Allgemeine multilineare Regression

Im allgemeinen Fall der multilinearen Regression (Funktion mit m Argumenten) ist ein Datensatz $((x_{11},\dots,x_{m1},y_1),\dots,(x_{n1},\dots,x_{nm},y_n))$ (oder auch dargestellt als $((\vec{x}_1,y_1),\dots,(\vec{x}_n,y_n))$ mit einem Eingabevektor $\vec{x}_i$ und der zugehörigen Ausgabe y_i, $1 \le i \le n$) gegeben, für den eine lineare Ausgleichsfunktion

$$y = f(x_1,\dots,x_m) = a_0 + \sum_{k=1}^{m} a_k x_k$$

gesucht ist. Zur Ableitung der Normalgleichungen stellt man in diesem Fall das zu minimierende Funktional bequemer in Matrixform dar, nämlich als

$$F(\vec{a}) = (\mathbf{X}\vec{a} - \vec{y})^\top (\mathbf{X}\vec{a} - \vec{y}),$$

wobei

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1m} \\ \vdots & & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{nm} \end{pmatrix} \qquad \text{und} \qquad \vec{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

den Datensatz wiedergeben und $\vec{a} = (a_0, a_1, \dots, a_m)^\top$ der Vektor der zu bestimmenden Koeffizienten ist.[1] (Man beachte, daß die Einsen in der ersten Spalte der Matrix $\mathbf{X}$ dem Koeffizienten a_0 zugehören.) Wieder ist eine notwendige Bedingung für ein Minimum, daß die partiellen Ableitungen nach den Koeffizienten a_k, $0 \le k \le m$, verschwinden, was wir mit Hilfe des Differentialoperators ∇ (gesprochen: nabla) schreiben können als

$$\nabla_{\vec{a}}\, F(\vec{a}) = \frac{\mathrm{d}}{\mathrm{d}\vec{a}} F(\vec{a}) = \left(\frac{\partial}{\partial a_0} F(\vec{a}), \frac{\partial}{\partial a_1} F(\vec{a}), \dots, \frac{\partial}{\partial a_m} F(\vec{a}) \right) = \vec{0}.$$

Die Ableitung läßt sich am leichtesten berechnen, wenn man sich klar macht (wie man durch elementweises Ausschreiben leicht nachrechnet), daß sich der Differentialoperator

$$\nabla_{\vec{a}} = \left(\frac{\partial}{\partial a_0}, \frac{\partial}{\partial a_1}, \dots, \frac{\partial}{\partial a_m} \right)$$

formal wie ein Vektor verhält, der von links an die Summe der Fehlerquadrate „heranmultipliziert“ wird. Alternativ kann man die Ableitung komponentenweise ausschreiben. Wir verwenden hier jedoch die erstere, wesentlich

[1] $\top$ bedeutet die Transponierung eines Vektors oder einer Matrix, also die Vertauschung von Zeilen und Spalten.

bequemere Methode und erhalten

$$\begin{aligned}
\vec{0} &= \nabla_{\vec{a}}\, (\mathbf{X}\vec{a}-\vec{y})^\top(\mathbf{X}\vec{a}-\vec{y}) \\
&= \left(\nabla_{\vec{a}}\, (\mathbf{X}\vec{a}-\vec{y})\right)^\top (\mathbf{X}\vec{a}-\vec{y}) + \left((\mathbf{X}\vec{a}-\vec{y})^\top \left(\nabla_{\vec{a}}\, (\mathbf{X}\vec{a}-\vec{y})\right)\right)^\top \\
&= \left(\nabla_{\vec{a}}\, (\mathbf{X}\vec{a}-\vec{y})\right)^\top (\mathbf{X}\vec{a}-\vec{y}) + \left(\nabla_{\vec{a}}\, (\mathbf{X}\vec{a}-\vec{y})\right)^\top (\mathbf{X}\vec{a}-\vec{y}) \\
&= 2\mathbf{X}^\top(\mathbf{X}\vec{a}-\vec{y}) \\
&= 2\mathbf{X}^\top\mathbf{X}\vec{a} - 2\mathbf{X}^\top\vec{y},
\end{aligned}$$

woraus sich unmittelbar das System

$$\mathbf{X}^\top\mathbf{X}\vec{a} = \mathbf{X}^\top\vec{y}$$

der Normalgleichungen ergibt. Dieses System ist offenbar lösbar, wenn $\mathbf{X}^\top\mathbf{X}$ invertierbar ist. Dann gilt

$$\vec{a} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\vec{y}.$$

Den Ausdruck $(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$ nennt man auch die (Moore-Penrose-) **Pseudoinverse** der Matrix $\mathbf{X}$ [Albert 1972]. Mit ihr kann man unmittelbar die Lösung der Regressionsaufgabe angeben.

Es dürfte klar sein, daß sich die Methode der kleinsten Quadrate auch auf Polynome in mehreren Variablen erweitern läßt. Am einfachsten geht man dabei ebenfalls von der oben verwendeten Matrixdarstellung aus, wobei man in die Matrix $\mathbf{X}$ als Eingangsgrößen auch die zu verwendenden Potenzprodukte der unabhängigen Variablen einträgt. Die Ableitung der Normalgleichungen kann dann einfach übernommen werden.