Seminararbeit:

Rousseeuw (1984)
Least Median of Squares Regression

Autor: Miriam Buß

Matrikelnummer: 10549807

Betreuer: Andrea Wiencierz

04.06.2014

In der vorliegenden Seminararbeit soll der von Rousseeuw eingeführte "Least Median of Squares"-Schätzer genauer untersucht werden, vor allem in Hinblick auf seinen Bruchpunkt und seine Robustheit.
Er minimiert nicht wie andere ähnliche Schätzer die Summe der quadrierten Residuen, sondern es wird vielmehr die Summe durch den Median ersetzt. Der dadurch erhaltene Schätzer kann damit einer fast 50%igen Verunreinigung der Daten standhalten. Im Falle der einfachen Regression entspricht dies der Methode, einen möglichst schmalen Streifen zu finden, der die Hälfte der Daten abdeckt.
Außerdem wird gezeigt, dass sich dieser Schätzer sehr gut eignet, um Ausreißer ausfindig zu machen, die ein Modell mit einem anderen Schätzer verzerren würden.
Im Weiteren wird noch auf die Effizienz eingegangen und wie man sie verbessern kann. Eine große Rolle spielt dabei der "Least Trimmed Squares"-Schätzer, der die Datenhälfte mit den kleinsten quadrierten Residuen minimiert.
Und zuletzt werden die möglichen Algorithmen zur Berechnung des LMS-Schätzers untersucht.

Inhaltsverzeichnis

1 Einführung

1.1 Autor

Der Autor Peter Rousseeuw wurde am 13. Oktober 1956 in Antwerpen in Belgien geboren. Um seine Doktorarbeit zu schreiben entschied er sich für den bedeutenden Professor Frank Hampel an der Universität Zürich als Betreuer. Mit diesem arbeitete er an der Einflussfunktion für Tests und der Änderung der Varianzfunktion. Nach dem Abschluss seiner Arbeit 1981 war er im Bereich der angewandten Wahrscheinlichkeit und statistischer Mechanik tätig, bevor er die Forschung für die folgende Arbeit begann. Außerdem arbeitete er an Clusteranalysen die zu L_1-basierten Differenziertheitsmethoden, dem Silhoutte-Plot und neuen unscharfen Clustertechniken führten. Momentan lehrt er am Institut fr Mathematik an der Universität Antwerpen.

1.2 Historischer Hintergrund

Die ersten Theorien robuster Statistik entstanden in den 1960er und 1970er Jahren. Von großem Interesse war dabei, welche Effekte Ausreißer haben können. Um diese zu messen, führte der Wissenschaftler Hampel 1974 den Bruchpunkt ein. Dabei handelt es sich um den höchsten Anteil schlechter Daten an der Gesamtstichprobe, die eine Anwendung bewältigen kann. Rousseeuw suchte nach einem robusten Schätzer, der noch dazu einen Bruchpunkt von 50% haben sollte. Erste Erfolge dabei erzielten Stahel und Donoho, allerdings wurde dieser Schätzer nicht sehr genutzt, genau so wie der von Siegel 1982 vorgeschlagene Schätzer mit 50%igem Bruchpunkt für Regressionen. Die von Rousseeuw 1984 entwickelte Methode war die Erste, die große Erfolge erzielte und deshalb auch die Vorlage fr viele weitere Forscher war. Als Beispiele von Forschungsgebieten in denen der Schätzer benutzt wurde seien die Rechenstatistik, Umweltforschung und Chemometrie genannt. Vor allem im Bereich der Statistik mit positivem Bruchpunkt wurde viel geforscht, um einen gegen Ausreißer resistenten Schätzer zu finden.

1.3 Definitionen

Bruchpunkt:

Als Bruchpunkt wird der größte Anteil von verunreinigten Daten bezeichnet, den eine Anwendung bewältigen kann. Er wurde von Hampel eingeführt und kann beliebig große Werte annehmen, je höher dieser Wert ist desto besser. Im Falle der kleinsten Quadrate ist der Bruchpunkt $\epsilon^* = 0$. Er ist ein sehr wichtiges Kriterium zur Beurteilung ob eine Methode gut oder schlecht ist, allerdings alleine nicht ausreichend als Bedingung.
Rousseeuw und Leroy benutzen den Bruchpunkt als Indikator für Robustheit. Man liest oft, dass der Median robust sei, der Mittelwert aber nicht. Das kommt daher, dass der Median einen Bruchpunkt von 50% hat, was ihn sehr robust gegen Ausreißer macht. Der Mittelwert dagegen einen Bruchpunkt von 0, somit reicht schon ein einziger Ausreißer, um den Schätzer zum Zusammenbruch zu bringen.

Robustheit:

Unter dem Begriff der Robustheit versteht man, dass ein Schätzer auch dann noch zuverlässig arbeitet, wenn er durch bestimmte Faktoren wie z.B. Ausreißer oder eine zu kleine Stichprobe gestört werden könnte.

2 Hauptteil

2.1 Grundlegende Schätzer

Grundlage für die folgenden Schätzer war das klassische lineare Modell

$$y_i = x_{i1}\theta_1 + ... + x_{ip}\theta_p + e_i$$

dessen Fehler e_i als normalverteilt angenommen wird mit Mittelwert 0 und Standardabweichung σ. Das Ziel dieses Modells ist es, $\theta = (\theta_1, ..., \theta_p)^T$ von den Daten $(x_{i1}, ..., x_{ip}; y_i)$ abzuschätzen.

Der bekannteste Schätzer dafür ist die Kleinste-Quadrate-Methode oder auch LS-Schätzer genannt

$$\underset{\hat{\theta}}{minimize} \sum_{i=1}^{n} r_i(\theta)^2$$

Dabei sollen die quadrierten Residuen durch die Hyperebene θ abgeschätzt werden.

Definition (Residuum) *Ein Residuum ist der Abstand zur Hyperebene und entspricht* $r_i = y_i - x_{i1}\theta_1 - ... - x_{ip}\theta_p$.

Allerdings ist der LS-Schätzer durch seinen Mangel an Robustheit nicht sehr zuverlässig, da ein einziger Ausreißer schon einen großen Effekt auf die Schätzung haben kann.
In diesem Zusammenhang wurde von Hampel der Bruchpunkt ϵ^* eingeführt, der in diesem Fall bei 1/n liegt. Wenn also n gegen unendlich läuft gilt ϵ^*=0, was zur Folge hat, dass die Schätzung stark verfälscht werden kann.
Daraufhin entwickelte Edgeworth das $L_1 - Kriterium$

$$\underset{\hat{\theta}}{minimize} \sum_{i=1}^{n} |r_i|$$

welches ein robusterer Schätzer sein sollte. Bei diesem Kriterium wurden die Residuen nicht mehr quadriert, sondern in den Betrag gesetzt. Außderm war es der erste Schritt,

den Median in einen Schätzer mit aufzunehmen, da das $L_1 - Kriterium$ den Median einer eindimensionalen Stichprobe generalisiert und dafür einheitlich gemacht werden muss. Zwar liegt der Bruchpunkt für den Median bei 50%, dennoch erreicht er hier nur den gleichen Wert von ϵ*=0 wie der LS-Schätzer.
Es ist weniger anfällig für Ausreißer, allerdings nicht robust genug gegenüber Ausreißern in den Werten der unabhängigen Variablen.
Statistisch gesehen ist es auch deshalb nicht optimal, da es keine großen abweichenden Werte von x_i bewältigen kann.
Langsam kam also die Frage auf, ob es überhaupt möglich sei einen robusten Schätzer mit hohem Bruchpunkt zu finden.
Abhilfe konnte Siegel 1982 mit seinem "*Repeated Median*" schaffen, welcher der erste Regressionsschätzer mit hohem Bruchpunkt war. Er basiert auf der Berechnung perfekt angepasster Modelle für alle Teilmengen der Größe p mit den Beobachtungen $(x_{i1}, y_{i1}), ..., (x_{ip}, y_{ip})$, die einen eindeutigen Parametervektor festlegen. Die j-te Koordinate dieses Vektors wird mit $\theta_j(i_1, ..., i_p)$ bezeichnet. Außerdem dient der Repeated Median der Gewinnung eines Regressionsmodells durch die koordinatenweise Berechnung des Medians und ist definiert als

$$\hat{\theta} = \underset{i_1}{med}(...(\underset{i_p-1}{med}(\underset{i_p}{med}\, \theta_j(i_1, ..., i_p)))...).$$

Der Bruchpunkt liegt zwar bei $\epsilon* = 50\%$, allerdings ist der Schätzer nicht geeignet für lineare Transformationen von x_i.

Das Ziel war jedoch, einen Schätzer zu finden, der sowohl ein möglichst hohen Bruchpunkt erreicht und noch dazu sehr robust ist, was aber von keinem der zuvor genannten Schätzer erfllt wurde. Bis Rousseeuw schlielich auf die Idee kam, die Summe durch den Median zu ersetzen, was den "*Least Median of Squares*" Schätzer ergab.

$$\underset{\hat{\theta}}{minimize}\ \underset{i}{med}\, r_i(\theta)^2$$

Dieser minimiert den Median der einzelnen Residuen r_i und erreicht einen Bruchpunkt von 50%, der beste zu erwartende Wert. Außerdem kann einer 50%igen Datenverunreinigung standhalten, was zeigt, dass das Kriterium der Robustheit auch erfüllt ist.

Definition (Median) *Der Median einer geordneten Stichproben von n Beobachtungen ist definiert als*

$$med(x_i) = \begin{cases} x_{\frac{n+1}{2}} & \text{für n ungerade} \\ \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}) & \text{für n gerade} \end{cases}$$

Welchen Effekt Ausreißer auf den LS-Schätzer und LMS-Schätzer haben, bzw. wie robust die beiden Schätzer sind, sieht man an der folgenden Grafik.

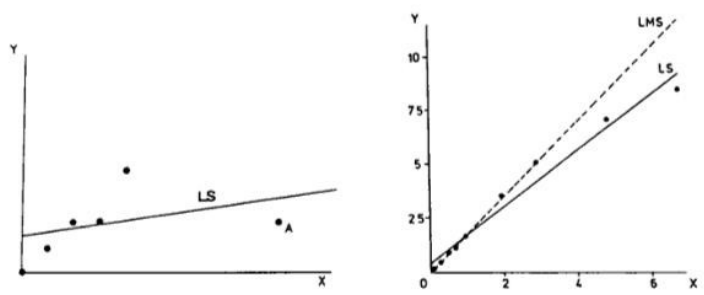


Fig. 1. Effect of outlier on a least-squares line.

Während der LMS-Schätzer von den Ausreißern so gut wie nicht beeinflusst wird, wird der LS-Schätzer jedoch sichtlich verzerrt.

2.2 LMS-Schätzer

2.2.1 Das LMS-Problem

Es seien $x_i^T = (x_{i1}, ..., x_{ip})$ und $y = (y_1, ..., y_n)$ reale Vektoren, sowie der unbekannte Regressionsparameter θ ein p-dimensionaler Spaltenvektor $(\theta_1, ..., \theta_p)^T$.
Außerdem bewegen sich die Vektoren in einem Raum der Größe p+1, in dem der Schätzer eine Zielfunktion besitzt, die auf halben Stichproben basiert, also [n/2]. In diesem (p+1)-dimensionalen Raum gibt es vertikale Hyperebenen, welche p-dimensionale Unterräume sind und (0,...,0) und (0,...,0,1) enthalten. Hyperebenen mit mehr als [n/2] Beobachtungen existieren nicht.
Alle Beobachtungen mit $x_i = 0$ wurden gelöscht, da diese keine Information über θ geben.
Es hieß, ein Schätzer sei dann robust, wenn er eine begrenzte Einflussfunktion hat. Allerdings war diese Definition so nicht mehr länger gültig, da der LMS-Schätzer nur sehr langsam mit $n^{-1/3}$ konvergiert (zum Vergleich: normal konvergieren Schätzer mit $n^{-1/2}$). Nachdem er aber hauptsächlich als diagnostische Methode und Versuchswerkzeug verwendet wird, ist das nicht von großer Bedeutung.

Theorem 1 *Wenn $p > 1$ und die Beobachtungen in allgemeiner Position (bezüglich der x-Koordinate sind), dann ist der Bruchpunkt der LMS Methode gleich* $([n/2] - p + 2)/n$.
Wie beim LS-Schätzer lässt man auch hier n $\to \infty$ laufen, nur dass der Bruchpunkt diesmal nicht bei 0, sondern bei 50% liegt. Einen besseren Wert kann man nicht erreichen.

Die obere Grenze $[(n-p)/2+1]/n$ erhält man, indem man statt des Medians der geordneten quadrierten Residuen die k-te geordnete Statistik der absoluten Residuen $(r^2)_{k:n}$ minimiert. Dabei sei $k = [n/2] + [(p+1)/2]$.

Theorem 2 *Sei T ein Regressionsschätzer, $p = 1$ und alle $x_i = 1$, somit reduziert sich die Stichprobe zu $(y_i)_{i=1,...,n}$.*
Wenn

$$m_T^2 : med\ r_i^2 = med(y_i - T)^2$$

$$min\ med(y_i - \theta)^2$$

gleicht, dann sind sowohl $T - m_T$ als auch $T + m_T$ Beobachtungen in der Stichprobe.
Damit ist es leicht T zu bestimmen, da man nur die kleinste Stichprobenhälfte der geordneten Beobachtungen betrachten muss. T ist dann der Mittelpunkt des kleinsten

Intervalls.

Mit dem Theorem 2 besteht die LMS-Lösung darin, einen möglichst schmalen Streifen zu finden, der die Hälfte der Daten abdeckt. Die Grundidee dabei ist, dass sich in diesem Streifen nur die "guten"Daten befinden, die restlichen werden ignoriert. Um zu entscheiden, was die "guten"Punkte sind, wird jede Teilmenge S eines Datensatzes $\mathcal{P}$ durch eine Funktion f: $\mathbb{P}(\mathcal{P}) \to \mathbb{R}^+$ bewertet und die Teilmenge mit dem besten Wert ausgewählt. $\mathbb{P}(\mathcal{P})$ ist die Menge aller Teilmengen von $\mathcal{P}$.
Eine Definition von Rousseeuw und Leroy für dieses Problem lautet folgendermaßen:

Definition (Statistische Betrachtungsweise) *Gegeben sei ein Datensatz* $\mathcal{P} = \{P_1, ...P_n\}$ *mit n Beobachtungen,* $P_i \in \mathbb{R}^d$ *und einer natürlichen Zahl h mit* $[n/2] \leq h \leq n$. *Finde eine Hyperebene L, so dass* $r_h(L)^2$ *minimiert wird. r bezeichne die Höhe in der der Punkt L liegt.*
Man setze nun $r = |r_h(L)|$ und betrachte die zwei Hyperebenen

$$L^+ : y = a_1x_1 + ... + a_{d-1}x_{d-1} + a_d + r$$

$$L^- : y = a_1x_1 + ... + a_{d-1}x_{d-1} + a_d - r$$

Diese zwei parallelen Hyperebenen bilden nun einen Hyperstreifen, in dem sich alle Residuen kleiner oder gleich r befinden. Die Breite des Streifens wird in vertikaler Richtung gemessen (also entlang der y-Achse) und beträgt hier 2r.

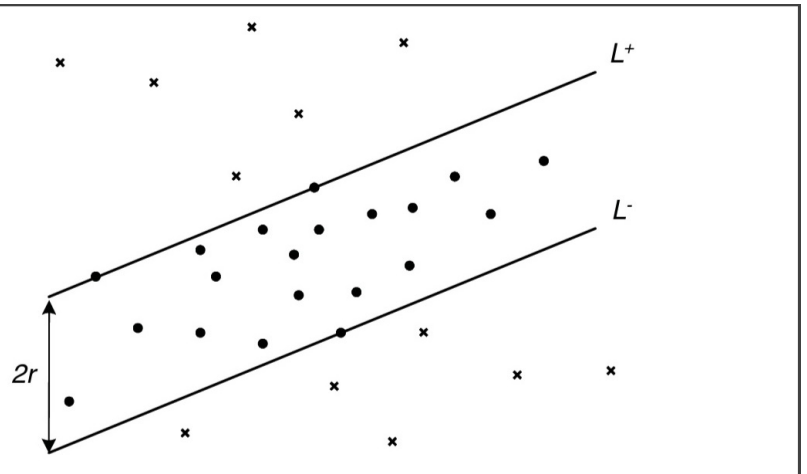


Fig. 2. The computational point of view (Definition 2) is displayed. The hyperstrip defined by L^+ and L^- divides the points in inlying points • and outlying points ✖. The half width of the hyperstrip is r. The solution for LMS is the hyperstrip with the smallest r containing h points

2.2.2 Ein Beispiel

Um in einer großen Stichprobe Ausreißer ausfindig zu machen, benötigt man einen sehr robusten Schätzer wie den LMS. Von Vorteil kann es sein, sowohl den LS als auch den LMS-Schätzer zu benutzen. Wenn beide das gleiche Ergebnis liefern kann man davon ausgehen, dass der LS-Schätzer richtig ist. Wenn es aber große Unterschiede gibt, sieht man an den LMS-Residuen, welche Daten verantwortlich dafür sind, was das nächste Beispiel veranschaulicht.
Die Idee war, zu zeigen, dass der LMS-Schätzer auch für kleinere Stichproben zuverlässig arbeitet. Dafür eignet sich sehr gut ein Datensatz von Draper und Smith (1966, p. 227), der 20 Daten enthält mit 6 zu schätzenden Parametern. Dieser wurde jedoch nicht in seiner ursprünglichen Form verwendet, sondern es wurden ein paar Beobachtungen durch Ausreißer ersetzt.
Die Anwendung des LS-Schätzers liefert

$$y_i = 0.44069x_{i1} - 1.47501x_{i2} - 0.26118x_{i3} + 0.02079x_{i4} + 0.17082x_{i5} + 0.42178$$

Zum Vergleich liefert der LMS-Schätzers folgendes Ergebnis

$$y_i = 0.26870x_{i1} - 0.23806x_{i2} - 0.53572x_{i3} - 0.29373x_{i4} + 0.45096x_{i5} + 0.43474$$

Die folgende Tabelle zeigt sowohl die LS-Residuen, die durch den LS-Skalenparameter $\sigma_{LS} = 0.02412$ geteilt wurden, als auch die durch den entsprechenden Skalenparameter $\sigma_{LMS} = 0.0195$ geteilten LMS-Residuen.
Wie man sieht ist der LS alleine nicht ausreichend um die Ausreißer ausfindig zu machen, da sich die Werte kaum unterscheiden. Betrachtet man dagegen die Werte die der LMS liefert erkennt man sofort die vier Ausreißer.

Table 2. Modified Data on Wood Specific Gravity With Standardized Residuals From Least Squares and Least Median of Squares

								Residual/Scale	
i	x_{i1}	x_{i2}	x_{i3}	x_{i4}	x_{i5}	x_{i6}	y_i	LS	LMS
1	.5730	.1059	.4650	.5380	.8410	1.000	.5340	−.7250	−.0827
2	.6510	.1356	.5270	.5450	.8870	1.000	.5350	.0472	.0013
3	.6060	.1273	.4940	.5210	.9200	1.000	.5700	1.2427	.2836
4	.4370	.1591	.4460	.4230	.9920	1.000	.4500	.3547	−7.6137
5	.5470	.1135	.5310	.5190	.9150	1.000	.5480	1.0024	.9020
6	.4440	.1628	.4290	.4110	.9840	1.000	.4310	−.4518	−9.1023
7	.4890	.1231	.5620	.4550	.8240	1.000	.4810	.9067	.3746
8	.4130	.1673	.4180	.4300	.9780	1.000	.4230	−.0349	−8.9077
9	.5360	.1182	.5920	.4640	.8540	1.000	.4750	−.3959	−.3746
10	.6850	.1564	.6310	.5640	.9140	1.000	.4860	−.4150	−.2071
11	.6640	.1588	.5060	.4810	.8670	1.000	.5540	1.9859	.0013
12	.7030	.1335	.5190	.4840	.8120	1.000	.5190	−1.1977	−.9656
13	.6530	.1395	.6250	.5190	.8920	1.000	.4920	−.4854	.0013
14	.5860	.1114	.5050	.5650	.8890	1.000	.5170	−1.2612	−.6709
15	.5340	.1143	.5210	.5700	.8890	1.000	.5020	−.5866	−.1733
16	.5230	.1320	.5050	.6120	.9190	1.000	.5080	.5237	.0013
17	.5800	.1249	.5460	.6080	.9540	1.000	.5200	−.2548	.0013
18	.4480	.1028	.5220	.5340	.9180	1.000	.5060	.2838	−.1090
19	.4170	.1687	.4050	.4150	.9810	1.000	.4010	−1.0837	−10.7265
20	.5280	.1057	.4240	.5660	.9090	1.000	.5680	.5450	.0013

2.2.3 Effizienz

Wie schon zuvor erwähnt, konvergiert der LMS-Schätzer nur sehr langsam mit $n^{-1/3}$, was eine Effizienz von Null hervorruft. Um diese zu verbessern gibt es mehrere Möglichkeiten.
Eine davon ist, einen gewichteten LS-Schätzer zu erstellen. Dafür wird jeder Beobachtung (x_i, y_i) ein Gewicht w_i zugeordnet, was eine nicht-wachsende Funktion ist. Anschließend werden alle Beobachtungen durch $(w_i^{1/2}x_i, w_i^{1/2}y_i)$ ersetzt, wodurch die Daten mit großen LMS-Residuen entfernt werden.
Betrachtet man nochmals das Beispiel von zuvor, würden mit dieser Methode die vier Ausreißer entfernt werden.

Eine andere effektive Methode ist, den "*Least Trimmed Squares*"-Schätzer (LTS-Schätzer) einzuführen. Dieser ist definiert als

$$minimize \sum_{i=1}^{h} (r^2)_i$$

bei dem die Residuen der Größe nach geordnet sind

$$(r^2)_1 \leq ... \leq (r^2)_i \leq ... \leq (r^2)_n.$$

h ist die Anzahl der Datenpunkte, die nicht aus dem Datensatz "getrimmt"wurden. Der LTS minimiert die 50% der kleinsten quadrierten Residuen, die restlichen Daten können mehr oder weniger beliebige Werte annehmen.
Wenn h=[n/2]+1 gesetzt wird, erreicht man den in Theorem 1 eingeführten Bruchpunkt ([n/2]-p+2)/n, der dem des LMS-Schätzers gleicht und somit bei 50% liegt.

2.2.4 Algorithmen zur Berechnung

Für die Erstellung von Algorithmen ist die O-Notation sehr hilfreich. Diese gibt die obere Schranke für die Laufzeit eines Algorithmus an und ist folgendermaßen definiert:

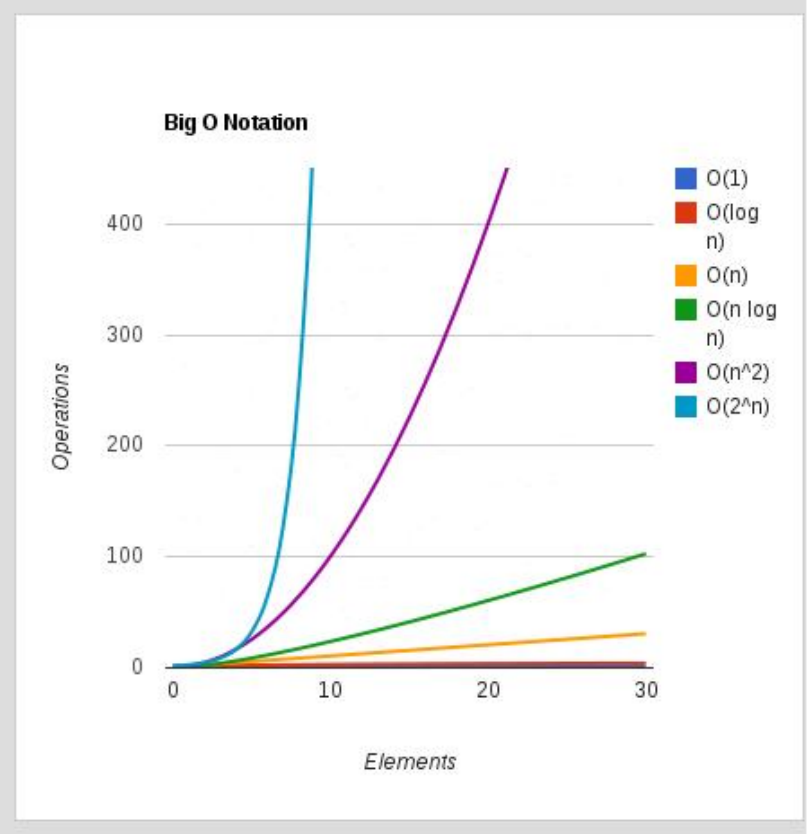
Definition (O-Notation) *Sei f:* $\mathbb{N} \to \mathbb{N}$ *eine Funktion. Die Menge* $O(f)$ *enthält alle Funktionen g, die ab einem gewissen* n_0 *höchstens so schnell wachsen wie f, abgesehen von jeweils einem konstanten Faktor c.*

$$O(f) = \{g : \mathbb{N} \to \mathbb{N} | \exists c > 0 \; \exists n_0 > 0, \; so\ dass \; \forall n \geq n_0 : g(n) \leq c * f(n)\}$$

Man will man mit der O-Notation zeigen, welchen Charakter die Laufzeit eines Algorithmus in Abhängigkeit der Anzahl der Eingabewerte besitzt, ob ihr Wachstumsverhalten z.B. linear, konstant oder exponentiell ist. Die wichtigsten Komplexitätsklassen sind:

$O(1)$	$O(log\ n)$	$O(n)$	$O(n^2)$	$O(2^n)$
konstant	logarithmisch	linear	quadratisch	exponentiell

Folgende Grafik veranschaulicht sehr gut das Verhalten der jeweiligen Funktionen:



Quelle: http://jsdo.it/Boony/big-o-notation

Der bekannteste Algorithmus wurde von Stromberg eingeführt. Dieser berechnet den LMS in einer Zeit von $O(n^{p+2} log\ n)$ und überprüft alle Teilmengen S des Datensatzes durch die eine Hyperebene gebildet wird.
Verbessert wurde dieser Algorithmus von Erickson, der die Rechenzeit zu $O(n^d log\ n)$ veränderte. Außerdem wurde ein willkürlicher Algorithmus mit Rechenzeit $O(n^d)$ für d Dimensionen entwickelt.
Ein lineares Gleichungssystem benötigt zum Beispiel eine Rechenzeit von $O(d^3)$, für den zweidimensionalen Fall wird dagegen ein Algorithmus mit $O(n^2)$ verwendet,welcher von Edelsbrunner und Souvaine eingeführt wurde. Wie letzterer funktioniert wird nun genauer erklärt.

Lemma 1 *Eine optimale Lösung für den LMS ist ein Punkt L' mit den folgenden Eigenschaften:*
1. Es gibt n+h Hyperebenen die unterhalb des Punktes L' liegen oder ihn schneiden
2. Die letzte Koordinate des Punktes ist minimal über alle Punkte mit der 1. Eigenschaft
3. L' ist der Schnittpunkt von mindestens d+1 Hyperebenen

Beweis: Sei der Hyperstreifen (L^+, L^-) eine optimale Lösung für den LMS. Dann existieren innerhalb dieses Streifens h Punkte und L dominiert n+h Hyperebenen.
Falls die Lösung (L^+, L^-) nicht minimal wäre, gäbe es einen schmaleren Hyperstreifen und somit wäre L' der Schnittpunkt von weniger als d+1 Hyperebenen. Ein solcher Punkt kann allerdings als Lösung nicht optimal sein. Daraus folgt, dass (L^+, L^-) eine minimale Lösung für den LMS sein muss.

3 Zusammenfassung

In dieser Arbeit wurden einige Schätzer vorgestellt, von denen aber die meisten nicht sehr zuverlässig und robust arbeiten. Ausreißer zum Beispiel sind Punkte, die eine große Abweichung zu den anderen Datenpunkten aufweisen. Dennoch ist es für einen Großteil der Schätzer nicht möglich, diese ausfindig zu machen, wodurch starke Verzerrungen im Regressionsmodell auftreten können.
Lediglich der von Rousseeuw eingeführte LMS-Schätzer ist in der Lage, diese durch geeignete Verfahren zu finden und zu entfernen. Danach können aber auch konventionelle Schätzer oder der LS berechnet werden, ohne mit Verzerrungen zu rechnen.
Wie man in der ersten Grafik gesehen hat, ist es empfehlenswert, ein Modell von sowohl dem LS als auch dem LMS erstellen, wenn sich die beiden Graphen unterscheiden, kann dem LMS-Schätzer mehr Vertrauen geschenkt werden.
Ein weiterer effektiver Schätzer den wir kennengelernt haben ist der LTS, der allerdings statt dem Median der Residuen wieder die Summe minimiert. Außerdem nimmt er in die Schätzung die 50% der Daten auf, die am kleinsten sind und beachtet nicht wie "gut" oder "schlecht" die Daten sind.
Zwar ist ein Schätzer mit großem Bruchpunkt sehr hilfreich um Ausreißer ausfindig zu machen, allerdings zieht er auch die Nachteile geringer Effizienz und Konvergenzrate mit sich.

Literaturverzeichnis

1 *Bernholt T.*, Computing the Least Median of Squares Estimator in Time $O(n^d)$. *Lecture notes in computer science*, **3480** (2005) 697-706.

2 *Davis P.L.*, Aspects of robust linear regression. *The annals of statistics*, **21** (1993) 1843-1899.

3 *Frosch Moller S., von Frese J., Bro R.*, Robust methods for multivariate data analysis. *Journal of Chemometrics*, **19** (2005) 549-563.

4 *Giloni A., Padberg M.*, Least trimmed squares regression, least median squares regression, and mathematical programming. *Mathematical and Computer Modelling*, **35** (2002) 1043-1060.

5 *Hettmannsperger T., Sheather S.*, A Cautionary Note on the Method of Least Median Squares. *The American Statistician*, **46** (1992) 79-83.

6 *Krivulin N.*, An analysis of the Least Median of Squares regression problem. *http://www.math.spbu.ru/user/krivulin/Publs/SCS19921.pdf* (23.05.2014).

7 *von Löwis F.*, Programmiertechnik II - Analyse von Algorithmen. *http://www.dcl.hpi.uni-potsdam.de/teaching/pt2_07/algorithms.pdf* (27.05.2014).

8 *Rousseeuw P.J.*, Least Median of Suares Regression. *Journal of the American Statistical Association*, **79** (1984) 871-880.

9 *RTC*, Laufzeitkomplexitt von Algorithmen - die O-Notation. *http://www.linux-related.de/index.html?/coding/o-notation.htm* (27.05.2014).

10 *Wainer H.*, Review of Robust Regression & Outlier Detection by Peter J. Rousseeuw; Annick M. Leroy. *Journal of Educational Statistics*, **13** (1988) 358-364.

11 *Wolf C., Best H.*, Handbuch der sozialwissenschaftlichen Datenanalyse. Berlin, Heidelberg: Springer, 2010.

Erklärung zur Urheberschaft

Hiermit versichere ich, dass ich die vorliegende Seminararbeit selbstständig aus den Vorlagen von Andrea Wiencierz angefertigt habe.

München, den 04.06.2014

(Miriam Buß)