

Određivanje širine h je jedini parametar KDE ocene gustine. Za Gausovo jezgro, pod pretpostavkom da je uzorak normalno raspodeljen, kao optimalna širina se uzima

$$h_{\text{opt}} = 1.06 \hat{\sigma} n^{-1/5}$$

gde je $\hat{\sigma}$ standardna devijacija uzorka. Sa ovom širinom i Gausovim jezgrom, KDE aproksimacija gustine raspodele za naš uzorak je data na slici 11.3.

11.2 Metoda najmanjih kvadrata

Pretpostavimo da imamo matematički model oblika

$$y(x) = y(x; a_1, \dots, a_M) = y(x; \mathbf{a})$$

gde M slobodnih parametara (konstanti) $\mathbf{a} = (a_1, \dots, a_M)$ ne znamo, a treba da ih nekako odredimo. U tom cilju korišćićemo N poznatih podataka (x_i, y_i) , $i = 1, \dots, N$ (obično dobijenih merenjima).

Postupak određivanja nepoznatih parametara u modelu na osnovu poznatih eksperimentalnih podataka se obično naziva fitovanjem. Ne postoji jednoznačan odgovor koji vektor parametara \mathbf{a} je najbolji. U tom cilju se obično definiše funkcija kvaliteta fitovanja, koja za date parametre \mathbf{a} daje numeričku vrednost kvaliteta modela u odnosu na podatke. Kao optimalne parametre ćemo uzeti one za koje ova funkcija uzima najbolju vrednost.

Metoda najmanjih kvadrata određuje optimalne parametre a_1, \dots, a_M tako da minimiziraju funkciju

$$\chi^2(\mathbf{a}) = \sum_{i=1}^N \left(\frac{y_i - y(x_i; a_1, \dots, a_M)}{\sigma_i} \right)^2 \quad (11.4)$$

gde su σ_i unapred poznate vrednosti. U slučaju da su sve te vrednosti jednake, $\sigma_i = \sigma$, $i = 1, \dots, M$, može se slobodno uzeti da je $\sigma = 1$, pošto optimalna vrednost \mathbf{a} ne zavisi od te vrednosti.

Pretpostavimo da je $y_i = y(x_i; \mathbf{a}) + \varepsilon_i$, gde su eksperimentalne greške ε_i normalno raspodeljene sa očekivanjem 0 i standardnom devijacijom σ_i . Pod tom pretpostavkom možemo za svaku vrednost parametara a_1, \dots, a_M izračunati verovatnoću da eksperimentalne vrednosti y_i uzimaju baš te vrednosti (odnosno pripadaju malom intervalu oko njih). Ocena maksimalne verodostojnosti kao optimalne parametre uzima one za koje je ta verovatnoća najveća. Pokazuje se da optimalni parametri upravo minimiziraju funkciju (11.4). Pod pretpostavkom normalno raspodeljenih grešaka, $\chi^2(\mathbf{a})$ ima χ^2 raspodelu sa $\nu = N - M$ stepeni slobode. Verovatnoća Q da χ^2 bude veće od

konkretno stačunate vrednosti $\chi^2(\mathbf{a})$ je data nekompletnom gama funkcijom $Q(\nu/2, \chi^2(\mathbf{a})/2)$ gde je

$$Q(a, x) = \frac{\Gamma(a, x)}{\Gamma(a)} = \frac{1}{\Gamma(a)} \int_x^\infty e^{-t} t^{a-1} dt, \quad (a > 0).$$

Vrednost Q daje meru kvaliteta fitovanja, tj. modela. Ukoliko je Q vrlo malo, uzrok može biti

- model je pogrešan, i može biti statistički odbačen,
- informacije o devijacijama σ_i su pogrešne i veće su nego što je navedeno,
- eksperimentalne greške nisu normalno raspodeljene.

Ukoliko je Q suviše veliko, blisko 1, uzrok može biti

- greške podataka su precenjene i stvarno su manje od navedenih,
- eksperimentator nije dobro obavio svoj posao, eksperimentalni podaci su sumnjivi.

Zbog činjenice da je Q malo ukoliko greške nisu normalno raspodeljene (recimo sistematske greške kod nekih nelinearnih modela), nije neobično da model prihvatimo i ako je $Q \sim 10^{-3}$, ali ipak treba biti oprezan. Stvarno pogrešni modeli daju mnogo manje vrednosti za Q .

Prema teoriji, za veliko N je χ^2 asimptotski normalno raspodeljeno sa očekivanjem ν i standardnom devijacijom $\sqrt{2\nu}$. Zato bi očekivane vrednosti za $\chi^2(\mathbf{a})$ trebale da budu $\chi^2(\mathbf{a}) \approx \nu$.

Ukoliko ne znamo veličine σ_i i ako pretpostavimo da su one jednake, $\sigma_i = \sigma$, možemo prvo u funkciji (11.4) zadati $\sigma_i = 1$, sračunati optimalne parametre \mathbf{a} koji minimiziraju tu funkciju, i naknadno sračunati

$$\sigma^2 = \frac{1}{N - M} \sum_{i=1}^N (y_i - y(x_i; \mathbf{a}))^2.$$

Kvalitet modela možemo verifikovati i statističkom analizom ostataka (razlika) $r_i = (y_i - y(x_i; \mathbf{a}))/\sigma_i$. Pravljenjem histograma ili KDE dijagrama za ostatke možemo videti da li su greške zaista normalno ili približno normalno raspodeljene. Kutijasti dijagram nam može pomoći u otkrivanju podataka sa velikim odstupanjima. Pošto je uticaj tih podataka u funkciji (11.4) značajan usled kvadratne zavisnosti, često se u postupku fitovanja

ovakve tačke eliminišu iz uzorka, i fitovanje ponovi bez njih. Motivacija za ovaj postupak je da je uzrok ovakvih grešaka drugačije prirode od normalnih eksperimentalnih grešaka. Takođe, ostaci r_i , kao slučajne veličine, treba da budu međusobno nezavisni. Međusobna nezavisnost ostataka se može statistički testirati autokorelacijom. Ukoliko se ispostavi da ostaci nisu međusobno nezavisni, to obično znači da model nije najbolje izabran. Do ovakve situacije se npr. dolazi ako kvadratnu funkciju fitujemo linearnom.

U zavisnosti od toga da li je funkcija $y(x; a_1, \dots, a_M)$ linearna po nepoznatim parametrima ili nije, razlikujemo linearne, odnosno nelinearne modele. Matematički aparat za dobijanje optimalnih parametara se u ta dva slučaja razlikuje.

11.2.1 Linearni model

U generalizovanom linearnom modelu je funkcija $y(x)$ linearna po parametrima a_1, \dots, a_M :

$$y(x) = \sum_{k=1}^M a_k X_k(x)$$

gde su $X_1(x), \dots, X_M(x)$ unapred zadate linearno nezavisne bazisne funkcije. Za fitovanje polinomima možemo uzeti $X_k(x) = x^{k-1}$ ili, bolje, ortogonalne polinome.

U skladu sa (11.4), treba minimizovati

$$\chi^2(\mathbf{a}) = \sum_{i=1}^N \left(\frac{y_i - \sum_{k=1}^M a_k X_k(x_i)}{\sigma_i} \right)^2. \quad (11.5)$$

Ukoliko su σ_i nepoznati ili međusobno jednaki, može se uzeti $\sigma_i = 1$.

Neka je A matrica dimenzije $N \times M$, i vektor b dužine N , sa elementima

$$A_{ij} = \frac{X_j(x_i)}{\sigma_i}, \quad b_i = \frac{y_i}{\sigma_i}.$$

U tački minimuma funkcije (11.5) su njeni parcijalni izvodi po a_k jednaki nuli, pa je

$$0 = \sum_{i=1}^N \frac{1}{\sigma_i^2} \left(y_i - \sum_{j=1}^M a_j X_j(x_i) \right) \cdot X_k(x_i), \quad k = 1, \dots, M,$$

što daje sistem linearnih jednačina

$$\sum_{j=1}^M \alpha_{kj} a_j = \beta_k \quad \text{ili} \quad \boldsymbol{\alpha} \mathbf{a} = \boldsymbol{\beta}$$

gde je

$$\alpha_{kj} = \sum_{i=1}^N \frac{X_j(x_i)X_k(x_i)}{\sigma_i^2}, \quad \beta_j = \sum_{i=1}^N \frac{y_i X_j(x_i)}{\sigma_i^2},$$

ili

$$\boldsymbol{\alpha} = A^T A, \quad \boldsymbol{\beta} = A^T \mathbf{b}.$$

Sistem $A^T A \mathbf{a} = A^T \mathbf{b}$ se naziva normalni sistem, i često je loše uslovljen, što predstavlja problem za nalaženje njegovog rešenja. Umesto standardnih metoda (Gausove ili metode Holeckog) tada je bolje koristiti iterativne metode (npr. Gaus-Zajdelovu metodu). Do istog rešenja dovodi i primena metode singularne dekompozicije (SVD) na pravougaoni sistem $A\mathbf{a} = \mathbf{b}$. Na uslovljenost matrice normalnog sistema bitno utiče izbor bazisnih funkcija $X_i(x)$, pa se pogodnijim izborom bazisa u istom prostoru može značajno popraviti uslovljenost sistema.

Inverzna matrica $C = \boldsymbol{\alpha}^{-1}$ je tesno povezana sa standardnim devijacijama fitovanih parametara a_j , a one daju meru stabilnosti parametara u odnosu na greške podataka. Pošto je $\mathbf{a} = C\boldsymbol{\beta}$, biće

$$a_j = \sum_{k=1}^M C_{jk} \left(\sum_{i=1}^N \frac{y_i X_k(x_i)}{\sigma_i^2} \right).$$

Varijansa parametra a_j se može izraziti kao

$$\sigma^2(a_j) = \sum_{i=1}^N \sigma_i^2 \left(\frac{\partial a_j}{\partial y_i} \right)^2.$$

Pošto α_{jk} a time i C_{jk} ne zavise od y_i

$$\frac{\partial a_j}{\partial y_i} = \sum_{k=1}^M \frac{C_{jk} X_k(x_i)}{\sigma_i^2},$$

pa je

$$\sigma^2(a_j) = \sum_{k=1}^M \sum_{l=1}^M C_{jk} C_{jl} \left(\sum_{i=1}^N \frac{X_k(x_i) X_l(x_i)}{\sigma_i^2} \right).$$

U zagradi su upravo elementi matrice $\boldsymbol{\alpha} = C^{-1}$, pa je

$$\sigma^2(a_j) = C_{jj}.$$

Elementi C_{jk} van dijagonale predstavljaju kovarijansu između a_j i a_k .

11.2.2 Nelinearni model

U slučaju modela koji nije linearan po parametrima

$$y = y(x, \mathbf{a}), \quad \mathbf{a} = (a_1, \dots, a_M)$$

traži se minimum funkcije

$$\begin{aligned} \chi^2(\mathbf{a}) &= \sum_{i=1}^N \left(\frac{y_i - y(x_i; \mathbf{a})}{\sigma_i} \right)^2 \\ &= \sum_{i=1}^N r_i^2(\mathbf{a}) \\ &= R(\mathbf{a})^T R(\mathbf{a}) \end{aligned} \tag{11.6}$$

gde smo označili $r_i(\mathbf{a}) = (y_i - y(x_i; \mathbf{a}))/\sigma_i$ i $R(\mathbf{a}) = (r_1(\mathbf{a}), \dots, r_N(\mathbf{a}))^T$. Zbog nelinearnosti, ova funkcija može imati više lokalnih minimuma, pa je ponekad nalaženje optimalnih parametara težak zadatak i moraju se primeniti metode globalne optimizacije. U svakom lokalnom, pa i globalnom minimumu, gradijent funkcije $\chi^2(\mathbf{a})$ mora biti jednak nuli. Označimo

$$J(\mathbf{a}) = \left[\frac{\partial r_i}{\partial a_j} \right], \quad i = 1, \dots, N, \quad j = 1, \dots, M.$$

Gradijent funkcije koju minimiziramo ima elemente

$$\begin{aligned} \frac{\partial \chi^2}{\partial a_k} &= -2 \sum_{i=1}^N \frac{y_i - y(x_i; \mathbf{a})}{\sigma_i^2} \frac{\partial y(x_i; \mathbf{a})}{\partial a_k} \\ &= 2 \sum_{i=1}^N r_i \frac{\partial r_i}{\partial a_k} \end{aligned}$$

pa je

$$\nabla \chi^2(\mathbf{a}) = 2J(\mathbf{a})^T R(\mathbf{a})$$

i

$$\frac{\partial^2 \chi^2}{\partial a_k \partial a_j} = 2 \sum_{i=1}^N \frac{1}{\sigma_i^2} \left[\frac{\partial y(x_i; \mathbf{a})}{\partial a_k} \frac{\partial y(x_i; \mathbf{a})}{\partial a_j} - (y_i - y(x_i; \mathbf{a})) \frac{\partial^2 y(x_i; \mathbf{a})}{\partial a_k \partial a_j} \right]$$

odakle sledi da je Hesijan funkcije jednak

$$\nabla^2 \chi^2(\mathbf{a}) = 2 \left(J(\mathbf{a})^T J(\mathbf{a}) + S(\mathbf{a}) \right), \quad S(\mathbf{a}) \equiv \sum_{i=1}^N r_i(\mathbf{a}) \nabla^2 r_i(\mathbf{a})$$

Kvadratna aproksimacija funkcije $\chi^2(\mathbf{a})$ je data sa

$$\begin{aligned}\chi^2(\mathbf{a} + \mathbf{a}_\delta) &= \chi^2(\mathbf{a}) + \nabla \chi^2(\mathbf{a}) \cdot \mathbf{a}_\delta + \frac{1}{2} \mathbf{a}_\delta^T \cdot \nabla^2 \chi^2(\mathbf{a}) \cdot \mathbf{a}_\delta \\ &= R(\mathbf{a})^T R(\mathbf{a}) + 2R(\mathbf{a})^T J(\mathbf{a}) \mathbf{a}_\delta + \\ &\quad + \mathbf{a}_\delta^T (J(\mathbf{a})^T J(\mathbf{a}) + S(\mathbf{a})) \mathbf{a}_\delta\end{aligned}$$

Ukoliko je funkcija $y(x; \mathbf{a})$ dovoljno glatka, za minimizaciju funkcije (11.6) se mogu koristiti kako gradijentne, tako i metode drugog reda tačnosti. Sledi kraći pregled metoda koje se koriste u praksi.

Njutnova metoda

Primena Njutnove metode za minimizaciju funkcije daje

$$\mathbf{a}_\delta = - (J(\mathbf{a})^T J(\mathbf{a}) + S(\mathbf{a}))^{-1} J(\mathbf{a})^T R(\mathbf{a}).$$

I pored brze konvergencije u okolini rešenja, metodu komplikuje potreba računanja drugih izvoda od kojih zavisi $S(\mathbf{a})$.

Gaus-Njutnova metoda

Za dovoljno malo \mathbf{a}_δ aproksimiraćemo linearizacijom $R(\mathbf{a} + \mathbf{a}_\delta) = R(\mathbf{a}) + J(\mathbf{a})\mathbf{a}_\delta$, što za posledicu ima da je $\nabla^2 r_i(\mathbf{a}) = 0$, pa je $S(\mathbf{a}) \equiv 0$ i

$$\mathbf{a}_\delta = - (J(\mathbf{a})^T J(\mathbf{a}))^{-1} J(\mathbf{a})^T R(\mathbf{a}),$$

tj. \mathbf{a}_δ je rešenje linearnog sistema

$$(J(\mathbf{a})^T J(\mathbf{a})) \mathbf{a}_\delta = J(\mathbf{a})^T R(\mathbf{a})$$

ili

$$\alpha \mathbf{a}_\delta = \beta$$

gde je

$$\begin{aligned}\alpha_{kl} &\equiv \frac{1}{2} \frac{\partial^2 \chi^2}{\partial a_k \partial a_l} = \sum_{i=1}^N \frac{1}{\sigma_i^2} \frac{\partial y(x_i; \mathbf{a})}{\partial a_k} \frac{\partial y(x_i; \mathbf{a})}{\partial a_l}, \\ \beta_k &\equiv -\frac{1}{2} \frac{\partial \chi^2}{\partial a_k} = \sum_{i=1}^N \frac{y_i - y(x_i; \mathbf{a})}{\sigma_i^2} \frac{\partial y(x_i; \mathbf{a})}{\partial a_k}.\end{aligned}$$

Matrica α se u slučaju da je $y(x; \mathbf{a})$ linearna po parametrima poklapa sa istom matricom iz prethodnog odeljka, pa matrica $C = \alpha^{-1}$ predstavlja matricu varijansi i kovarijansi parametara a_1, \dots, a_M .

Gaus-Njutnova metoda je jednostavnija za primenu jer ne zahteva računanje drugih izvoda. Opravdanje za aproksimaciju koja je u njoj primenjena je u činjenici da će

$$S(\mathbf{a}) \equiv \sum_{i=1}^N r_i(\mathbf{a}) \nabla^2 r_i(\mathbf{a})$$

u slučaju nezavisnih i slučajno raspodeljenih grešaka r_i biti malo usled međusobnog potiranja članova u sumi. Prednosti Gaus-Njutnove metode su

- lokalna kvadratna konvergencija u zadacima sa nultom optimalnom greškom,
- brza lokalna konvergencija za zadatke koji nisu mnogo nelinearni ili koji imaju male lokalne greške r_i ,

a mane su

- slaba konvergencija za izrazito nelinearne modele i za velike optimalne greške r_i ,
- odsustvo konvergencije u nekim slučajevima iz prethodne tačke,
- nije korektno definisana kada $J(\mathbf{a})$ nema puni rang,
- ne garantuje se globalna konvergencija.

Dampirana Gaus-Njutnova metoda

U ovoj varijanti prethodne metode je

$$\mathbf{a} + \mathbf{a}_\delta = \mathbf{a} - \lambda (J(\mathbf{a})^T J(\mathbf{a}))^{-1} J(\mathbf{a})^T R(\mathbf{a})$$

gde se λ bira po kriterijumu jednodimenzione optimizacije. Parametrom λ izbegavamo prevelike vrednosti za \mathbf{a}_δ , a metoda je korektno definisana i ako $J(\mathbf{a})$ nema puni rang.

Levenberg-Markartova metoda

U Levenberg-Markartovoj metodi (Levenberg 1944., Marquardt 1963.) se nova iteracija u optimizaciji dobija po formuli

$$\mathbf{a} + \mathbf{a}_\delta = \mathbf{a} - (J(\mathbf{a})^T J(\mathbf{a}) + \lambda I)^{-1} J(\mathbf{a})^T R(\mathbf{a})$$

gde je $\lambda \geq 0$ parametar metode, a I jedinična matrica. U graničnom slučaju za $\lambda = 0$ se svodi na Gaus-Njutnovu metodu, a za velike vrednosti λ se vektor \mathbf{a}_δ po pravcu asimptotski približava antigradijentu sa sve manjom dužinom, što garantuje uspešnost koraka za dovoljno veliko λ .

Druga interpretacija metode je da je ona teorijski ekvivalentna zadatku

$$\min_{\|\mathbf{a}_\delta\| \leq \varepsilon} \|R(\mathbf{a}) + J(\mathbf{a}) \mathbf{a}_\delta\|_{L_2}$$

gde je veličina ε povezana sa parametrom λ (tačnije $1/\lambda$).

U ovoj metodi se, koristeći oznake od ranije, \mathbf{a}_δ dobija kao rešenje sistema

$$(\boldsymbol{\alpha} + \lambda I) \mathbf{a}_\delta = \boldsymbol{\beta}. \quad (11.7)$$

Za pozitivno λ je matrica sistema pozitivno definitna i povećavanjem ovog parametra poboljšavamo uslovljenost sistema. Variranjem parametra obezbeđujemo kompromis između brze konvergencije i stabilnosti metode.

Algoritam metode se sastoji od sledećih koraka

1. Izabrati početnu aproksimaciju \mathbf{a} .
2. Sračunati $\chi^2(\mathbf{a})$.
3. Postaviti početnu vrednost $\lambda = 0.001$.
4. Proveriti kriterijum za zaustavljanje algoritma.
5. Rešiti sistem (11.7) i sračunati $\chi^2(\mathbf{a} + \mathbf{a}_\delta)$.
6. Ako je $\chi^2(\mathbf{a} + \mathbf{a}_\delta) \geq \chi^2(\mathbf{a})$, povećati λ (npr. 10 puta) i ići na korak 4.
7. Ako je $\chi^2(\mathbf{a} + \mathbf{a}_\delta) < \chi^2(\mathbf{a})$, smanjiti λ (npr. 10 puta), prihvatiti novu aproksimaciju $\mathbf{a} + \mathbf{a}_\delta$, i ići na korak 4.

Kriterijum za zaustavljanje obično kombinuje sledeće kriterijume: maksimalan dozvoljen broj iteracija, $\|\mathbf{a}_\delta\| \leq \varepsilon$, $|\chi^2(\mathbf{a} + \mathbf{a}_\delta) - \chi^2(\mathbf{a})| \leq \varepsilon$.

Po završetku algoritma, invertovanjem matrice sistema $\boldsymbol{\alpha}$ (za $\lambda = 0$), dobijamo matricu C varijansi i kovarijansi parametara a_1, \dots, a_M .

Ovde je metoda izložena u prvobitnoj, Levenbergovoj formi, a Markart je kasnije modifikovao metodu tako što zamenom jedinične matrice I dijagonalnom $\text{diag}(J(\mathbf{a})^T J(\mathbf{a}))$ ostvaruje poboljšanu stabilnost za velike vrednosti parametra λ .

11.3 Robusno fitovanje

Metodu za procenu parametara modela smatramo robusnom ukoliko mali poremećaji vrednosti eksperimentalnih tačaka, ili veći poremećaji na manjem broju tačaka, rezultuju malim razlikama u procenjenim parametrima.

Metoda najmanjih kvadrata pretpostavlja da su greške normalno raspodeljene. Kod normalne raspodele velike vrednosti su vrlo malo verovatne. Velike greške pojedinih tačaka bitno utiču na vrednost funkcije χ^2 zbog kvadratne zavisnosti u njoj, i rezultat fitovanja usled toga može biti nezadovoljavajući.

Metode ocene (fitovanja) parametara metodom maksimalne verodostojnosti se, polazeći od pretpostavljene raspodele eksperimentalnih grešaka, svode na minimizaciju

$$\min_{(a_1, \dots, a_M)} \sum_{i=1}^N \rho \left(\frac{y_i - y(x_i; \mathbf{a})}{\sigma_i} \right)$$

ili

$$\min_{(a_1, \dots, a_M)} \sum_{i=1}^N \rho(r_i), \quad r_i = \frac{y_i - y(x_i; \mathbf{a})}{\sigma_i},$$

za neku funkciju $\rho(r)$.

Za normalnu raspodelu grešaka sa normalnom gustinom raspodele oblika $C_1 \exp(-C_2 r^2)$ metoda maksimalne verodostojnosti daje $\rho(r) = r^2/2$, i to predstavlja metodu najmanjih kvadrata.

U slučaju da greške nisu normalno raspodeljene, već imaju dvostranu eksponencijalnu raspodelu sa gustinom oblika $C_1 \exp(-|r|)$, tada se dobija da je $\rho(r) = |r|$. Treba primetiti da u ovom slučaju $\rho(r)$ uzima manje vrednosti za veliko r nego u prethodnom slučaju, pa zato veće greške u ovom slučaju imaju manji uticaj na ukupnu funkciju greške koja se minimizira. Sa druge strane, ova funkcija nije glatka, pa se ne mogu koristiti gradijentne i metode višeg reda za minimizaciju.

Za Košijevu (ili Lorencovu) raspodelu sa gustinom oblika $C/(1 + r^2/2)$ se dobija da je $\rho(r) = \ln(1 + r^2/2)$. U ovom slučaju veće greške imaju još manji uticaj.

Nevezano za metodu maksimalne verodostojnosti, robusno fitovanje možemo izvesti pogodnim izborom funkcije $\rho(r)$. Poželjno je da za veliko r ova funkcija ne uzima velike vrednosti, čime se neutrališe preveliki uticaj velikih grešaka u podacima. Sa druge strane, ako su greške približno normalno raspodeljene, poželjno je da ova funkcija za male vrednosti r liči

na kvadratnu. Jedna od takvih funkcija (Andrew's sine) koja se koristi za robusno fitovanje je

$$\rho(r) = \begin{cases} C(1 - \cos(r/C)), & |r| \leq C\pi, \\ C, & |r| > C\pi. \end{cases}$$

Ako su greške normalno raspodeljene, najbolje je uzeti da je $C = 2.1$. Druga pogodna funkcija za robusno fitovanje (Tuckey's biweight) je

$$\rho(r) = \begin{cases} \frac{C^2}{6} \left[1 - \left(1 - \frac{r^2}{C^2} \right)^3 \right], & |r| \leq C, \\ \frac{C^2}{6}, & |r| > C. \end{cases}$$

Za normalno raspodeljene greške je najbolje uzeti $C = 6$.

11.4 Pouzdanost dobijenih parametara

Kada smo u modelu odredili optimalne parametre (a_1, \dots, a_M) , postavlja se pitanje koliko su oni osetljivi na male varijacije ulaznih podataka. Naime, ukoliko su eksperimentalne greške isto raspodeljene, ali su drugačije (recimo kod drugog kompleta merenja), izračunati parametri će se malo razlikovati. Nas interesuje mera njihove promene, a nju dobro opisuje standardna devijacija parametara, kao i intervali poverenja koji se dobijaju ako je poznata raspodela parametara.

U slučaju metode najmanjih kvadrata i linearnog modela, pod pretpostavkom nezavisnih normalno raspodeljenih grešaka, teorijski rezultat je da su standardne devijacije parametara dijagonalni elementi matrice kovarijansi $C = \alpha^{-1}$. Ova procena se koristi i kod nelinearnih modela. Međutim, ukoliko greške nemaju normalnu raspodelu, ukoliko nisu nezavisne, ili ako koristimo metode robusnog fitovanja, ovaj rezultat ne smemo koristiti. Na koji način ipak proceniti osetljivost dobijenih parametara?

Komplet podataka koji imamo označimo sa $D_{(0)}$, a izračunate parametre sa $\mathbf{a}_{(0)}$. Da smo imali dodatne kompletne podataka $D_{(k)}$, $k = 1, \dots, K$ iste veličine i sa isto raspodeljenim slučajnim greškama, oni bi dali za rezultate $\mathbf{a}_{(k)}$, $k = 1, \dots, K$. Osetljivost dobijenih parametara se onda svodi na statističku obradu rezultata $\mathbf{a}_{(k)}$. Međutim, mi takve izvorne kompletne podataka nemamo, ali ih možemo simulirati na sledeća dva načina.

Monte Carlo simulacija podataka

Iako znamo da parametri $\mathbf{a}_{(0)}$ nisu pravi, već imaju neku grešku, pretpostavićemo za potrebe simulacije podataka da su to baš pravi parametri.

Sa njima ćemo generisati nove kolekcije podataka iste dužine preko

$$y_i = y(x_i; \mathbf{a}_{(0)}) + \varepsilon_i$$

gde slučajne veličine ε_i imaju istu raspodelu kao u izvornim podacima. Na taj način dobijamo kompletne podatke $D_{(k)}$, $k = 1, \dots, K$ koji daju optimalne parametre $\mathbf{a}_{(k)}$, $k = 1, \dots, K$.

Bootstrep metoda

Bootstrep metoda je još jednostavnija za primenu. U njoj ne generišemo nikakve nove podatke, već nove kompletne pravimo od istog polaznog kompleta podataka. Pošto nije pametno polazni komplet deliti na više delova jer se tako smanjuje broj tačaka u kompletu, postupićemo na sledeći način. Iz polaznog kompleta podataka na slučajan način biramo (sa vraćanjem) onoliki broj podataka koliko je sadržao polazni komplet. Na ovaj način će se neki podaci uzeti više puta, a neki će biti izostavljeni. Ponavljanjem ovog postupka dobijamo kompletne podatke $D_{(k)}$, $k = 1, \dots, K$, i za njih sračunate optimalne parametre $\mathbf{a}_{(k)}$, $k = 1, \dots, K$.

11.4.1 Višedimenziona normalna raspodela

U ovom poglavlju ćemo izložiti definiciju i osnovne osobine višedimenzionalne normalne raspodele.

Jednodimenziona normalna raspodela slučajne promenljive X sa matematičkim očekivanjem μ i varijansom σ^2 ima gustinu raspodele

$$g(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-[(x-\mu)/\sigma]^2/2}$$

Ukoliko imamo uzorak (x_1, \dots, x_N) koji je normalno raspodeljen, ocene za μ i σ^2 su date formulama (11.1) i (11.2). Kao intervali poverenja se mogu uzeti $[\mu - c\sigma, \mu + c\sigma]$ za $c > 0$. Što je c veće, veća je i verovatnoća da X pripada tom intervalu. Za $c = 1$ ta verovatnoća je 68.27%, za $c = 2$ je 95.45%, a za $c = 3$ je 99.73%.

Višedimenziona slučajna promenljiva $X = (X_1, \dots, X_p)'$ ima p -dimenzionu normalnu (Gausovu) raspodelu, ukoliko je njena gustina raspodele

$$g(x) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-(x-\mu)'\Sigma^{-1}(x-\mu)/2}$$

gde je $x = (x_1, \dots, x_p)$, $-\infty < x_i < \infty$, $i = 1, \dots, p$, μ je matematičko očekivanje promenljive X , a Σ je simetrična, pozitivno definitna kovarijantna

matrica vektora X . Zbog pozitivne definitnosti je $|\Sigma| = \det(\Sigma) \neq 0$, pa Σ^{-1} postoji. Ovu raspodelu ćemo označavati sa $N_p(\mu, \Sigma)$. Navedimo osnovne osobine ove raspodele.

- Konturne hiperpovršni iste vrednosti gustine raspodele su elipsoidi dati formulom $(x - \mu)' \Sigma^{-1} (x - \mu) = c^2$. Elipsoidima je centar tačka μ , a za ose imaju $\pm c\sqrt{\lambda_i} e_i$, gde su λ_i sopstvene vrednosti matrice Σ , a e_i odgovarajući normirani sopstveni vektori, $\Sigma e_i = \lambda_i e_i$, $i = 1, \dots, p$.
- Ako X ima $N_p(\mu, \Sigma)$ raspodelu, za matricu A dimenzije $q \times p$ važi da AX ima raspodelu $N_q(A\mu, A\Sigma A')$. Za konstantan vektor d dimenzije p će $X + d$ imati raspodelu $N_p(\mu + d, \Sigma)$.
- Neka je $\Sigma = LL'$ dekompozicija Holeckog, gde je L donja trougaona matrica. Ako X ima $N_p(\mu, \Sigma)$ raspodelu, na osnovu prethodnog sledi da $L'(X - \mu)$ ima $N_p(0, I_p)$ raspodelu, gde je I_p jedinična matrica dimenzije p .
- Ako Z ima $N_p(0, I_p)$ raspodelu, tada su njene koordinate nezavisne i imaju $N(0, 1)$ raspodelu. Tada će $X = \mu + LZ$ imati $N_p(\mu, \Sigma)$ raspodelu. Ovaj rezultat služi za generisanje slučajnih tačaka sa ovom raspodelom.
- Ako X ima $N_p(\mu, \Sigma)$ raspodelu, tada svaka koordinata X_i ima $N(\mu_i, \Sigma_{ii})$ raspodelu. Svaki izabrani podskup parametara imaće takođe višedimenzionu normalnu raspodelu sa podmatricom matrice Σ koja sadrži odgovarajuće kolone i vrste tih izabranih parametara.
- Ako X ima $N_p(\mu, \Sigma)$ raspodelu, tada $(X - \mu)' \Sigma^{-1} (X - \mu)$ ima χ_p^2 raspodelu sa p stepeni slobode.
- Ako X ima $N_p(\mu, \Sigma)$ raspodelu, i ako $\chi_p^2(\alpha)$ označava gornji (100α) -ti percentil, tada je verovatnoća da X pripada elipsoidu $(x - \mu)' \Sigma^{-1} (x - \mu) \leq \chi_p^2(\alpha)$ jednaka $1 - \alpha$.

Za uzorak nezavisnih višedimenziono normalno raspodeljenih vektora (x_1, \dots, x_N) , svaki dimenzije M , statističke ocene parametara će biti

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i \quad (11.8)$$

i

$$\hat{\Sigma} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \hat{\mu})' (x_i - \hat{\mu}). \quad (11.9)$$

11.4.2 Intervali i regioni poverenja

Ako smo nekom od prethodno opisanih metoda dobili veći broj optimalnih parametara $\mathbf{a}_{(k)}$, $k = 1, \dots, K$ za isti broj simuliranih kompleta podataka,

pretpostavićemo da je raspodela ovih vektora parametara višedimenziona normalna.

Ukoliko nas zanimaju samo pojedinačni parametri, možemo jednostavno oceniti parametre jednodimenzione normalne raspodele za odgovarajuću koordinatu vektora parametara, po formulama (11.1) i (11.2). Nakon toga možemo odrediti odgovarajuće intervale poverenja oko dobijenih vrednosti parametra u funkciji izračunate standardne devijacije.

p	$\nu = 1$	$\nu = 2$	$\nu = 3$	$\nu = 4$	$\nu = 5$	$\nu = 6$
68.27%	1.00	2.30	3.53	4.72	5.89	7.04
90%	2.71	4.61	6.25	7.78	9.24	10.6
95.45%	4.00	6.18	8.02	9.72	11.3	12.8
99%	6.63	9.21	11.3	13.3	15.1	16.8
99.73%	9.00	11.8	14.2	16.3	18.2	20.1
99.99%	15.1	18.4	21.1	23.5	25.7	27.9

Tabela 11.1: Vrednosti $\Delta\chi^2$ za razne nivoe poverenja p i stepena slobode ν

Ukoliko nas, međutim, interesuju oblasti poverenja za više parametara istovremeno, možemo odrediti regione (elipsoide) poverenja na osnovu prethodno navedenih osobina višedimenzione normalne raspodele. Prvo se naprave statističke ocene po formulama (11.8) i (11.9). Vrednost funkcije $\chi^2(\mathbf{a})$ za originalne podatke će biti najmanja kada je $\mathbf{a} = \mathbf{a}_{(0)}$ i tu vrednost označimo sa $\chi_{\min}^2 = \chi^2(\mathbf{a}_{(0)})$. Za sve ostale vrednosti $\mathbf{a}_{(k)}$, $k = 1, \dots, K$, vrednost će biti veća. Veličina $\Delta\chi^2 \equiv \chi^2(\mathbf{a}_{(j)}) - \chi^2(\mathbf{a}_{(0)})$ će imati χ^2 raspodelu sa M stepeni slobode.

U slučaju da nas ne interesuje region poverenja za sve, nego samo za $\nu < M$ parametara, postupićemo na sledeći način. Nakon dobijanja $\mathbf{a}_{(0)}$, za sve druge simulirane kompletne podatke vršićemo optimizaciju samo po parametrima koji nas interesuju, dok preostale parametre fiksiramo na vrednosti iz $\mathbf{a}_{(0)}$.

Određivanje elipsoida poverenja se vrši na sledeći način.

1. Neka nas interesuje zajednička oblast poverenja za $\nu \leq M$ izabranih parametara.
2. Izabрати nivo poverenja p , recimo 0.9, 0.95, 0.99 ili neki drugi.
3. Izračunati takvo Δ da je verovatnoća da je χ^2 promenljiva sa ν stepeni slobode manja od Δ tačno jednaka p . Za češće korišćene verovatnoće se te vrednosti mogu naći u Tabeli 11.1.

4. Iz matrice kovarijansi $C = \alpha^{-1}$, dobijene u metodi najmanjih kvadrata, izdvojiti podmatricu dimenzije ν koja odgovara parametrima koji nas interesuju, i označimo je sa C_{proj} .
5. Invertovati matricu C_{proj} .
6. Jednačina elipsoida poverenja za ν izabranih parametara je

$$(\mathbf{a} - \mathbf{a}_{(0)})' \cdot C_{\text{proj}}^{-1} \cdot (\mathbf{a} - \mathbf{a}_{(0)}) \leq \Delta.$$

11.5 Kuda ide ovaj svet?

Ljudi su oduvek bili fascinirani lepotom nebeskog svoda, a Suncu, Mesecu i zvezdama, naročito pokretnim (planete) su pridavali božanske osobine. Periodične pojave pomračenja Sunca i Meseca su izazivale strah u drevnim vremenima a, neposredno smo se uvelili, i u bliskoj prošlosti (11. avgusta 1999. u Srbiji). Onaj ko je takve pojave mogao da predvidi, nesumnjivo je posedovao veliki ugled i moć. Radi predviđanja astronomskih pojava, naučnici su se uvek služili matematičkim modelima, koji su zavisili od nivoa naučnih znanja kao i preciznosti raspoloživih instrumenata. Crkve su bile prirodno zainteresovane za ovladavanje ovim znanjima, centri nauke i pismenosti su bili ranije prvenstveno pri manastirima, ali čim su naučne istine došle u sukob sa verskim dogmama, nauci je bilo sve teže da se razvija.

Za opisivanje kretanja je potrebno odrediti tačku (centar) oko koga se tela kreću. Matematički je prihvatljivo staviti koordinatni početak u proizvoljnu tačku, ali različita rešenja daju više ili manje komplikovane putanje (orbite) tela koje se kreću. Za drevne ljude je bilo očigledno da je Zemlja "dole", a nebo "gore", i da se sve na nebu kreće oko njih i Zemlje. Ubrzo su otkrili da je Zemlja zakrivljena, a Eratosten je dosta tačno izračunao i njen obim. Iako je još u helenizmu postojala heliocentrična teorija, u srednjem veku je geocentrična teorija bila zvanično prihvaćena, prvenstveno iz verskih razloga.

Sa matematičke strane, posmatrano kroz tačnost matematičkih modela, geocentrična, Braheova geoheliocentrična i Kopernikova heliocentrična teorija su bile ekvivalentne, tj. davale su isti nivo grešaka u predviđanjima zbog pretpostavke o kružnim orbitama (uz pomoćne epicikle). Sa fizičke strane postoji bitna razlika između geocentrične i heliocentrične teorije. Kod epicikala (kružnice čiji se centar kreće po kružnici oko Zemlje), po kojima se kreću planete u geocentričnoj teoriji, nije jasno šta bi mogao biti uzrok kretanja, jer centri epicikala nisu bili vezani za neka fizička tela. U Kopernikovoj