Neuronale Netzwerke in der Ökonometrie
Die Entmythologisierung ihrer Anwendung

Ulrich Anders

Zentrum für Europäische
Wirtschaftsforschung (ZEW)

Mannheim im Juni 1995

Kaiserring 14-16
Postfach 10 34 34
68163 Mannheim

Tel: 0621/1235-141
Fax: 0621/1235-223
Email: anders@zew.de

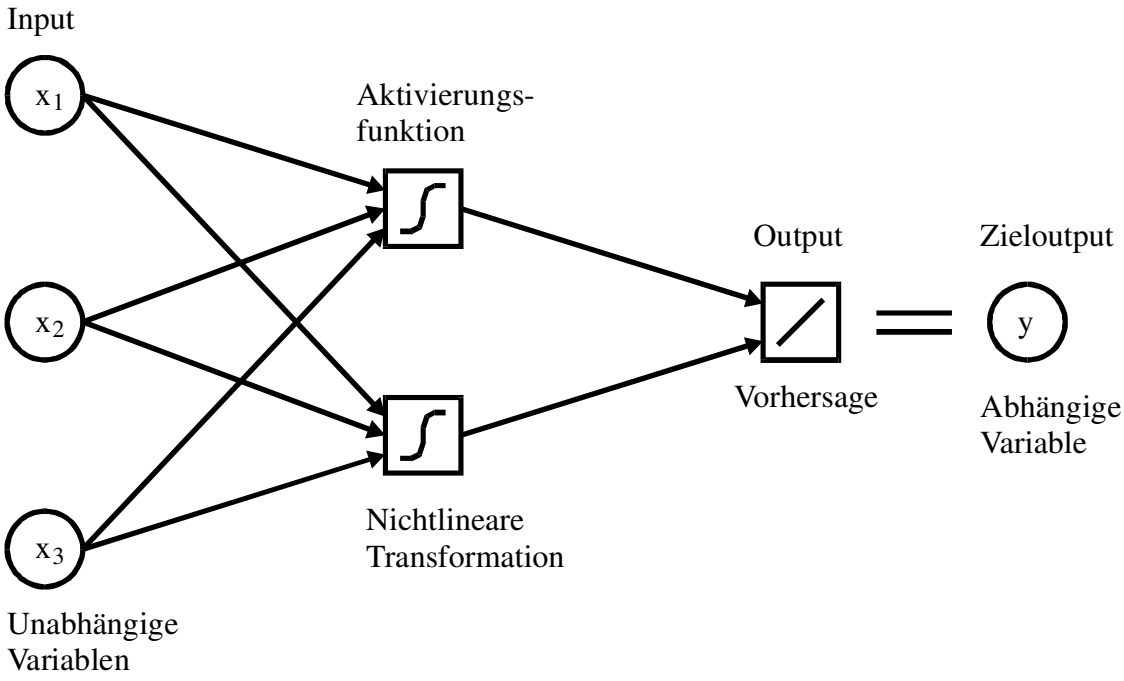
---

ftp.zew.de
login: anonymous

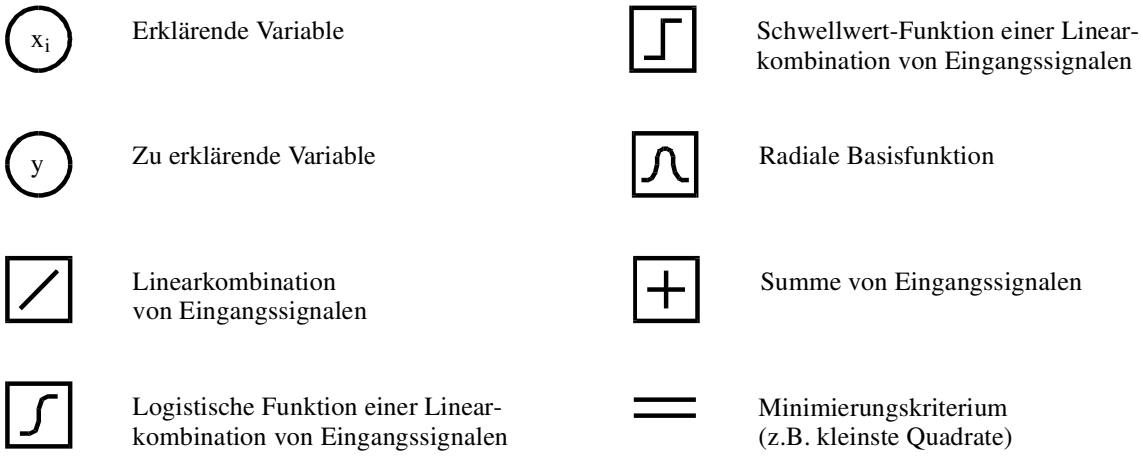
/incoming/uan/nninoeko.ps
/incoming/uan/nnfolien.ps

© Ulrich Anders

Was sind Neuronale Netzwerke?



Standardnetzwerk mit drei Schichten



Die Bausteine Neuronaler Netzwerke

© Ulrich Anders

- Typische Problemstellung:

  $y = F(X) + \varepsilon \qquad X = [X_1, X_2, \ldots, X_I]$

  $E[\varepsilon | X] = 0 \qquad \varepsilon$ ist ein *i.i.d.*-Störterm

  $F(X) = E[y | X]$ heißt Regression von y auf X.

- Ziel jeder ökonometrischen Anwendung:
  Approximation von *F(X)* durch $f(X, w)$.

$$f(X, w) = g_o\left(\sum_h \beta_h g_h\left(\sum_i \gamma_i x_i\right)\right) + \sum_i \alpha_i x_i$$

- Neuronale Netzwerke sind universale Approximatoren und können daher beliebige Funktionen *F(X)* abbilden.

- Neuronale Netzwerke machen keine expliziten Annahmen über die Modellstruktur von *F(X)*.

© Ulrich Anders

Parametrische und nichtparametrische Verfahren (I)

- Bei **parametrischen Verfahren** wird dem zu modellierenden Zusammenhang eine bestimmte funktionale Form unterstellt (Problem: Bias).

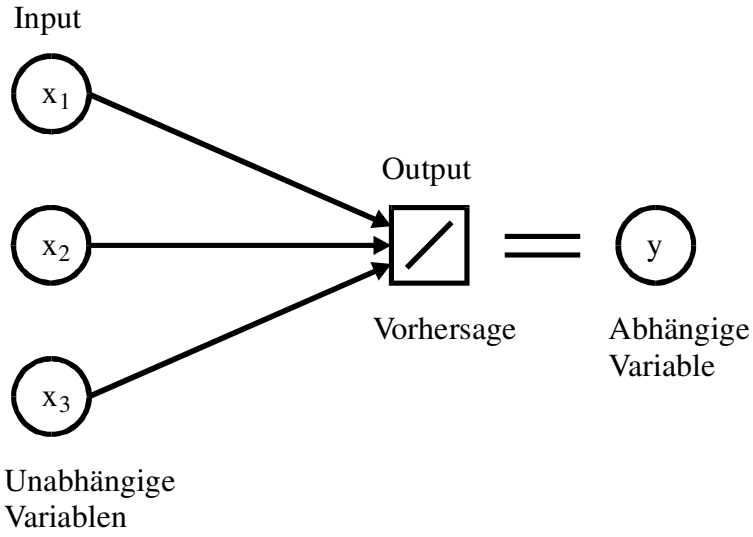
- Bei **nichtparametrischen Verfahren** werden keine Annahmen über die funktionale Form des Zusammenhangs gemacht (Problem: Varianz).

- **Semiparametrische Verfahren** sind Mischformen.

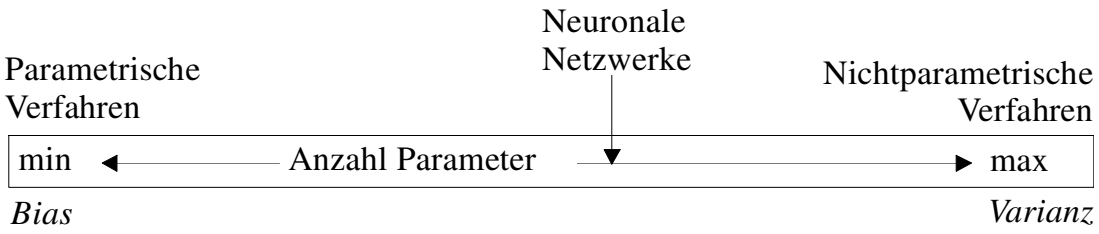
© Ulrich Anders

Parametrische und nichtparametrische Verfahren (II)

- Neuronale Netzwerke sind parametrische Verfahren mit einer nichtparametrischen Interpretation.



Das einfache lineare Regressionsmodell
als Neuronales Netzwerk



Kontinuum zwischen parametrischen
und nichtparametrischen Verfahren

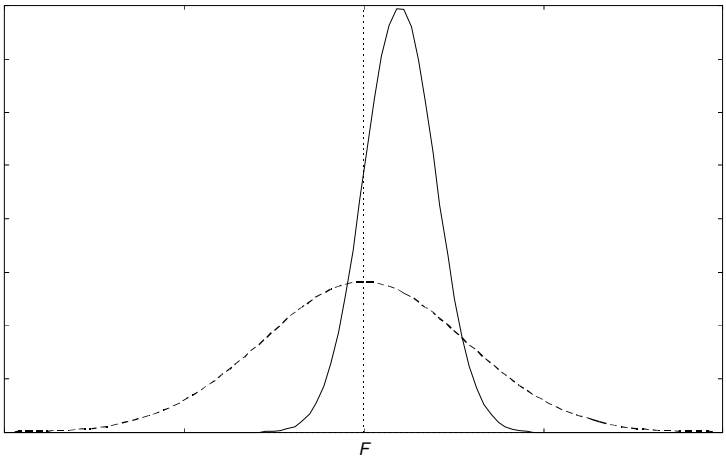
© Ulrich Anders

Exkurs: Das Bias/Varianz-Dilemma

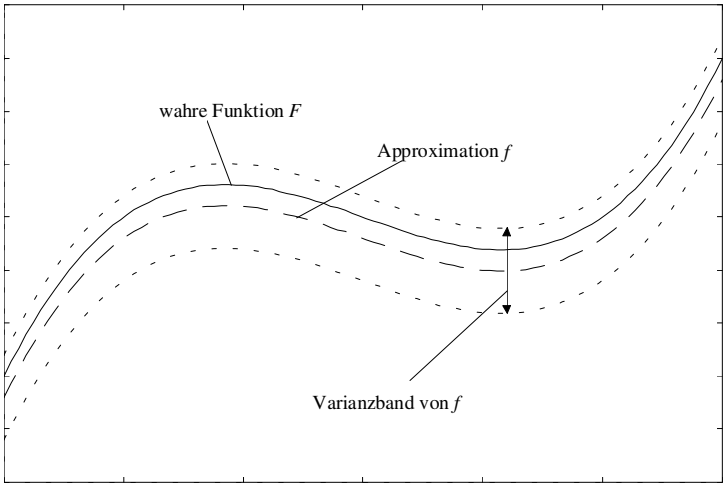
$$\begin{aligned}\mathrm{MSE} &= E[(y - f(X,w))^2] \\ &= E[(y - F(X))^2] + E[(F(X) - f(X,w))^2] \\ &= \mathrm{MSE}_u + \mathrm{MSE}_s\end{aligned}$$

$$\mathrm{MSE}_s = \mathrm{MSE}[f] = \mathrm{Bias}[f]^2 + \mathrm{Var}[f]$$

$$\mathrm{Bias}[f] = E[f] - F \qquad \mathrm{Var}[f] = \tfrac{1}{T}\sum_{t=1}^{T}(f_t - \bar{f})^2$$



Zwei mögliche Verteilungen eines Schätzers f



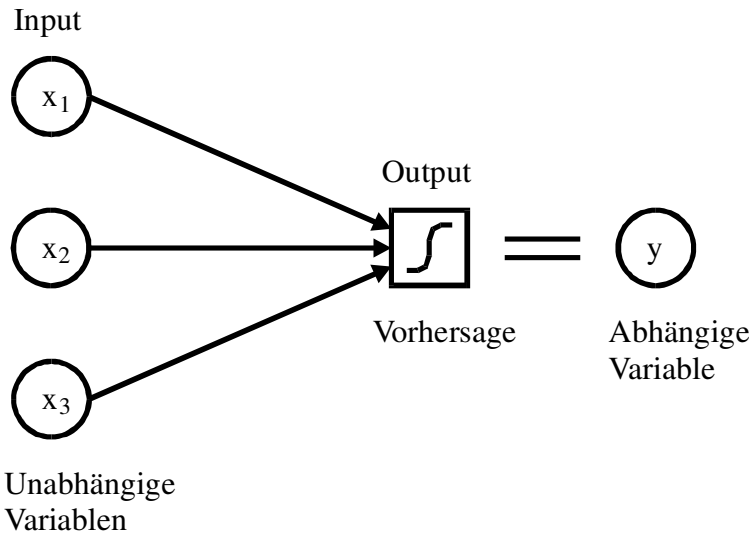
Bias und Standardabweichung eines Funktionsschätzers f

© Ulrich Anders

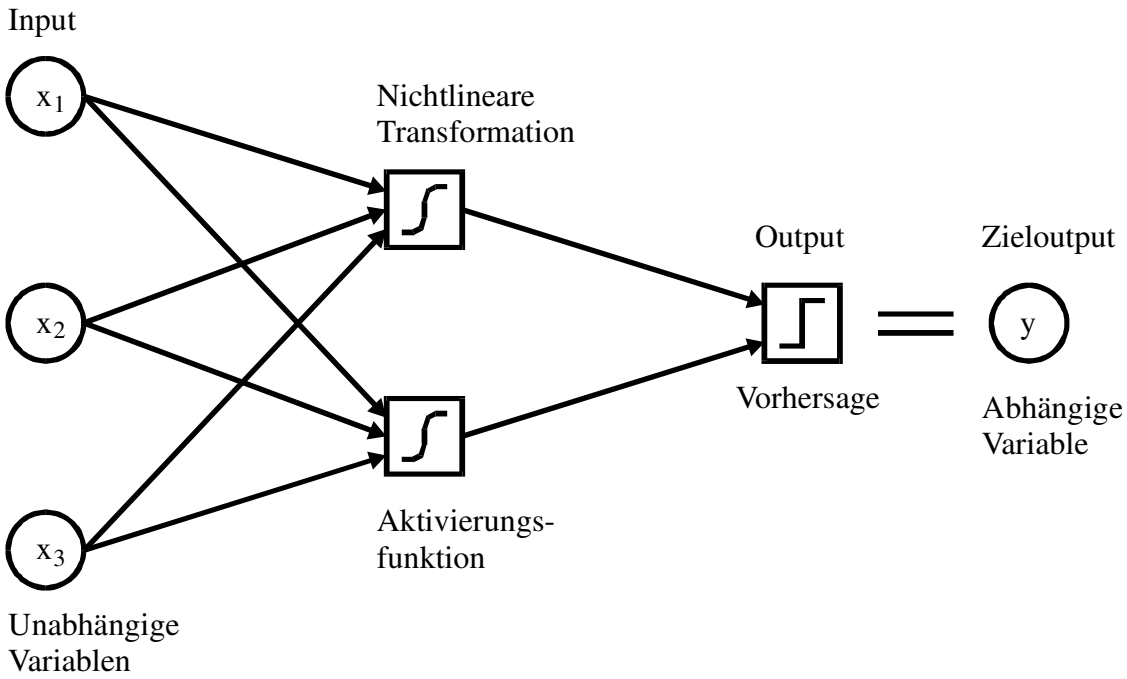
Glossar

Neuronales Netz	Parametrisches Modell
Netzwerkarchitektur	Modellspezifikation
Ungeeignete Netzarchitektur	Über- bzw. Unterparametrisierung
Aktivierungsfunktion	Nichtl. Transformationsfkt.
Inputs	Erklärende (unabhängige) Variablen
Outputs	Erklärte (abhängige) Variable
Gewicht	Parameter
Training	Parameterschätzung
Konvergenz	*in sample*-Qualität
Generalisierung	*out of sample*-Qualität
Muster	Beobachtung

© Ulrich Anders

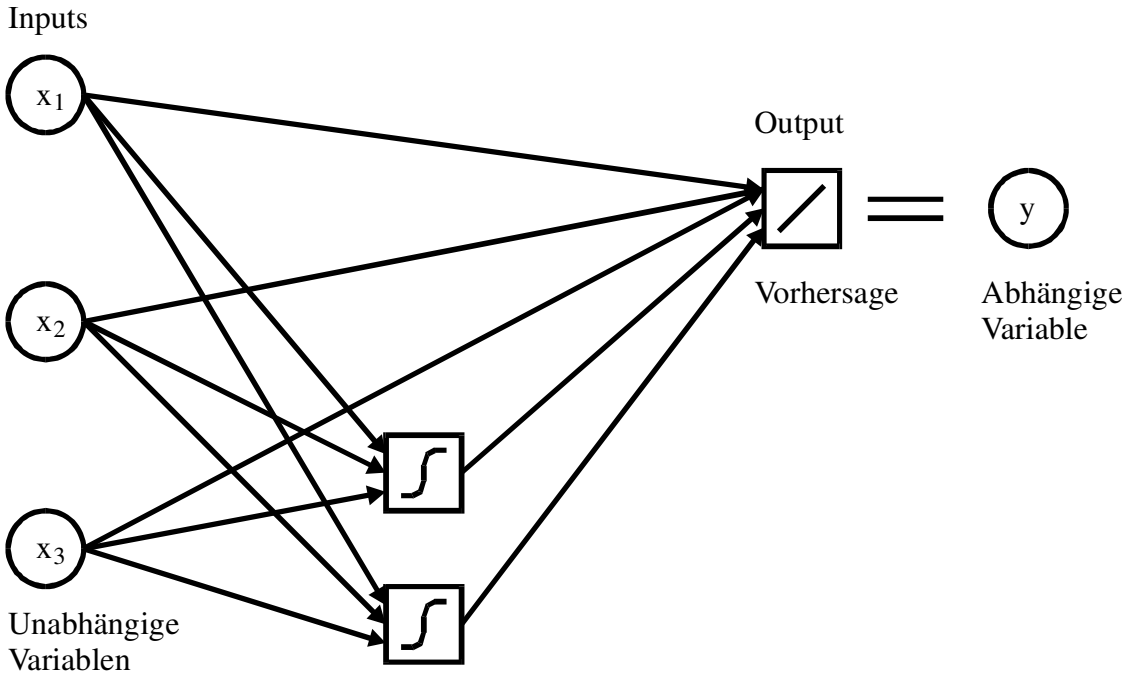


Logistisches Regressionsmodell

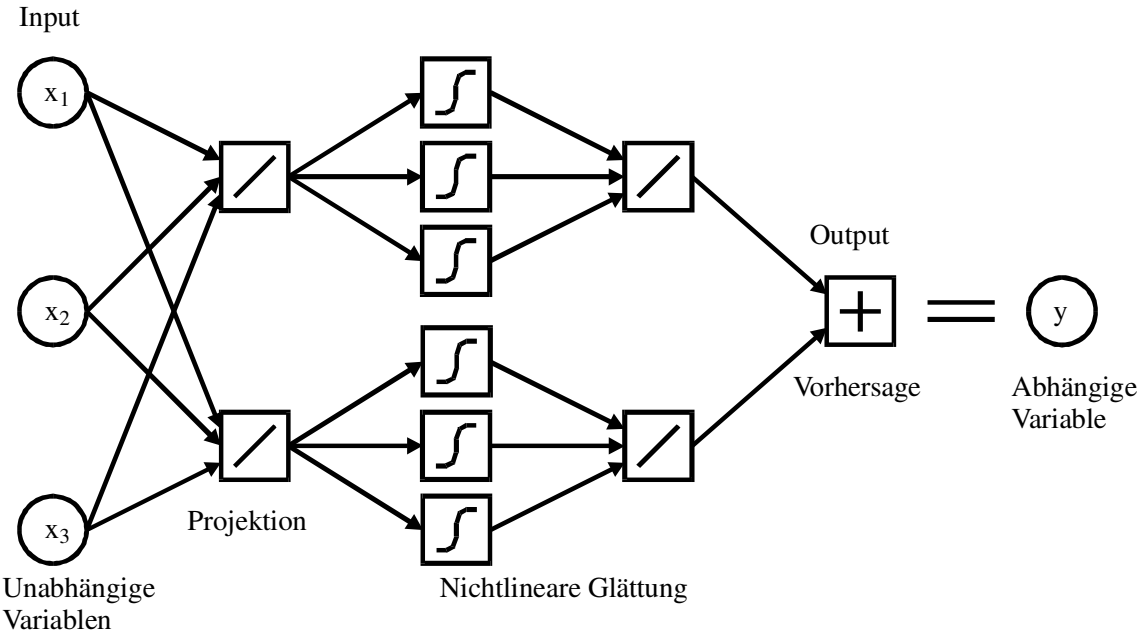


Nichtlineare Diskriminanzanalyse

© Ulrich Anders

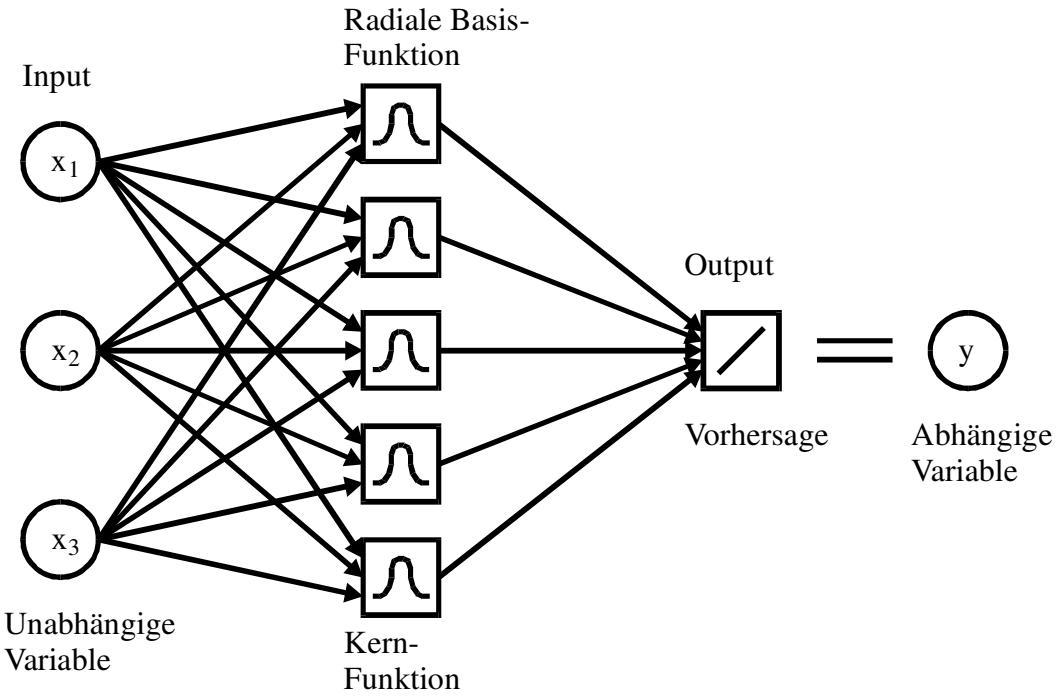


„Genestetes“ lineares Regressionsmodell



Projection Pursuit–Regression

© Ulrich Anders



Kern–Regression mit zusätzlichen Freiheiten

© Ulrich Anders

Asymptotik neuronaler Netzwerke (I)

Intention der Vorgehensweise:

1. Sind die erlernten Gewichte $\hat{w}$ (global) identifiziert, dann ist $\hat{w}$ ein konsistenter Schätzer für w (White, 1992).

2. Errechne asymptotische Kovarianzmatrix $\hat{C}$ von $\hat{w}$. $\hat{C}$ ist normal, wenn 1. gilt, ansonsten gehört $\hat{C}$ zu LMG-Familie (Phillips, 1989).

3. Führe Tests durch, z.B. Wald-Test:

   $$(R\hat{w} - r)'(R\hat{C}R')^{-1}(R\hat{w} - r) \rightarrow \chi^2_q$$

   insbesondere interessant
   - Irrelevante *input unit*
   - Irrelevante *hidden unit*

4. Schätzung von $\hat{C}$ durch QMLE–Informationssandwich

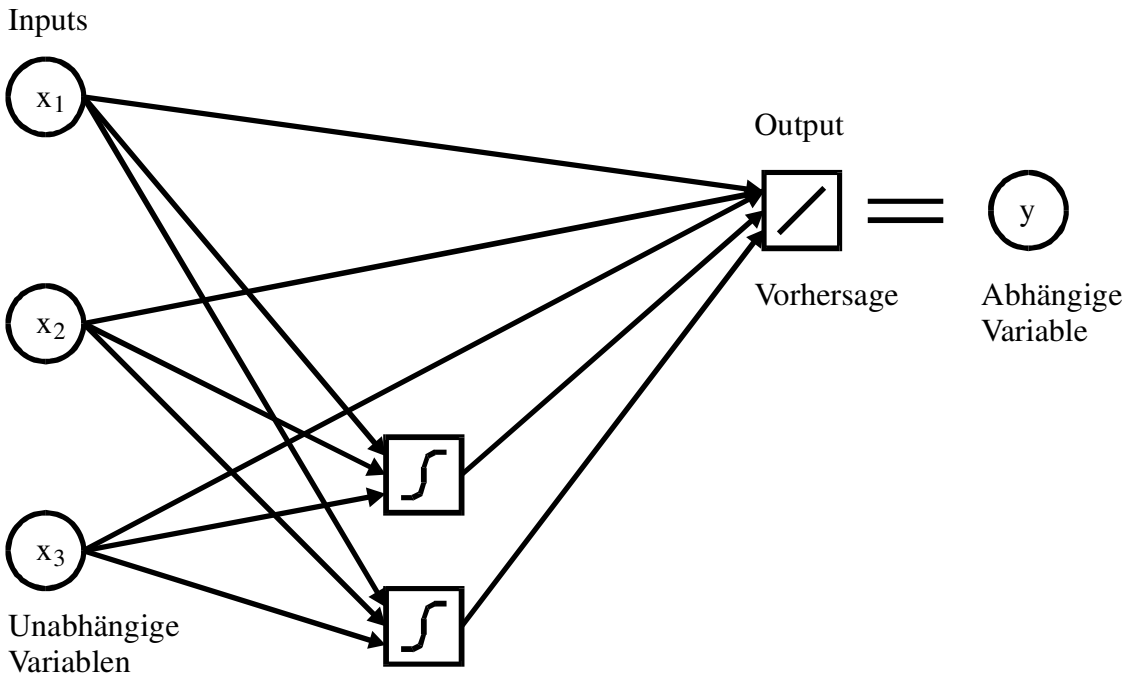
   $\hat{C} = \frac{1}{T}\hat{A}^{-1}\hat{B}\hat{A}^{-1}$, wobei:

$$A = -E\left[\frac{\partial^2 \ln L(w)}{\partial w \partial w'}\right] \qquad B = E\left[\left(\frac{\partial \ln L(w)}{\partial w}\right)\left(\frac{\partial \ln L(w)}{\partial w}\right)'\right]$$

Hypothesentests eignen sich als Nichtlinearitätstests und für die Bestimmung einer geeigneten Netzwerkarchitektur.

© Ulrich Anders

Asymptotik neuronaler Netzwerke (II)

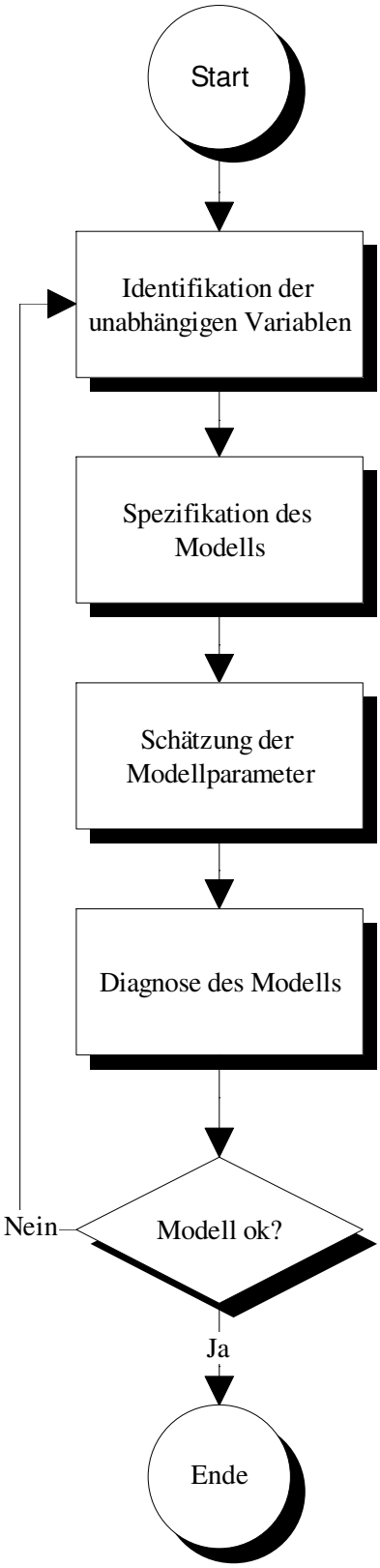


Problem: Identifikation

- Symmetrien (lokale Identifikation ausreichend).
- *nuisance parameter* (Davies, 1977; Davies, 1978)
  Lösungen: White (1989) und Teräsvirta (1992).

© Ulrich Anders

Modellbildung mit Neuronalen Netzwerken



Ökonomische Fundierung
Grafische Analyse
Integration/Kointegration
Multikollinearitäten
ommitted variables
Normierung

Daumenregeln
IC, R2, PE
Regularisierung
Pruning
Cross Validation
Data Splitting

BFGS, DFP
konj. Gradientenabstieg
Newton-Raphson
Backpropagation

Residualanalyse
2-Stufen-Schätzer
Signifikanz/Sensitivität
Parametertests
Fehlermaße
Konfidenzintervalle
Robustheit
Bootstrapping

© Ulrich Anders

Identifikation

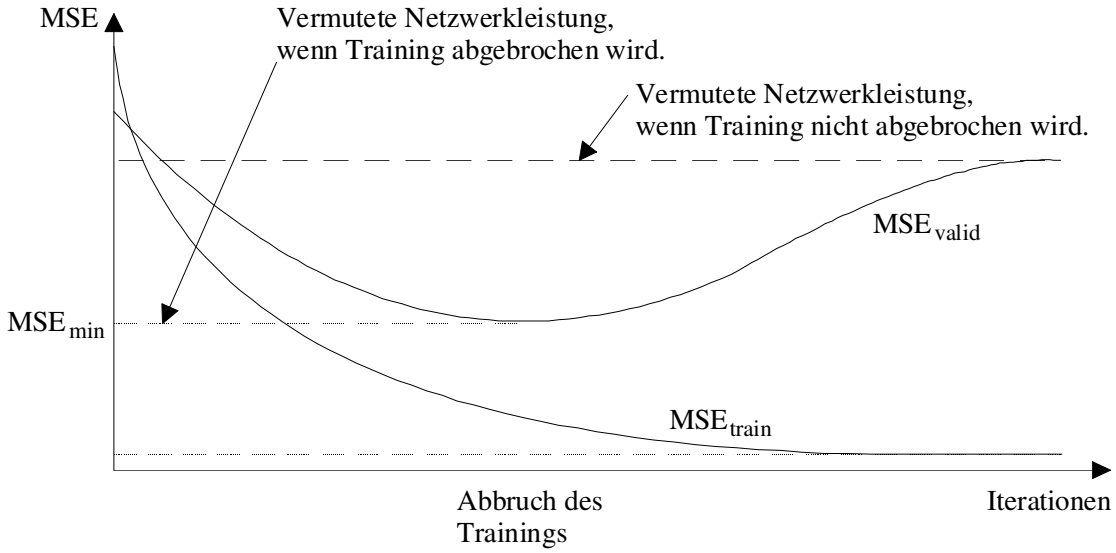
- Auswahl der Daten
- Integration/Kointegration
- Multikollinearitäten
- *Ommitted Variables*
- Normierung

Guter Start: Lineares Modell als Ausgangsbasis und Referenzmodell

© Ulrich Anders

Spezifikation der Netzwerkarchitektur

- Daumenregeln
- Informations- und Selektionskriterien (AIC, SIC, NIC, $\overline{R}^2$, PC, FPE, etc.)
- Regularisierung
- *Pruning*
- *Generalized Cross Validation*
- *Data Splitting (Stopped Training)*



Die Methode des *Stopped Training*

© Ulrich Anders

Parameterschätzung (I): Optimierungskriterien

- Regressionsmodell

  $y = f(X,w) + \varepsilon \qquad \varepsilon$ ist *i.i.d.* $\qquad \mathrm{Var}[\varepsilon] = \sigma \qquad E[\varepsilon] = 0$

- Nichtlineare kleinste Quadrate (NLS)

  $\mathrm{SSE}(w) = [y - f(X,w)]'[y - f(X,w)] \rightarrow \mathrm{Min!}$

- Maximum Likelihood (ML)

  allgemein:

  $$\ln L(\theta) = \sum_{t=1}^{T} P[\mathrm{E}_t = \varepsilon_t | \theta] \rightarrow \mathrm{Max!}$$

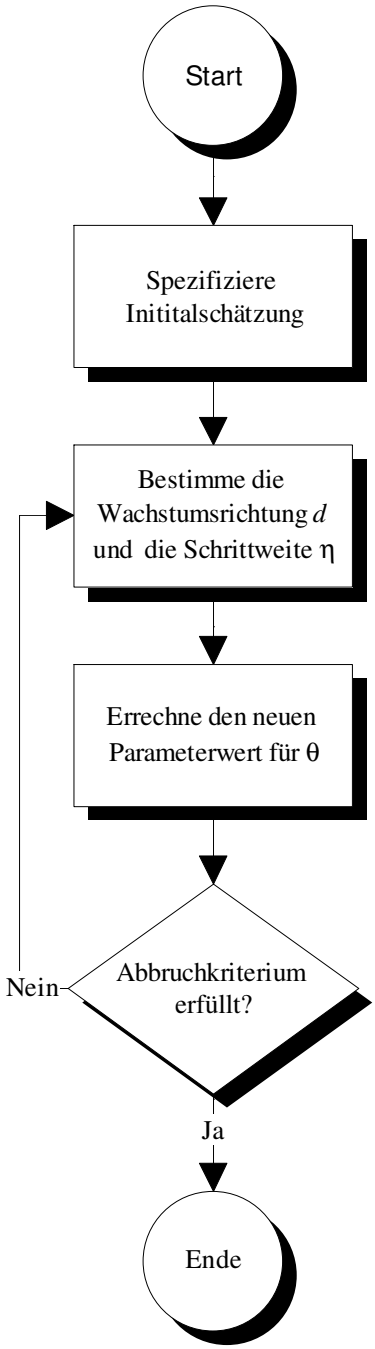
  bei Normalverteilung der Störterme:

  $$\ln L(w,\Omega) = -\ln |\Omega| - [y - f(X,w)]' \Omega^{-1} [y - f(X,w)] \rightarrow \mathrm{Max!}$$

- Auch bei Neuronalen Netzwerken wird eine Annahme über die Verteilung der Störterme getroffen!

© Ulrich Anders

Parameterschätzung (II): numerische Schätzverfahren



Verfahren 2. Ordnung:
Newton-Rhaphson
Quasi-Newton (DFP, BFGS)
Gauss-Newton
Scoring-Verfahren (BHHH)

Verfahren 1. Ordnung:
Steilster Abstieg
(Konjugierter) Grad.-abstieg
Backpropagation

Erweiterungen:
Levenberg-Marquardt
Line Search
Stabilisierung

Globale Verfahren:
Genetische Algorithmen
Simulated Annealing
Sintflut–Algorithmus

θ: Parameterwert
η: Schrittweite
d: Abstiegsrichtung
Q: Verfahrenstyp. Richtungsmatrix
$\dot{q}$: Gradient der Zielfunktion

Es gilt immer: $\theta_{(n+1)} = \theta_{(n)} - \eta_{(n)} \cdot d_{(n)}(\theta) = \theta_{(n)} - \eta_{(n)} \cdot Q_{(n)} \dot{q}_{(n)}$

© Ulrich Anders

Parameterschätzung (III)

- Kombination von Schätzalgorithmen
- Gradient und Hessematrix
- Monitoring

© Ulrich Anders

Diagnose

- Residualanalyse:

  1. Modell fehlspezifiziert
  2. Annahme $E[u'u] = \Omega \neq \sigma I$ verletzt

- Zwei-Stufen-Schätzung:

  1. $y = f(X, \hat{w}_{(1)}) + u_{(1)}$
  2. modelliere $u_{(1)}^2$ und schätze damit Ω
  3. $\text{SSE}(w_{(2)}) = [y - f(X, w_{(2)})]' \hat{\Omega}^{-1} [y - f(X, w_{(2)})] \rightarrow \text{Min!}$

- Relevanz- und Sensitivitätsanalyse

- Signifikanztests für Parameter (Wald, LR, LM)

- Prognosequalität im Erwartungswert
  (MSE, RMSE, Theils *U,* MAPE, etc.)

- Konfidenzintervalle

- Robustheit

- *Bootstrapping*

© Ulrich Anders

Zusammenfassung

- Neuronale Netzwerke stellen eine neue und flexible Klasse von statistischen Modellen dar.

- Neuronale Netzwerke benötigen keine Annahme über die Struktur des zu modellierenden Zusammenhangs.

- Neuronale Netze sind parametrische Verfahren mit einer nichtparametrischen Interpretation.

- Die Modellierung Neuronaler Netzwerke sollte in vier Schritten geschehen:

    1. Identifikation der Daten
    2. Spezifikation des Modells
    3. Schätzung der Modellparameter
    4. Diagnose des Modells

- Verfahren der Statistik und der Netzwerkliteratur kommen dabei gemeinsam zum Einsatz.

Umfangreiche Simulationsstudien sind notwendig.

© Ulrich Anders

Davies R.B. (1987): *Hypothesis Testing when a Nuisance Parameter is present only under the Alternative.* Biometrika, 74, 33-34.

Davies R.B. (197): *Hypothesis Testing when a Nuisance Parameter is present only under the Alternative.* Biometrika, 64, 247-254.

Lee T.-H., White H., Granger C.W. (1993): *Testing for Neglected Nonlinearity in Time Series Models.* Journal of Econometrics, 56, 269-290.

Murata N., Yoshizawa S., Amari S. (1994): *A Criterion for Determining the Number of Parameters in an Artificial Neural Network Model.* IEEE Trans. Neural Networks, Vol 5, No 6, 865-872.

Phillips P.C.B. (1989): *Partially Identified Econometric Modells.* Econometric Theory, Vol 5, 181-240.

Teräsvirta T., Lin C.-F., Granger C.W. (1993): *Power of the Neural Network Linearity Test.* Journal of Time Series Analysis, Vol 14, No 2, 209-220.

White H. (1992): *Artifical Neural Networks: Approximation and Learning Theory.* Blackwell.

White H. (1989): *An Additional Hidden Unit Test for Neglected Nonlinearity in Multilayer Feedforward Networks.* Proceedings of the International Joint Conference on Neural Networks, Washington, DC. San Diego: SOS Printing, II, 451-455.

© Ulrich Anders