

Statistische Modellierung mit Neuronalen Netzen und Anwendungen in der geographischen Informationsverarbeitung

[Czeranka, M.](#) und [A. Trapletti](#)

Gilgegasse 16/7
A-1090 Wien

Inhalt:

1. [Zusammenfassung](#)
2. [Einleitung](#)
3. [Von linearer zu nicht-linearer Regression](#)
4. [Modellieren mittels Feedforward Neuronalen Netzen](#)
5. [Anwendungsbeispiele für NN im Bereich der Geoinformationsverarbeitung](#)
6. [Schlußworte](#)
7. [Einführende Literatur und Online-Ressourcen](#)
8. [Verwendete Literatur](#)

[Zum Index!](#) - [Zur AGIT!](#)

1. Zusammenfassung

Dieser Beitrag gibt eine Einführung in das Gebiet der nichtlinearen Regressionsmodellierung mit Hilfe von Neuronalen Netzen (NN). Eine solchermaßen angestrebte quantitative Modellierung hat üblicherweise nicht die perfekte Repräsentation von beobachteten Daten zum Ziel, sondern verfolgt die Entwicklung eines statistischen Modells des zugrundeliegenden Prozesses. Die hier betrachtete Klasse der Vorwärtsgerichteten NN (Feedforward NN) erlaubt eine sehr allgemeine Modellierung des datengenerierenden Mechanismus; die Feedforward NN werden als eine allgemeine Form der nichtlinearen Regressionsmodelle vorgestellt. Der oftmals angeführte biologisch inspirierte Erklärungsansatz zur Funktionsweise von NN wird hier nicht explizit verfolgt. Auch eine allgemeine Einführung in die Theorie der Neuronalen Netze insgesamt wird nicht angestrebt. Hingegen wird der Einsatzbereich dieser NN-Technologie im Kontext der geographischen Informationsverarbeitung umrissen.

[Zum Inhalt!](#)

2. Einleitung

Der Begriff 'Neuronale Netze' (NN) findet sich immer öfter im Zusammenhang mit quantitativen und multivariaten Datenanalysen verschiedener Anwendungsfächer, wie z.B. auch im Zusammenhang mit

geographischer Informationsverarbeitung. Die Hintergründe bzw. die Funktionsweisen dieser modernen Technologie sind jedoch bisher relativ wenig bekannt. Für Nicht-Experten stellen Neuronale Netze noch eine magische Methode dar, mit welcher sich auf Tastendruck Massendaten verarbeiten lassen, gewünschte Werte in beliebiger Genauigkeit berechnet werden können oder Klassifikationen durchführbar sind, für die eine Vielzahl von Variablen gleichzeitig auszuwerten ist.

In diesem Aufsatz wird der Versuch unternommen, eine bestimmte Klasse von NN, die Vorwärtsgerichteten Neuronalen Netze (*Feedforward Neural Nets*) nicht nur als Black-Box Technologie vorzustellen, sondern sie in den Kontext von Statistik und Regression einzubetten. Der Einsatzbereich dieser NN-Architektur liegt insbesondere im Bereich der nichtlinearen Regression.

Nach einer Einführung in die klassische Regressionsanalyse werden die Besonderheiten dieser Modelle vorgestellt. Dabei werden interessante Punkte bezüglich der Umsetzung der NN-Modellierung herausgearbeitet. Auf eine tiefergehende Erläuterung der Arbeitsweise der Netze muß hier allerdings verzichtet werden. Abschließend wird über die theoretischen Erläuterungen hinaus ein Überblick über den Einsatzbereich von NN im Kontext der Geoinformationsverarbeitung gegeben.

[Zum Inhalt!](#)

3. Von linearer zu nicht-linearer Regression

Angewandte Aufgabestellungen im Bereich der quantitativen Datenverarbeitung haben oftmals das Ziel, den Einfluß bestimmter Variablen auf untersuchte Zielvariablen festzustellen. Es wird also eine mathematische Funktion gesucht, mit welcher die Beziehung zwischen den Variablen möglichst gut ausdrückt wird. Eine deterministische Beziehung ist nur bei exakten Zusammenhängen zwischen den untersuchten Variablen formulierbar. Häufig wird jedoch die Zielvariable nicht vollständig durch die Einflußvariablen erklärt, was die Verwendung eines statistischen Modells erzwingt. Ein einfaches statistisches Modell, das die Zielgröße y in Abhängigkeit von K Einflußvariablen (x_1, \dots, x_K) beschreibt, ist

$$y_t = f(x_{1t}, \dots, x_{Kt}, \mathbf{B}) + \varepsilon_t \quad \text{mit } t = 1, \dots, T, \quad (1)$$

wobei T die Anzahl der Beobachtungen angibt. \mathbf{B} ist der Vektor aller unbekannt Parameter und bestimmt den spezifischen Einfluß der Variablen x_k auf die Zielvariable y . ε ist der Fehlerterm, der unter anderem von Meßungenauigkeiten herrühren kann. Auch die Nichtberücksichtigung von Einflußvariablen wird mit diesem Fehlerterm abgedeckt. Im Regressionskontext nimmt man üblicherweise an, daß ε normalverteilt ist, da davon ausgegangen wird, daß sich ε aus vielen kleinen Störgrößen additiv zusammensetzt.

Im klassischen linearen Regressionsmodell ist der Zusammenhang zwischen den Einflußvariablen x_1, \dots, x_K und der Zielvariable y linear:

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_K x_{Kt} + \varepsilon_t. \quad (2)$$

Im einfachsten Fall - der Bestimmung des Einflusses einer einzigen Variable auf eine Zielvariable - charakterisieren nur zwei unbekannt Parameter den Zusammenhang. Dabei definieren die Regressionskonstante β_1 und der Regressionskoeffizient β_2 die Lage der Regressionsgeraden in der Ebene. Im multivariaten Fall wird der Zusammenhang zwischen den Einfluß- und Zielvariablen durch eine Ebene (bei zwei Einflußvariablen) bzw. bei weiteren Einflußvariablen durch eine sogenannte Hyperebene beschrieben. Die in [Gleichung \(2\)](#) aufgeführte erste unabhängige Variable x_1 ist üblicherweise eine Konstante, die gleich 1 gesetzt wird. β_1 stellt also auch im multivariaten Fall die Regressionskonstante dar.

Um nun die unbekannt Parameter $\mathbf{B} = (\beta_1, \dots, \beta_K)'$ aus den beobachteten Daten bestimmen zu können, muß ein sogenanntes Schätzverfahren verwendet werden. Dies geschieht üblicherweise mittels der Methode der kleinsten Fehlerquadrate. Die stochastische Begründung hierfür liefert die Maximum-Likelihood Methode ([vgl. Harvey,](#)

1990). Die intuitive Begründung hingegen basiert darauf, daß die Abweichung der beobachteten Werte zu jenen Werten, die mit dem Modell für y berechnet werden, möglichst minimiert werden soll.

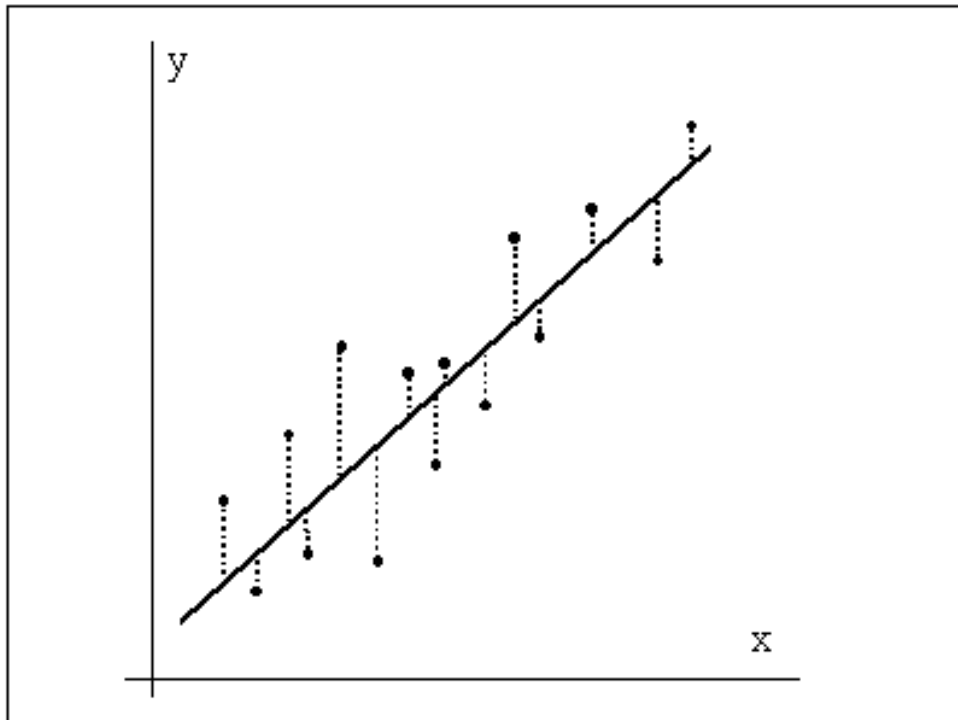


Abb. 2.1. Residuen: Abstände der beobachteten Werte (•) von der Regressionsgeraden

Anschaulich läßt sich die Methode der kleinsten Fehlerquadrate folgendermaßen erläutern: die in [Abb. 2.1.](#) dargestellte Gerade ist der geschätzte Zusammenhang zwischen den Variablen unter Annahme bestimmter Parameterwerte für β . Die Punkte repräsentieren die tatsächlichen Beobachtungen der Einfluß- und Zielvariablen. Die Länge der vertikalen Entfernungen dieser Punkte zur Regressionsgerade entspricht der Größe der Residuen. Andere Werte der Regressionsparameter β würden eine andere Schätzgerade und damit auch ein anderes Set an Residuen produzieren. Die Regression verfolgt nun das Ziel, β so zu schätzen, daß die Summe der quadrierten Residuen (also die Summe der quadrierten Fehler) minimiert wird. Dazu muß die Fehlerfunktion E in Abhängigkeit von β minimiert werden. Diese Fehlerfunktion E ist:

$$E(\beta) = \sum_{t=1}^T \varepsilon_t^2 = \sum_{t=1}^T (y_t - \beta_1 x_{1t} - \dots - \beta_K x_{Kt})^2 \quad (3)$$

Zur Minimierung der Fehlerfunktion $E(\beta)$ wird der Gradient von $E(\beta)$ (also die partiellen Ableitungen von $E(\beta)$ nach β) gleich Null gesetzt. Das Auflösen dieses Gleichungssystems ergibt sodann die Lösungen für β .

Die Beschränkung eines Modells auf den bisher betrachteten Fall des linearen Zusammenhanges zwischen den Variablen kann jedoch unter Umständen eine zu restriktive Annahme sein. Darum werden bei nicht-linearen Regressionsmodellen allgemeine Funktionen $f(\cdot)$ in Gleichung (1) erwogen. Diese Verallgemeinerung des Modells ist allerdings mit erheblichem Mehraufwand für die Schätzung der Modellparameter β verbunden. Falls nämlich der Zusammenhang zwischen den Variablen nichtlinear ist, lassen sich im Allgemeinen die Regressionsparameter nicht mehr, wie beschrieben, analytisch bestimmen: das Nullsetzen der partiellen Ableitungen führt nicht mehr zu geschlossenen Lösungen für die Parameterschätzer. Daher müssen die Lösungen mittels numerischer Verfahren gesucht werden. Der Weg dahin führt - ebenso wie bei den linearen Modellen - über die kleinsten Fehlerquadrate sowie den Gradienten von $E(\beta)$ ([s. Kap. 4.3.](#)).

Eine interessante Klasse nichtlinearer Regressionsmodelle stellen die im folgenden vorgestellten Feedforward NN dar.

4. Modellieren mittels Feedforward Neuronalen Netzen

Neuronale Netze bieten sich für die nichtlineare Regressionsmodellierung an, wenn der funktionale Zusammenhang zwischen den Variablen nicht explizit bekannt ist oder wenn dieser Zusammenhang aufgrund seiner Komplexität nicht ausreichend genau formulierbar ist. Innerhalb der Familie der NN ist insbesondere die Klasse der *Feedforward NN* (Vorwärtsgerichtete NN) interessant: bereits mit einer Zwischenschicht ([s. Kap. 4.2.](#)) und unter Verwendung einer nichtlinearen Aktivierungsfunktion ([s. Kap. 4.1.](#)) können sie jeden beliebigen Zusammenhang zwischen Einfluß- und Zielvariablen beliebig genau abbilden (universale Funktionsapproximation, [Hornik et al. 1989](#)).

4.1. Der grundlegende Baustein eines NN: das Verarbeitungselement

Wie der Name 'Neuronales' Netz schon andeutet, war die ursprüngliche Entwicklung ihrer Struktur wie auch ihrer Funktionsweise von biologischen neuronalen Netzen inspiriert (z.B. Funktionsweise des Gehirns). Oftmals wird daher zur Unterscheidung der Begriff 'künstliche' NN verwendet. Die NN bestehen aus miteinander vernetzten Verarbeitungselementen (auch: *Units*, *Cells*, *Processing Elements*, *Neurons*). Über die Verbindungen zwischen den einzelnen Elementen werden Impulse weitergegeben, wodurch das Netz letztlich Information generieren kann (z.B. Muster erkennen oder Regressionen durchführen kann).

Die Grundrechenritte, die zur Funktionsmodellierung notwendig sind, finden in den Verarbeitungselementen statt ([Abb. 3.1.](#)). Die Variablenwerte aller eintreffenden Verbindungen (im Falle von nur einer Zwischenschicht, [s. Kap. 4.2.](#), die Eingabewerte bzw. die Werte der Einflußvariablen) werden auf dem Weg zum Verarbeitungselement gewichtet und dort aufsummiert. Im Verarbeitungselement wird eine Funktion (sog. Aktivierungsfunktion) auf diese Summe angewendet, wodurch die Stärke des weiterzugebenden Impulses bestimmt wird. Wenn die Aktivierungsfunktion eine Schwellwertfunktion wäre, würde so bestimmt, ob überhaupt ein Impuls weitergeleitet wird. Der mittels dieser Aktivierungsfunktion berechnete Ausgabewert pro *Unit* wird dann an alle *Units* der nächsten Verarbeitungsebene weitergegeben.

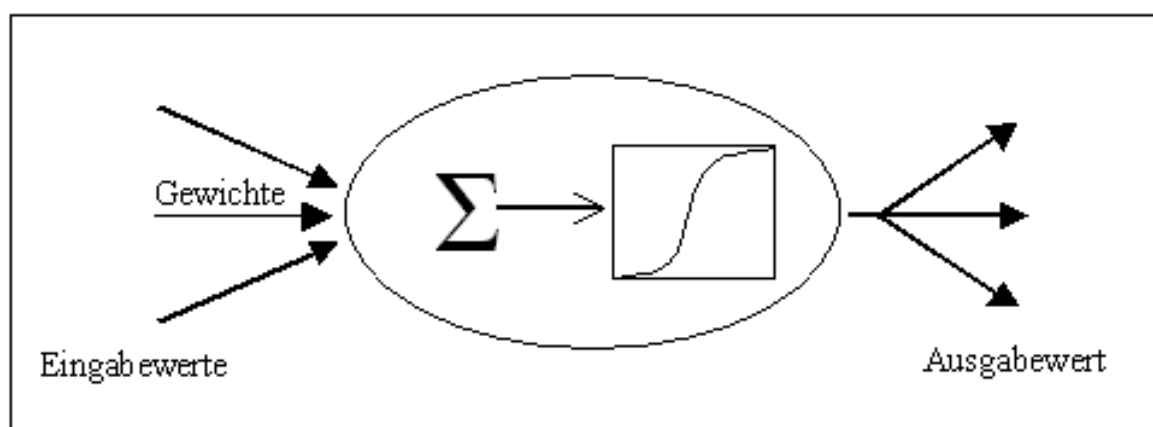


Abb. 3.1: Schematische Darstellung eines Verarbeitungselements der Zwischenschicht (Hidden Unit)

Die Nichtlinearität im NN wird durch die spezielle Form der Aktivierungsfunktion in den Verarbeitungselementen der Zwischenschicht (und evtl. auch der Ausgabeschicht) erreicht. Wie in [Abb. 3.1.](#) angedeutet wird, ist der Verlauf dieser Funktion üblicherweise sigmoid.

4.2. Die Netztopologie von Feedforward NN

Im Feedforward NN (auch: Mehrebenen Perzeptron, *Multilayer Perceptron*) sind die Verarbeitungselemente in Schichten (*Layers*) angeordnet. Die Variablen, welche die Daten charakterisieren (also die aus der Regression

bekannten Einflußvariablen x_k), sind als Eingabeeinheiten (*Input Units*) in der Eingabeschicht angeordnet. Die Zielvariablen (*Output Units*) - bzw. im Falle der Regression häufig nur die eine gesuchte Zielvariable y (vgl. [Abb. 3.2.](#)) - finden sich in der Ausgabeschicht wieder.

Das in [Abb. 3.2.](#) verdeutlichte Zusammenspiel von Verarbeitungselementen und Verarbeitungsschichten ergibt in Kombination mit [Gleichung \(1\)](#) das Feedforward NN Regressionsmodell:

$$f(x_{1t}, \dots, x_{kt}, \beta, \gamma) = F \left[\beta_0 + \sum_{j=1}^Q \beta_j G \left(\gamma_{j0} + \sum_{k=1}^K \gamma_{jk} x_{kt} \right) \right]. \quad (4)$$

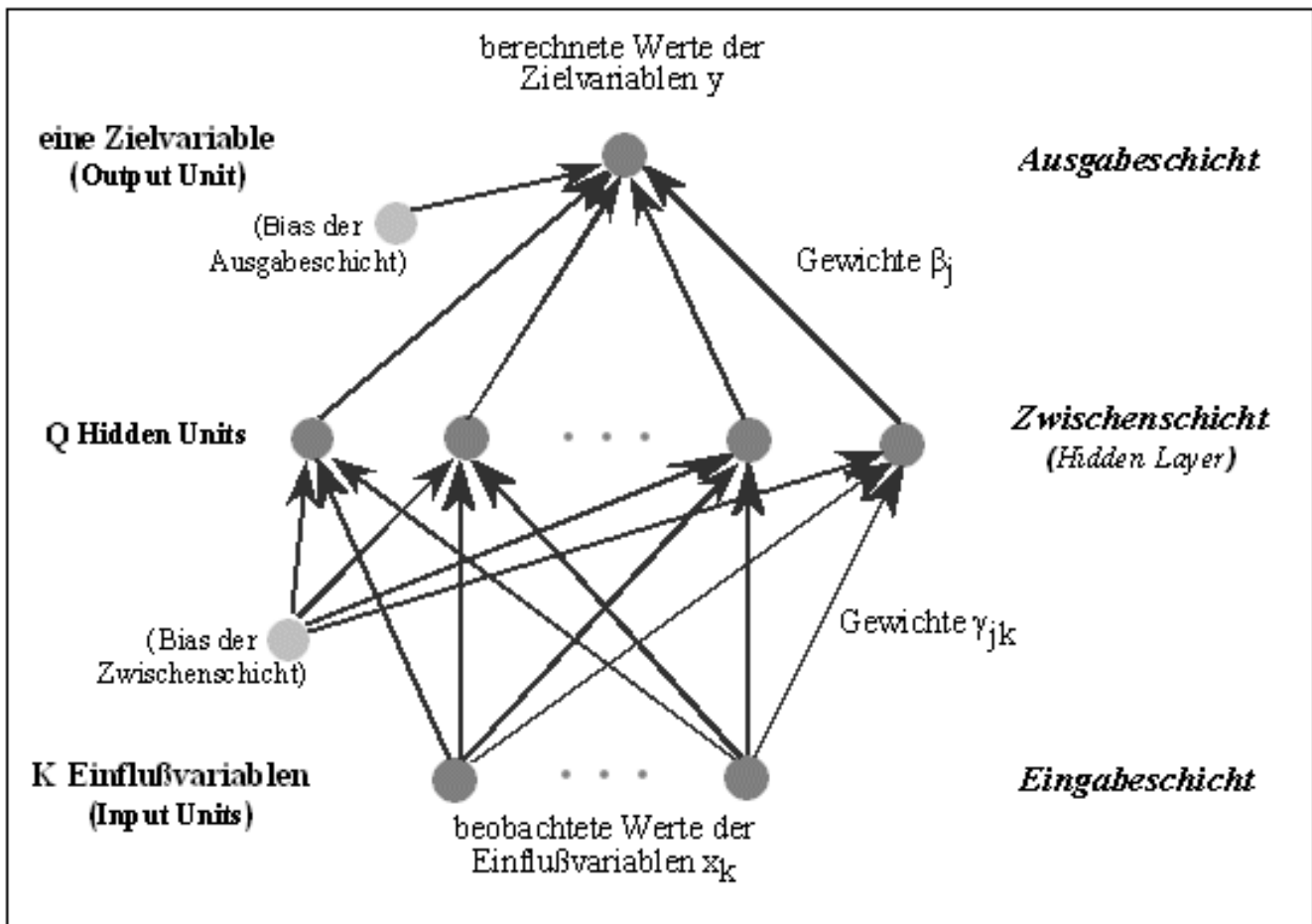


Abb. 3.2: Architektur eines Feedforward Neuronales Netzes

Der Wert der Zielvariablen hängt also von den Werten der Einflußvariablen sowie von den Gewichten (vgl. [Regressionskoeffizienten, Kap. 3](#)) zwischen den Verarbeitungsebenen ab. Die Gewichte β_0 und γ_{j0} entsprechen den Regressionskonstanten im klassischen linearen Regressionsmodell. Sie heißen in einem NN *Bias* und sorgen dafür, daß die Aktivierungsfunktion pro Unit beliebig im Raum liegen kann. G bezeichnet die Aktivierungsfunktion für die Zwischenschicht und F die Aktivierungsfunktion für die Ausgabeschicht (bezüglich der universalen Funktionsapproximationseigenschaft muß F nicht unbedingt nichtlinear sein, z.B. könnte F die Identitätsfunktion sein). Q gibt die Anzahl der *Hidden Units* an und K die Anzahl der Einflußvariablen. γ ist der Vektor aller Parameter (Gewichte) von der Eingabeschicht zur Zwischenschicht und β ist der Vektor aller Parameter (Gewichte) von der Zwischenschicht zur Ausgabeschicht (im hier betrachteten Fall einer einzigen Zielvariablen also zur einzigen Ausgabeeinheit y).

Üblicherweise wird zur Regressionsmodellierung ein voll vernetztes Feedforward NN mit einer Zwischenschicht verwendet. In dieser Netzarchitektur gibt es keine Rückkopplung zwischen Ausgabeschicht, Zwischenschicht und

Eingabeschicht (*feedforward*) und alle Units einer Schicht sind mit allen Units der jeweils nachfolgenden verbunden. Um nun ein bestimmtes Problem mittels eines NN modellieren zu können, müssen darüber hinaus weitere Entscheidungen getroffen werden: für die erreichbare Güte der Modellierung ist insbesondere ausschlaggebend, wieviele freie Parameter das Netz besitzt. Unter der Annahme, daß die Anzahl der Einfluß- und Zielvariablen für die Regression jeweils fix vorgegeben ist, ist also die Anzahl der *Hidden Units* maßgebend ([Q in Gleichung \(4\)](#)). Die Bestimmung der angemessenen Anzahl dieser *Hidden Units* gehört zum Modellselektionsproblem ([Kap. 4.4.](#)).

4.3. Schätzung der freien Parameter im NN

Analog zur Bestimmung der Regressionskoeffizienten beim klassischen Regressionsmodell sind auch bei der NN-Modellierung die Netzwerkparameter so zu schätzen, daß die Summe der Fehlerquadrate - also die Abweichung der prognostizierten von den beobachteten Zielwerten - möglichst gering wird ([s. Kap. 3](#); diese Methode ist auch hier nur angebracht, wenn die Fehlergrößen normalverteilt sind). Aufgrund der (gewünschten) Nichtlinearität ist allerdings eine analytische Bestimmung der optimalen Parameter, wie bei den linearen Modellen, nicht mehr möglich. Durch numerische Verfahren lassen sich die Parameter dennoch schätzen (in NN-Terminologie: Trainieren oder Lernen der Gewichte). Dieses Anpassen der Gewichte ist eine Optimierungsaufgabe, wozu entweder deterministische (s.u.) oder stochastische Methoden verwendet werden können. Zu letzteren gehören Suchverfahren, wie z.B. Genetische Algorithmen, welche die globale Minimierung der Fehlerfunktion (bzw. die Maximierung einer Fitnessfunktion) zum Ziel haben.

Ebenso wie bei der klassischen linearen Regression ist die Fehlerfunktion E abhängig von den Parametern (nunmehr β und γ); sie lautet analog zu [Gleichung \(3\)](#):

$$E(\beta, \gamma) = \sum_{t=1}^T (y_t - f(x_{1t}, \dots, x_{kt}, \beta, \gamma))^2 . \quad (5)$$

Im folgenden werden nur deterministische Verfahren zur Fehlerminimierung betrachtet. Dies sind üblicherweise Gradientenabstiegsverfahren (*Gradient Descent Techniques*). Eine hinsichtlich des Berechnungsaufwandes effiziente Methode zur Berechnung des Gradienten ist (*Error-Backpropagation* (Fehlerrückvermittlung)). Das einfachste Gradientenabstiegsverfahren ist ein Verfahren erster Ordnung, wobei die Parameter iterativ in Richtung des negativen Gradienten angepaßt werden (erster Ordnung, da nur die ersten Ableitungen der Fehlerfunktion bzgl. der Parameter verwendet werden).

Die Fehlerfunktion ([Gleichung \(5\)](#)) kann sehr kompliziert sein, mit vielen lokalen Minima oder flachen Tälern: der Gradient zeigt also nicht notwendigerweise in Richtung des globalen Minimums. Deterministische Verfahren zur Fehlerminimierung enden daher typischerweise nur in einem lokalen Minimum, wobei der Lernprozeß zudem sehr langsam sein kann. Verschiedenste Varianten an Gradientenabstiegsmethoden, wie z.B. die konjugierte Gradientenmethode (*Conjugate Gradient Method*), weisen aber zumindest ein weitaus besseres Konvergenzverhalten als das einfache Gradientenabstiegsverfahren auf.

Eine sehr einfache Topographie der Fehlerlandschaft ist in [Abb. 3.3](#) skizziert: der Fehler ist in Abhängigkeit von β und γ (gezwungenermaßen nur 3-dimensional) dargestellt. Der negative Gradient ist mit $-\nabla E(\beta, \gamma)$ an der Stelle (β, γ) eingezeichnet und zeigt in Richtung des Minimums. n ist die Anzahl aller Schätzparameter des NN (also die Anzahl aller Gewichte β und γ).

Hier sei darauf hingewiesen, daß die Effizienz der gewählten Lernstrategie von verschiedenen weiteren Faktoren und Implementierungsdetails beeinflusst wird, wie z.B. von der Verteilung der Anfangsgewichte und der Art der Präsentation der Lernbeispiele (Reihenfolge oder Anzahl der gleichzeitig präsentierten Beispieldaten, etc.). Tiefergehende Erläuterungen hierzu finden sich in einschlägigen Lehrbüchern; einige sind am Ende des Beitrages aufgeführt.

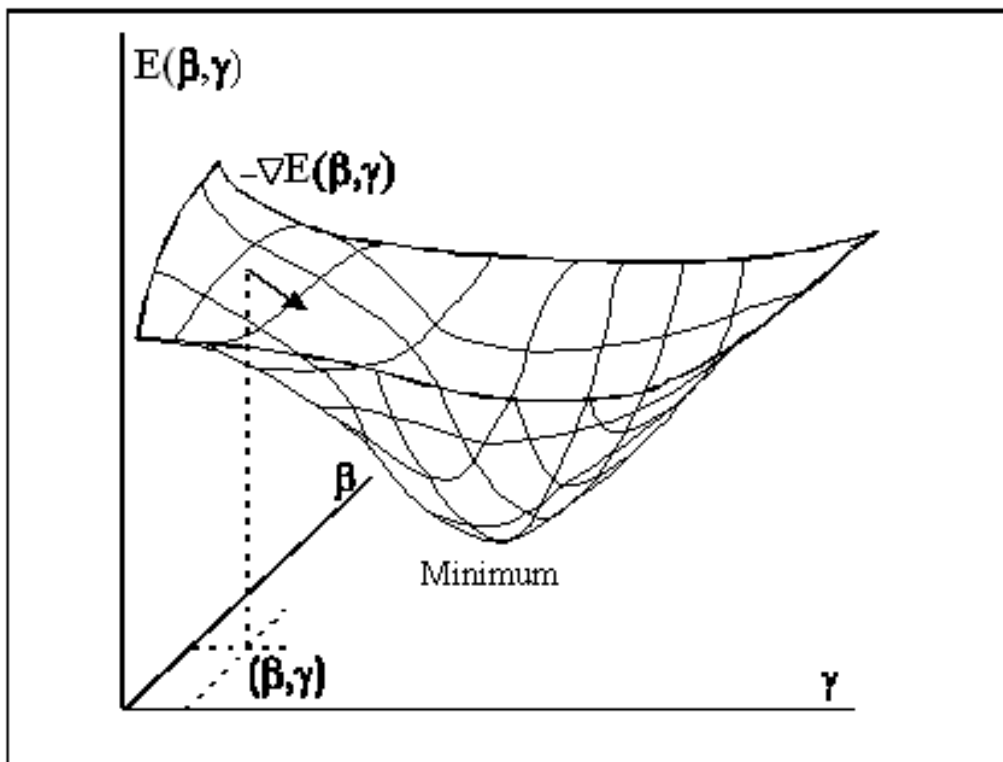


Abb. 3.3: Der Fehler als n -dimensionale Hyperfläche im $n+1$ -dimensionalen Raum

4.4. Bestimmung der Anzahl der Hidden Units

Güte, Genauigkeit und speziell die Generalisierungsfähigkeit eines Netzes hängen, neben der bereits diskutierten Art der Parameterschätzung, fundamental von der Anzahl der Hidden Units ab (diese begrenzen die mögliche Komplexität der Funktion). Ebenso ist das Verhältnis der Anzahl dieser freien Parameter zur Größe des beobachteten Datensatzes bedeutend. Zu wenig adaptive Parameter (Gewichte zwischen den einzelnen Units) bedingen eine ungenügende Lernfähigkeit des Netzes; zu viele adaptive Parameter bewirken ein *Overfitting* der Beispieldaten (Trainingsset). Beides würde eine schlechte Prognosequalität des Netzes bedeuten.

Je mehr Freiheitsgrade existieren, desto besser kann die Anpassung des Modells an die Daten erfolgen. Hingegen verfolgt eine Kurve (bzw. im multivariaten nichtlinearen Fall eine Hyperfläche), die von zu vielen Parametern abhängt, jedes kleine Detail der Beispieldaten ([vgl. Abb. 3.4](#)). Diese Details sind aber nicht unbedingt von Bedeutung, sondern können einzig und allein durch den Fehlerterm bedingt sein (*Noise*) und damit für generelle Aussagen unerwünscht sein.

Trotz der grundlegenden Bedeutung, die der Bestimmung der angemessenen Anzahl von Modellparametern zukommt, gibt es jedoch noch keine vollends befriedigende Lösung dieses Problems. Die bekanntesten Ansätze zur Optimierung der Modellkomplexität sind Kreuzvalidierung (*Crossvalidation*), *Network Pruning* und diverse *Information Criteria* ([vgl. Bishop 1995](#)). Abgesehen vom verwendeten Verfahren ist es aber grundsätzlich notwendig, die Generalisierungsfähigkeit des geschätzten Modells mit Hilfe eines unabhängigen Testsets beobachteter Daten zu validieren.

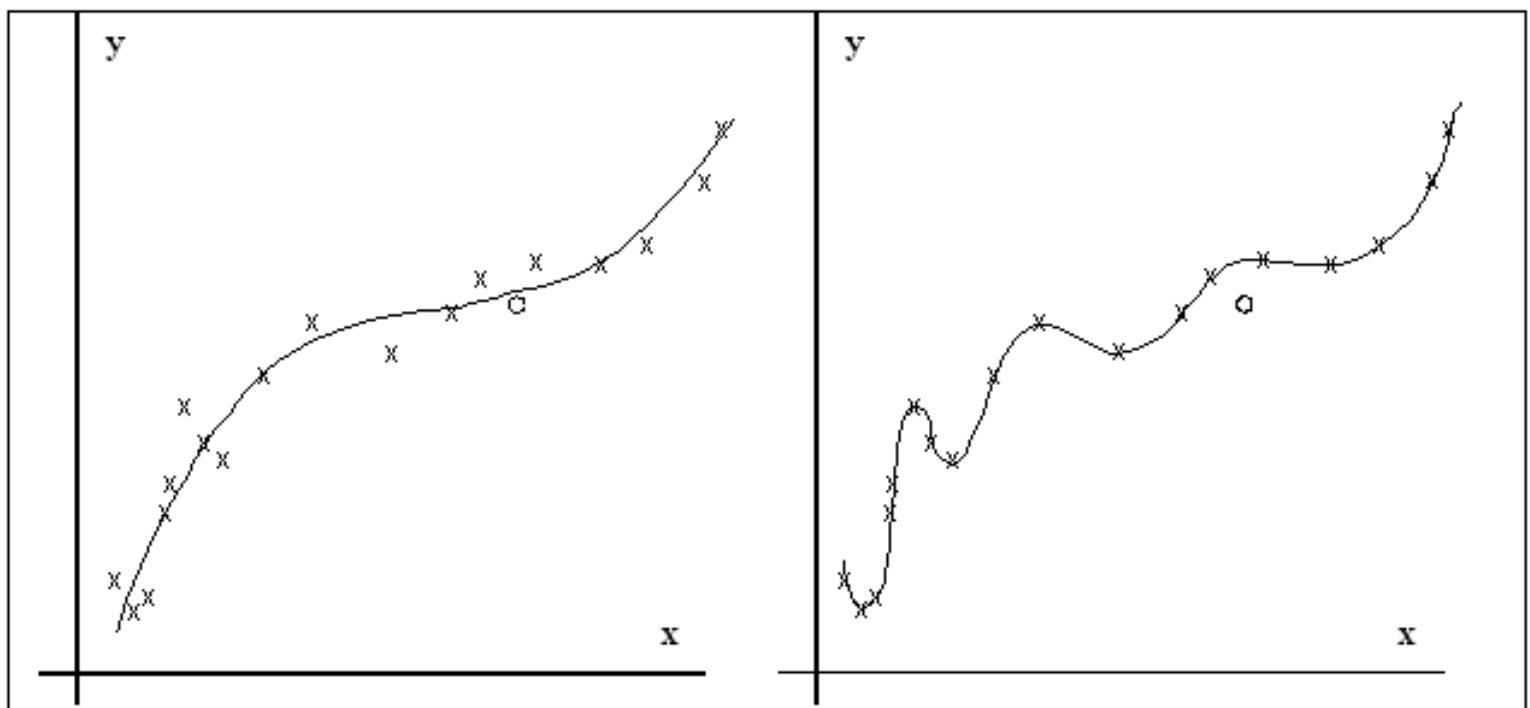


Abb. 3.4: Gute Anpassung (links) der Regressionskurve an verrauschte Daten versus Overfitting (rechts)

4.5. Vor- und Nachteile von (Feedforward) Neuronalen Netzen

Die Stärke von NN liegt darin begründet, daß sie jeden beliebigen Zusammenhang beliebig genau abbilden können. Ihr großer Vorteil gegenüber klassischen linearen Modellen besteht zudem darin, daß die Zusammenhänge zwischen den Variablen völlig unbekannt und beliebig komplex sein dürfen. Allerdings liegt genau hierin gleichzeitig auch die Schwäche der NN-Modellierung: da a priori eine Quantifizierung der stochastischen Modellkomponente (also des Fehlerterms) nicht möglich ist, besteht die Gefahr, daß das entwickelte Modell die Beispieldaten überbestimmt (*Overfitting*) und somit keine guten Generalisierungsfähigkeiten besitzt. Die Modellgüte hängt daher ganz entscheidend davon ab, welche Methode zur Modellschätzung geführt hat, wie aussagekräftig die für die Modellschätzung herangezogenen Beispieldaten sind und wie groß das Verhältnis von freien Parametern zu beobachteten Datensets ist.

Ein weiterer Nachteil der NN-Modellierung ist der für die Modellschätzung benötigte Zeitaufwand, da anstelle von analytischen Lösungen iterative Schätzverfahren verwendet werden müssen. Diese können vor allem für größere Netze sehr rechenintensiv sein. Darüber hinaus sei hier erwähnt, daß ein NN-Modell grundsätzlich nur schwer inhaltlich/fachlich interpretierbar ist. Zumindest aber können anhand der ermittelten Netzgewichte Rückschlüsse auf die Bedeutung bestimmter Einflußvariablen gezogen werden.

[Zum Inhalt!](#)

5. Anwendungsbeispiele für NN im Bereich der Geoinformationsverarbeitung

Die beiden Bücher [Hewitson und Crane \(1994\)](#) sowie [Openshaw und Openshaw \(1997\)](#) bieten sich für einen Überblick über den Einsatzbereich von NN in der geographischen Informationsverarbeitung an. Darüber hinaus finden sich mittlerweile relativ viele Fachbeiträge zum Einsatz von NN (zumeist handelt es sich dabei um die hier behandelten Feedforward Modelle) in verschiedensten Zeitschriften und Tagungsbänden in den Bereichen: Modellierung, Simulation, GeoComputation, sowie vermehrt auch in einigen Anwendungsfächern, speziell in den Fachgebieten Hydrologie und Klimatologie. Zumeist handelt es sich dabei (noch) um Arbeiten im Rahmen von Forschungsprojekten.

Neben der Regressionsmodellierung werden die NN-Modelle zur quantitativen Datenverarbeitung bei Zeitreihenanalysen eingesetzt: die hier vorgestellten Feedforward Modelle sind quasi unverändert auf die autoregressive Zeitreihenmodellierung übertragbar. Auch zur multivariaten Klassifikation werden diese Feedforward Modelle verwendet: die Modellarchitektur wird nur dahingehend erweitert, daß mehr als nur eine Zielvariable existiert (Einsatzbereich in der Geoinformationsverarbeitung: z.B. digitale Satellitenbilddauswertung). Für die Lösung dieser letztgenannten Aufgaben existieren über die Feedforward Modelle hinaus verschiedene weitere NN-Modelle, wie z.B. Rekurrente Netze, Hopfield-Netze oder ART-Modelle.

Im folgenden sind einige Einsatzbereiche von Feedforward NN in der Geoinformationsverarbeitung zusammengestellt. So z.B. sind diese NN-Modelle bereits teilweise an die Stelle von traditionellen Regenwasser-Abflußmodellen getreten. Dabei werden Zeitreihen z.B. von Abflußmenge, Niederschlag und Temperatur als Einflußvariablen für die NN-Modellierung verwendet (u.a. bei [Minns u. Hall 1996](#)). Auch die Vorhersage täglicher Wasserstände, Flutvorhersagen oder Trinkwasserversorgungsprognosen sowie Vorhersagen der Lufttemperatur oder der Sonneneinstrahlung wurden mittlerweile mit NN durchgeführt und als erfolgreich dokumentiert. Im Bereich der Anthropogeographie gibt es vergleichbare NN-Anwendungen für Autoverkehrsprognosen oder Stromverbrauchsprognosen. Bei diesen Analyse- und Prognosemodellen werden z.T. Zeitreihen von Messungen an nur einem Ort, z.T. auch räumlich verteilte Datenreihen herangezogen.

NN werden also auch in der Geoinformationsverarbeitung verwendet, um neue Lösungsansätze zu erarbeiten, wenn die exakten Modellzusammenhänge nicht bekannt sind (und evtl. nichtlineare Zusammenhänge angenommen werden müssen). Auf der Grundlage von solchermaßen entwickelten NN-Modellen, die anhand präsentierter Beispieldatensätze implizit die Zusammenhänge gelernt haben (vorausgesetzt, die gewählten Einflußvariablen sind aussagekräftig für die gesuchten Zielvariablen), lassen sich spezifischere Aussagen ableiten und eventuell neuartige deterministische Modelle entwickeln. Ein Beispiel hierzu geben [Roadknight et al. \(1997\)](#), die ein NN zur Modellierung des Eintretens von Pflanzenkrankheiten auf der Grundlage von Klimadaten eingesetzt und anschließend Regeln aus den Netzgewichten abgeleitet haben. Ein weiteres Beispiel zur explorativen Datenanalyse mittels NN findet sich bei [Recknagel et al. \(1997\)](#). Dort wurde das Auftreten der Algenblüte in Seen modelliert. In Ermangelung eines deterministischen Modells haben die Autoren eine Vielzahl von Merkmalen abwechselnd in separaten NN miteinander kombiniert und so auf ihre Bedeutung als Einflußvariable hin untersucht (u.a. Wassertemperatur, pH-Wert, Regen, Einstrahlung, Stickstoff).

Auch räumliche Interaktionsmodelle sind ein vielversprechendes Anwendungsgebiet, bei welchem die Adaptierbarkeit und die Nichtlinearität der NN besondere Bedeutung erlangt haben (vgl. [Openshaw 1993](#); [Fischer und Gopal, 1994](#)). In diesen Interaktionsmodellen wird, wie bei den traditionellen Interaktionsmodellen, der Raumbezug über eine der Einflußvariablen hergestellt (z.B. Entfernung in Straßenkilometern zwischen den untersuchten Orten/Raumeinheiten).

[Zum Inhalt!](#)

6. Schlußworte

Die besonderen Probleme, die raumbezogene Fragestellungen aufgeben (wie die Modellierung räumlicher Heterogenität oder räumlicher Autokorrelation) werden bisher nicht explizit bei der NN-Modellentwicklung berücksichtigt. Hingegen wird davon ausgegangen, daß ein NN-Regressions- bzw. Klassifikationsmodell diese Zusammenhänge immanent anhand der Beispieldaten lernt. Inwiefern diese Annahme berechtigt ist, bzw. ob es nicht sogar effizientere Möglichkeiten der raumbezogenen NN-Modellierung gibt, ist noch nicht eingehend untersucht worden. Grundlagenforschung, welche auf die speziellen Bedürfnisse der Geoinformationsverarbeitung eingeht, ist also durchaus noch sinnvoll und notwendig.

Abschließend sei festgehalten, daß NN sehr mächtige Modellierungsinstrumente sind, welche allerdings - ohne entsprechende Kenntnisse eingesetzt - nicht unbedingt sinnvolle Ergebnisse produzieren. Eine unkritische Anwendung auf Datenmengen, insbesondere mit zu wenig Beispieldatensätzen und unter Verwendung zu vieler Freiheitsgrade, führt nicht zu generellen Erkenntnissen über die gesuchten Funktionszusammenhänge. Hingegen würde so nur ein Netz konstruiert, welches die Beispieldaten 'auswendig' lernt. Genau wie für die

Geoinformationsverarbeitung gilt natürlich auch für die hier vorgestellten neuartigen statistischen Analysewerkzeuge, daß es bei weitem nicht ausreicht, bestimmte Knöpfe einer Software bedienen zu können. Hingegen sollten die Umstände und die passenden Verarbeitungsschritte bekannt sein, unter denen mit aussagekräftigen Ergebnissen gerechnet werden kann.

[Zum Inhalt!](#)

7. Einführende Literatur und Online-Ressourcen

Bishop, C. (1995): Neural Networks for Pattern Recognition. Clarendon, Oxford (ein Standardwerk).

Dorffner, G. (1991): Konnektionismus. Teubner, Stuttgart (eines der wenigen deutschsprachigen Fachbücher).

Kennedy, P. (1992): A Guide to Econometrics. Blackwell, Oxford (3. Aufl.) (u.a. Einführung in Regression).

Hertz, J., Krogh, A. und Palmer, R.G. (1991): Introduction to the Theory of Neural Computing. Addison Wesley, Reading (ein Standardwerk).

Masters, T. (1993): Practical Neural Network Recipes in C++. Academic Press, San Diego (inkl. Diskette, unter Dos oder Unix verwendbar).

NN-FAQ: <ftp://ftp.sas.com/pub/neural/FAQ.html> (inkl. ausführlichen Literatur- und Softwarehinweisen und mit zahllosen weiterführenden Links).

Smith, M. (1993). Neural Networks for Statistical Modeling. Van Nostrand Reinhold, New York (relativ 'untechnische' Einführung mit wenigen Formeln, aber vielen anschaulichen Graphiken).

[Zum Inhalt!](#)

8. Verwendete Literatur

Bishop, C. (1995): Neural Networks for Pattern Recognition. Clarendon, Oxford.

Fischer, M.M. and S. Gopal (1994): Artificial Neural Networks: A New Approach to Modelling Interregional Telecommunication Flows, Journal of Regional Science 34, pp. 503-527.

French, M., W. Krajewski and R.R. Cuykendall (1992): Rainfall-Forecasting in Space and Time Using a Neural Network. Journal of Hydrology 173, pp. 1-31.

Harvey, A. (1990): The Econometric Analysis of Time Series. Hilip Allan, New York.

Hewitson, B.C. and R.G. Crane (eds.) (1994): Neural Nets: Applications in Geography. Kluwer, Dordrecht.

Hornik, K., Stinchcombe, M. and White, H. (1989): Multilayer Feedforward Networks are Universal Approximators. Neural Networks 2, pp. 359-366.

Minns, A.W. and M.J. Hall (1996): Artificial Neural Networks as Rainfall-Runoff-Models. Hydrological Sciences 41, pp. 399-417.

Openshaw, S. and C. Openshaw (1997): Artificial Intelligence in Geography. John Wiley & Sons, Chichester.

Openshaw, S. (1993): Modelling Spatial Interaction Using a Neural Net. In: M.M. Fischer and P. Nijkamp (eds.): GIS, Spatial Modelling and Policy Evaluation. Springer, Berlin, pp. 147-164.

Recknagel, F., M. French, P. Harkonen and K. Yabunaka (1996): Artificial Neural Network Approach for Modelling and Prediction of Algal Blooms. Ecological Modelling 96(1-3), pp. 11-28.

Roadknight, C.M., D. Palmer-Brown and G.E. Mills (1997): The Analysis of Artificial Neural Network Data Models. In: X. Liu, P. Cohen and M. Berthold (eds.): *Advances in Intelligent Data Analysis*. Springer, Berlin, pp. 337-346.

[Zum Inhalt!](#)

© 1998 *Institut für Geographie der Universität Salzburg*