3. KI Übung: Einfache Lineare Regression

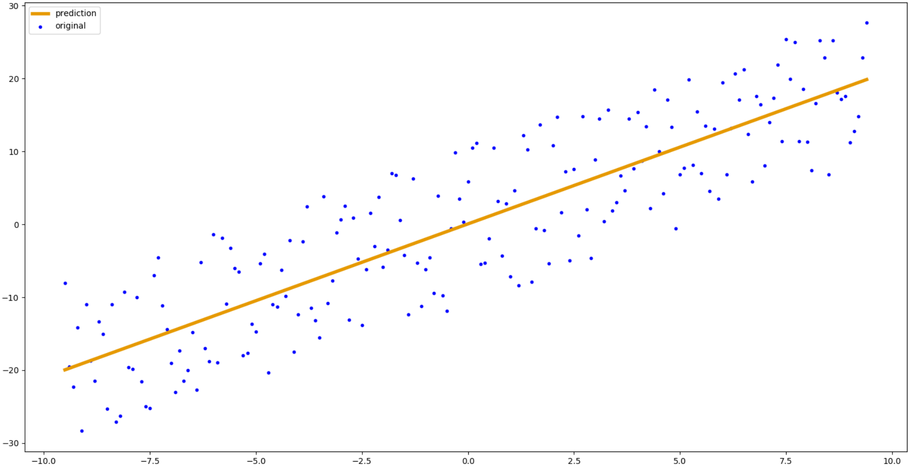
Matthias Tschöpe, Kunal Oberoi

11. November 2019

Namensherkunft: Einfache lineare Regression

einfach da wir eine affin-lineare Approximation suchen

lineare da unser Model eine Lineare Abbildung ist.



Idee:

Gegeben:

Ein **Input-Datensatz** $\mathcal{X} := \{x^{(1)}, x^{(2)}, \dots, x^{(n)}\}$, mit $x^{(i)} \in \mathbb{R}$, von Datenpunkten (oft auch **Samples** genannt) und ein **Label-Datensatz** $\mathcal{Y} := \{y^{(1)}, y^{(2)}, \dots, y^{(n)}\}$, wobei $y^{(i)} \in \mathbb{R}$ das **Label** (manchmal auch Ground Truth genannt) von $x^{(i)}$ ist.

Gesucht:

Eine affin-lineare Funktion $f : \mathbb{R} \to \mathbb{R}$ so, dass für alle $x^{(i)} \in \mathcal{X}$ gilt:

$$f(x^{(i)}) := \widehat{y}^{(i)} = ax^{(i)} + b \approx y^{(i)} \tag{1}$$

und der "*Fehler*" zwischen $\widehat{y}^{(i)}$ und $y^{(i)}$ minimal ist. Das heißt, wir möchten den Faktor a und die Konstante b approximieren. Als nächstes bringen wir die Gleichung (1) in Matrix-Vektor-Form.

Normalengleichung

Dazu definieren wir zunächst:

$$X := \begin{bmatrix} 1 & x^{(1)} \\ 1 & x^{(2)} \\ \vdots & \vdots \\ 1 & x^{(n)} \end{bmatrix} \qquad w := \begin{bmatrix} b \\ a \end{bmatrix} \qquad y := \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(n)} \end{bmatrix} \tag{2}$$

Wenn $y^{(i)} = \widehat{y}^{(i)}$ (d.h. unsere Daten sind nicht verrauscht) und w ist optimal, dann gilt:

$$Xw = y \tag{3}$$

Doch wie lösen wir das Problem? Existiert die Inverse von X immer?

Normalengleichung

X ist nicht notwendigerweise quadratisch. Das ist aber eine Voraussetzung dafür, dass für X eine Inverse existiert. Also müssen wir einen anderen Weg finden die Gleichung (3) nach w aufzulösen. Dazu multiplizieren wir X^T von links und erhalten damit:

$$\underbrace{X^T X}_{\in \mathbb{R}^{2\times 2}} w = X^T y \tag{4}$$

Das LGS in (4) nennt man auch **Normalengleichungen**. Offensichtlich ist $X^T X$ eine quadratische Matrix, aber ist sie auch invertierbar? Dafür untersuchen wir die Definitheit der Matrix $X^T X$. Zur Erinnerung: Eine Matrix $A := X^T X \in \mathbb{R}^{k\times k}$ heißt:

$$\begin{aligned}
&\textit{positiv definit} && \text{wenn für alle } u \in \mathbb{R}^k : u^T A u > 0, \text{mit } u \neq \mathbf{0}_k \\
&\textit{positiv semidefinit} && \text{wenn für alle } u \in \mathbb{R}^k : u^T A u \geq 0, \text{mit } u \neq \mathbf{0}_k \\
&\textit{negativ semidefinit} && \text{wenn für alle } u \in \mathbb{R}^k : u^T A u \leq 0, \text{mit } u \neq \mathbf{0}_k \\
&\textit{negativ definit} && \text{wenn für alle } u \in \mathbb{R}^k : u^T A u < 0, \text{mit } u \neq \mathbf{0}_k
\end{aligned} \tag{5}$$

Invertierbarkeit von Matrizen

Außerdem wissen wir:

(i) *Wenn A positiv definit, dann ist A invertierbar.*

(ii) *Die Zeilenvektoren von A sind linear unabhängig genau dann, wenn A invertierbar ist.*

(iii) *Die Spaltenvektoren von A sind linear unabhängig genau dann, wenn A invertierbar ist.*

Betrachten wir also nun den Ausdruck:

$$\begin{aligned} u^T A u &= u^T X^T X u \\ &= (Xu)^T (Xu) \\ &= \| Xu \|_2^2 \\ &\geq 0 \end{aligned} \tag{6}$$

Das heißt $\| Xu \|_2^2 = 0$ genau dann, wenn die Zeilen oder Spalten von $X^T X$ linear unabhängig sind.

Lösung der einfachen linearen Regression

<u>1. Fall:</u> $\| Xu \|_2^2 > 0$
Mit den zuvor gemachten Beobachtungen wissen wir, dass jetzt $X^T X$ invertierbar ist und wir können Gleichung (4) wie folgt umformen:

$$\begin{aligned} X^T X w &= X^T y \\ \iff \qquad w &= (X^T X)^{-1} X^T y \end{aligned} \tag{7}$$

<u>2. Fall:</u> $\| Xu \|_2^2 = 0$
Wir führen zunächst eine neue Variable $0 < \lambda \in \mathbb{R}$ ein. Für ein hinreichend großes λ wird:

$$X^T X + \lambda \cdot I_{2\times 2} \tag{8}$$

positiv definit und somit invertierbar. Die Lösung für Gleichung (4) kann dadruch mit:

$$w \approx w_{ridge} = \left(X^T X + \lambda \cdot I_{2\times 2} \right)^{-1} X^T y \tag{9}$$

approximiert werden. Hier steht $I_{k\times k}$ für die $(k \times k)-$Einheitsmatrix.

Aufgaben (i)

1. Aufgabe

In dieser Aufgabe soll der zuvor besprochene Ansatz implementiert werden.

(a) Lesen Sie die Daten der Datei `XY1.csv` ein. Diese hat die Form $x, y_1, y_2, \ldots y_{12}$. Die erste Spalte s_1 gibt ihnen die $x-$Werte. Kombinieren Sie s_1 mit genau einer weiteren Spalte s_i mit $i \geq 2$. Verwenden Sie die so konstruierte Funktion für die verbleibenden Aufgaben.

(b) Schreiben Sie eine Klasse Optimierer. Im Konstruktor soll festgelegt werden welcher Optimierer verwendet werden soll und eine Methode `fit(x_values, y_values, max_grad)` soll den entsprechenden Optimierer später starten.

(c) Bringen Sie ihre Daten in die Matrix-Vektor-Darstellung (siehe Folie 4).

(d) Schreiben Sie eine Funktion `vectorized_SLR(X,y)` die, die Normalengleichung wie zuvor gezeigt löst.

(e) Plotten Sie ihre Ergebnisse.

2. Aufgabe

In dieser Aufgabe soll der zuvor besprochene Ansatz ohne die Matrix-Vektordarstellung implementiert werden. Man kann zeigen, dass a und b auch wie folgt berechnet werden kann:

$$a = \frac{\sum_{i=1}^{n} (x^{(i)} - \bar{x})(y^{(i)} - \bar{y})}{\sum_{i=1}^{n} (x^{(i)} - \bar{x})^2} \tag{10}$$

$$b = \bar{y} - a\bar{x}$$

Hierbei bezeichnet $\bar{x}$ den Mittelwert des Input-Datensatzes und $\bar{y}$ den Mittelwert des Label-Datensatzes.

(a) Schreiben Sie eine Funktion `iterative_SLR(X,y)` die a und b wie in Gleichung (10) für die Datenpunkte aus Aufgabe 1 berechnet.

(b) Vergleichen Sie Ihre Ergebnisse und die Laufzeit. Das Package `time` liefert eine Funktion `time()`. Damit können Sie sich die aktuelle Zeit ausgeben lassen.

Quellen

1 Hands-On Machine Learning with Scikit-Learn & TensorFlow (Aurelien Geron)

2 Handbook of Medical Statistics (Ji-Qian Fang et al)

3 Deep Learning: Das umfassende Handbuch (Ian Goodfellow et al)