen und es gilt:

$$\begin{tikzcd}
V \arrow[r, "F"] & W \\
K^n \arrow[u, "T_A"] \arrow[r, "M"] & K^m \arrow[u, "T_B"']
\end{tikzcd}$$

$$M = T_B^{-1} \circ F \circ T_A$$

Rang

Definition: Der Rang einer Matrix ist die Maximalzahl linear unabhängiger Spalten (äquivalent: Zeilen).

Der Rang ist die Dimensionalität des Bildes von der durch A beschriebenen Abbildung.

Inverse Matrix

Definition: Eine quadratische Matrix $A \in \mathbb{R}^{n\times n}$ ist *invertierbar (regulär, nichtsingulär)*, wenn eine weitere Matrix A^{-1} existiert, sodass

$$AA^{-1} = A^{-1}A = I_n.$$

Es gilt:

- A ist genau dann invertierbar, wenn $\operatorname{rang}(A) = n$
- $(A \cdot B)^{-1} = B^{-1} \cdot A^{-1}$
- $(A^T)^{-1} = (A^{-1})^T$
- $\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$

Inverse Matrix Berechnung

Will man die Inverse von $A \in \mathbb{R}^{n \times n}$ ausrechenen, dann kann man folgendermassen vorgehen:

- Schreiben $(I_n|A)$ nebeneinander
- Fuehre synchron in beiden Matrizen die ueblichen Umformungen aus, bis auf der rechten Seite I_n erscheint
- Dann steht die Inverse von A, A^{-1} auf der linken Seite

Spezielle Matrizen

Sei $A \in \mathbb{R}^{n \times n}$ eine quadratische, reelle Matrix.

- A ist *orthogonal* wenn $AA^T = A^T A = I_n$.
  Orthogonale Matrizen stellen Spiegelungen und Drehungen im Raum dar.
- A ist *symmetrisch* wenn $A = A^T$
- A ist *antisymmetrisch* wenn $A = -A^T$
- A ist *diagonal* wenn alle Elemente außerhalb der Hauptdiagonale 0 sind

Determinante

Die Determinante einer quadratischen Matrix ist eine Zahl. Was bedeutet sie anschaulich?

- Absolutbetrag: das Volumen des Parallelotops das durch die Zeilen- oder Spaltenvektoren aufgespannt wird
- Vorzeichen: Orientierung des Parallelotops

Determinante

Satz und Definition: Es gibt genau eine Abbildung
det : $\mathbb{R}^{n\times n} \ni A \mapsto \det A \in \mathbb{R}$ mit

1. $\det$ ist linear in jeder Zeile
2. Ist der (Zeilen-) rang kleiner als n, so ist $\det A = 0$
3. $\det I_n = 1$

Diese Abbildung heisst *Determinante*.

Determinante - Regeln

Seien $A, B \in \mathbb{R}^{n \times n}$ zwei quadratische Matrizen und $\lambda \in \mathbb{R}$

- Ist A eine Dreiecksmatrix, dann ist die Determinante das Produkt der Hauptdiagonalelemente von A
- A ist genau dann invertierbar, wenn $\det A \neq 0$
- $\det AB = \det A \det B$
- $\det A^{-1} = (\det A)^{-1}$
- $\det A = \det A^T$
- $\det \lambda A = \lambda^n \det A$
- $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$