



Regressions modeller

Hvad regresserer vi på og hvorfor?

Stockmarr, Anders

Publication date:
2012

Document Version
Også kaldet Forlagets PDF

[Link back to DTU Orbit](#)

Citation (APA):
Stockmarr, A. (2012). Regressions modeller: Hvad regresserer vi på og hvorfor? [2D/3D (Fysisk produkt)]., Copenhagen, Danmark, 06/11/2012

DTU Library

Technical Information Center of Denmark

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Regressions modeller – Hvad regresserer vi på og hvorfor?

Anders Stockmarr

Axelborg statistikgruppe

6/11 2012

Generel Regression

- $Y_t = f(X_t) + \varepsilon_t, t=1, \dots, n$
- f er en UKENDT funktion, der beskriver relationen mellem den **uafhængige** variabel X , og den **afhængige** variabel Y .
- Vi vil gerne afdække hvorledes X og Y er relateret (dvs. undersøge egenskaber ved f), gennem analyser af parrede observationer $(X_t, Y_t)_{t=1, \dots, n}$

Lineær Regression

- Antag at $f \in C^\infty(\mathbb{R})$
- Så er

$$f(x) = \sum_{n=0}^{\infty} a_n x^n,$$

hvor $a_n = f^{(n)}(0) / n!$.

- Dermed er

$$f(x) = a_0 + a_1 x + o(x),$$

hvor $o(x) = x\varepsilon(x)$.

Lineær Regression – funktioner af flere variable

- 1. ordens Taylor-udvikling: I modeltermer erstattes f i

$$Y_i = f(X_i) + \varepsilon_i, \quad i=1, \dots, n$$

Med

$$f(x) = f(0) + \sum_i \frac{\partial f}{\partial x_i}(0)x_i + o(\|x\|) \stackrel{def}{=} a_0 + \sum_i a_{1,i}x_i + o(\|x\|)$$

hvor restleddet sættes til nul, og hvor a_0 , a_1 er ukendte, dvs. den multiple regressionsmodel

$$Y_t = \alpha + \beta_1 X_{1,t} + \dots + \beta_k X_{k,t} + \varepsilon_t, \quad t=1, \dots, n.$$

Lineær Regression –funktioner af flere variable

- 2. ordens Taylor-udvikling:

$$f(x) = f(0) + \sum_i \frac{\partial f}{\partial x_i}(0)x_i + \frac{1}{2} \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(0)x_i x_j + o(\|x\|^2)$$

- I modeltermer erstattes

$$Y_t = f(X_t) + \varepsilon_t, t=1, \dots, n$$

Med

$$Y_t = \alpha + \sum_i \beta_i X_{i,t} + \sum_{i,j} \beta_{i,j} X_{i,t} X_{j,t} + \varepsilon_t, t = 1, \dots, n.$$

Lineær Regression –funktioner af flere variable

- 2. ordens Taylor-udvikling:

$$f(x) = f(0) + \sum_i \frac{\partial f}{\partial x_i}(0)x_i + \frac{1}{2} \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(0)x_i x_j + o(\|x\|^2)$$

- I modeltermer erstattes

$$Y_t = f(X_t) + \varepsilon_t, t=1, \dots, n$$

Med

$$Y_t = \alpha + \sum_i \beta_i X_{i,t} + \sum_{i,j} \beta_{i,j} X_{i,t} X_{j,t} + \varepsilon_t, t = 1, \dots, n.$$

Lineær regressionsmodel med 1. ordens interaktioner (og kvadratiske effekter; udelades ofte).

Polynomiel regression af højere orden

- Princip:
- Ynøjagtigheden $o(\|x\|^n)$ mindskes til en pris af introduktion af flere forklarende variable;
- Når ynøjagtigheden $o(\|x\|^n)$ er af en størrelsesorden så den kan inkorporeres i residualvariansen ε , er modellen tilstrækkelig i modelleringsforstand.

Polynomiel regression af højere orden

- Princip:
- Ynøjagtigheden $o(\|x\|^n)$ mindskes til en pris af introduktion af flere forklarende variable;
- Når ynøjagtigheden $o(\|x\|^n)$ er af en størrelsesorden så den kan inkorporeres i residualvariansen ε , er modellen tilstrækkelig i modelleringsforstand.
- Det kan i praksis kræve mange led:
- Eksempel $f(x) = \exp(-x)$

Skalering

- Højere ordens regression ønsker vi ikke; meget vanskeligt at fortolke og kommunikere.
- Løsningen er data-transformation.
- Vi anstrenger os en del for at finde skalaer, hvor sammenhængen kan beskrives med en Taylor-approximation af lav orden;

”sammenhængen er approksimativt lineær”

- log-transformation,
Box-Cox transformation,
kvadratrods-transformation, etc.

Agenda

- Vi vil gerne erstatte ukendte funktioner med andre ukendte, som dog har en kendt struktur; polynomier.
- Formålet er selvfølgelig, som al modellering, at forenkle virkeligheden så man kan regne på den uden at begå for grove fejl.
- Men samtidigt skal vi også gerne kunne se og kommunikere logikken i vores approksimation, så den må ikke være for kompliceret.
- **Subjektiv konklusion:**
- Vi bør approksimere med en Taylor-udvikling der er af 1. eller 2. orden, nogen gange 3. orden og aldrig over 4. Data skal skaleres, så dette kan lade sig gøre.

Ortogonalisering

Modellen

$$Y_t = \alpha + \sum_i \beta_i X_{i,t} + \sum_{i,j} \beta_{i,j} X_{i,t} X_{j,t} + \varepsilon_t, t = 1, \dots, n.$$

Er af formen

$$Y_t = \alpha + \sum_i \beta_i X_{i,t} + \varepsilon_t, t = 1, \dots, n,$$

hvor vi blot lader $X_i X_j$ optræde som en selvstændig kovariat. Dette gør den lineære regressionsmodel meget generel.

Ortogonalisering II

I modellen

Som vi skriver
$$Y_t = \alpha + \sum_i \beta_i X_{i,t} + \varepsilon_t, t = 1, \dots, n,$$

$$Y_t = \beta^T X_t + \varepsilon_t, t = 1, \dots, n$$

på sædvanlig vis, benytter vi ML/LS/PE-estimatoren

$$\hat{\beta} = (A^T A)^{-1} A^T Y,$$

hvor A er matricen $A = (X_1 : \dots : X_k)$ bestående af søjlerne med værdierne for de enkelte kovariater.

Ortogonalisering II

Med normalfordelte støjled er $\hat{\beta}$ normalfordelt;

$$\hat{\beta} \sim N(\beta, \sigma^2 (A^T A)^{-1}).$$

Men nu er

$$A^T A = \begin{pmatrix} \|X_1\|^2 & \langle X_1, X_2 \rangle & \cdots & \langle X_1, X_k \rangle \\ \langle X_1, X_2 \rangle & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \langle X_1, X_k \rangle & \cdots & \cdots & \|X_k\|^2 \end{pmatrix}$$

hvorfor $\hat{\beta}_i$ uafh. af $\hat{\beta}_j$ hviss $\langle X_i, X_j \rangle = 0$.

Ortogonalisering III

Modellen

$$Y_t = \beta^T X_t + \varepsilon_t, t = 1, \dots, n$$

udtrykker jo blot at Y på nær støj er en linear-kombination af søjlerne i matricen A .

MAO: Finder vi en anden måde at udtrykke linearkombinationer af søjlerne i A , ændrer vi ikke modellen.

Ortogonalisering IV

- Ønsker vi stokastisk uafhængige estimater, kan vi derfor lave en ny design matrix B , således at søjlerne i B er ortogonale, og således at søjlerne i B og A udspænder det samme rum.

Ortogonalisering V

- Dette gøres rekursivt:

$$B_1 = A_1;$$

$$B_i = A_i - \sum_{j=1}^{i-1} \frac{\langle A_i, B_j \rangle}{\|B_j\|^2} B_j$$

Ortogonalisering VI

- Eksempel:

$$Y_t = \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + \varepsilon_t, t = 1, \dots, n$$

$$A_1 = \underline{1}, A_2 = x_t, A_3 = x_t^2 :$$

$$B_1 = \underline{1},$$

$$B_2 = x_t - \bar{x},$$

$$B_3 = x_t^2 - \bar{x}^2 - \frac{SPD_{x,x^2}}{SSD_x} (x_t - \bar{x}).$$

Ortogonalisering VII

- I modellen

$$Y_t = \gamma_0 + \gamma_1(x_t - \bar{x}) + \gamma_2(x_t^2 - \bar{x}^2 - \frac{SPD_{x,x^2}}{SSD_x}(x_t - \bar{x})) + \varepsilon_t,$$

$$t = 1, \dots, n$$

er estimerne derfor stokastisk uafhængige.

Tilbageregning:

$$\beta_2 = \gamma_2;$$

$$\beta_1 = \gamma_1 - \gamma_2 \frac{SPD_{x,x^2}}{SSD_x};$$

$$\beta_0 = \gamma_0 - \gamma_1 \bar{x} + \gamma_2 \left(\frac{SPD_{x,x^2}}{SSD_x} \bar{x} - \bar{x}^2 \right)$$

Ortogonalisering VIII

- **Hvilke fordele ser I??**

Regression på andet end polynomier

- Grunden til at vi kan bruge polynomier er at polynomierne udgør en basis for $C^\infty(\mathbb{R})$, udstyret med topologien for uniform konvergens på kompakte mængder;
- Men man kan forestille sig situationer, hvor det er mere naturligt at forlange, at f tilhører en anden klasse end $C^\infty(\mathbb{R})$, og hvor man derfor skal kigge på andre baser.

Regression på andet end polynomier

II

Eksempel:

Periodiske funktioner (de-trendede sæson-data).

Her udgør funktionerne

$$h_n(x) = \sin\left(\frac{2n\pi}{\omega} x\right)$$

hvor ω er perioden, en basis for en passende gruppe af funktioner; man kan derfor modellere a la

$$Y_t = \alpha + \sum_i \beta_i \sin\left(\frac{2i\pi}{\omega} x_t\right) + \varepsilon_t, t = 1, \dots, n,$$

hvor disse sinusfunktioner kan ortogonaliseres ligesom tidligere.

Regression hvor den afhængige variabel er stokastisk

- En forudsætning for at estimaterne i modellen

$$Y_t = \beta^T X_t + \varepsilon_t, t = 1, \dots, n$$

- er uafhængige, er at design-matricen er en diagonal-matrix.
- Men en anden, implicit, forudsætning er, at X_t er deterministisk. Hvis X_t er stokastisk, er sagen generelt en anden.

Regression hvor den afhængige variabel er stokastisk II

- Antag at både X og Y er stokastiske variable, med en kausal relation imellem sig givet ved at

$$Y_t = \beta^T X_t + \varepsilon_t, t = 1, \dots, n$$

- Her hvis gælder f.eks. at $X_t \sim N(\mu, \sigma^2)$, er f.eks

$$E(X^2) = \mu^2 + \sigma^2,$$

og dermed er det ikke oplagt at sædvanlige polynomier er den fornuftigste vej at gå, hvis man f.eks. interesserer sig for hvad effekten af X er i termer af potenser af μ .

Regression hvor den afhængige variabel er stokastisk III

- Samtidig kan man interessere sig for en helt anden form for ortogonalitet; nemlig om de uafhængige variable som man regresserer på, er uafhængige, eller i det mindste ukorrelerede.
- Dette er en ganske anden ortogonalitet end geometrisk ortogonalitet af n observationer, altså ortogonalitet i \mathbb{R}^n .

Regression hvor den afhængige variabel er stokastisk IV

Hvis X er normeret normalfordelt, er 2 funktioner f og g af X ukorrelerede, hvis

$$\int_{-\infty}^{\infty} f(x)g(x)e^{-x^2/2}dx = 0$$

Definerer vi det indre produkt

$$\langle f, g \rangle_1 \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f(x)g(x)e^{-x^2/2}dx,$$

er dette kriterium præcis ortogonalitet i L_2 -forstand.

Regression hvor den afhængige variabel er stokastisk V

En følger af funktioner der opfylder dette, er Hermite polynomierne He_n , givet ved

$$He_0(x) = 1;$$

$$He_1(x) = x;$$

$$He_2(x) = x^2 - 1;$$

$$He_3(x) = x^3 - 3x;$$

⋮

$$He_n(x) = xHe_{n-1}(x) - (n-1)He_{n-2}(x).$$

Egenskaber ved Hermite-polynomierne I

- Hermite-polynomierne udgør en basis for vektorrummet

$$\{f : \mathbb{R} \rightarrow \mathbb{R} : E(|f(X)|^2) < \infty, X \sim N(0,1)\}$$

Dermed kan de fleste funktioner approksimeres med summer af Hermite polynomier.

Egenskaber ved Hermitepolynomierne II

$$\langle He_n, He_m \rangle_1 \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} He_n(x) He_m(x) e^{-x^2/2} dx = 0$$

hvis $n \neq m$, således at $He_n(X)$ og $He_m(X)$ er ukorrelerede, hvis X er normeret normal-fordelt.

Hvis $X \sim N(\mu, 1)$, er $E(He_n(x)) = \mu^n$.

Defineres $He_n^{[\sigma^2]}(x) \stackrel{\text{def}}{=} \sigma^{-n/2} He_n(x/\sigma)$, er $He_n^{[\sigma^2]}(X)$ og $He_m^{[\sigma^2]}(X)$ ortogonale/ukorrelerede for $n \neq m$, hvis X er normalfordelt $(0, \sigma^2)$.

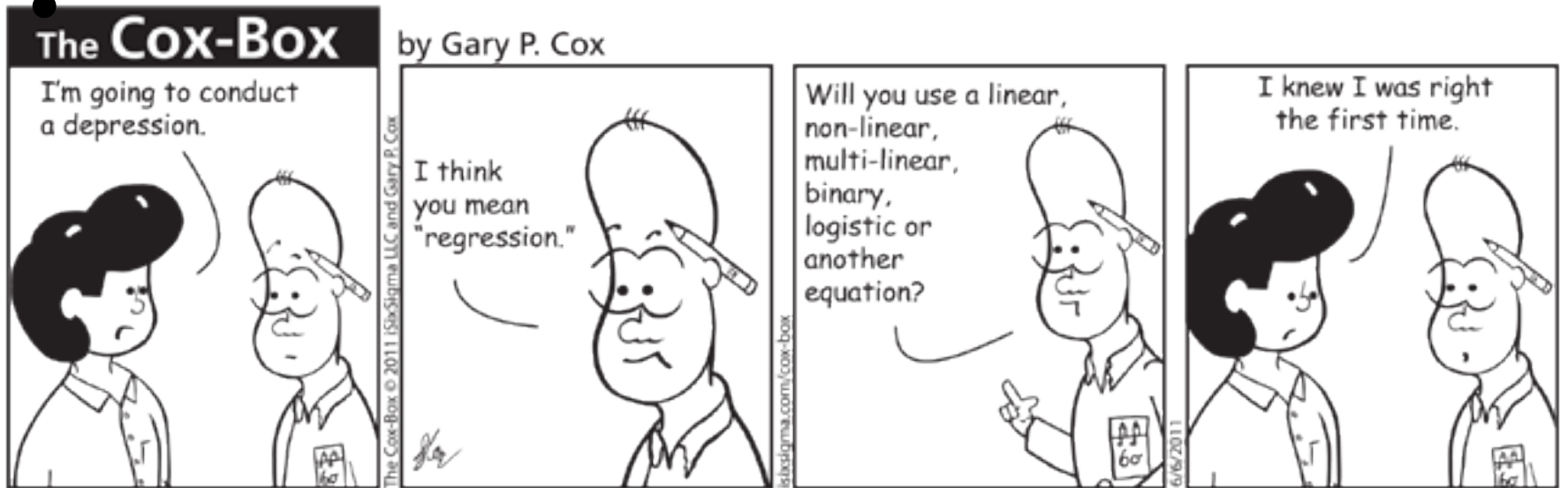
Egenskaber ved Hermite-polynomierne III

- Hermite-polynomier er altså skræddersyede til situationen, hvor man modellerer dynamiske systemer uden feedback mellem uafhængige og afhængige variable.
- Hermite polynomier har orden n , så Hermite polynomier op til orden n modellerer præcis også Taylorudvikling op til orden n (i 1 dimension).

Hermite-polynomier

- **Hvad er jeres erfaringer?**

Tak for opmærksomheden



Send comments and stories to Cox-Box@iSixSigma.com