Metod najmanjih kvadrata kroz primere

Zoran D. Jeličić
Mirna N. Kapetina
17. novembar 2021.

1 Uvodna razmatranja

U okviru ovog poglavlja namera nam je da detaljnije izučimo metod najmnjih kvadrata, kao jedan od najpoznatijih alata u problemima regresije. Počećemo naše razmatranje sa problemom linearne regeresije, nastaviti sa problemima koji se mogu svesti na problem linearne regresije, zatim ćemo nastaviti sa polinomijalnom regresijonom analizom višeg reda, da bi se nakraju poglavlja skocentrisali na studiju strogo nelinearnih problema.

1.1 Linearna regresija i metod najmanjih kvadrata

Počećemo od postupka linearne regresije uz oslonac na metod najmanjih kvadrata. Ova dva postupka predstavljaju osnovu mašinskog učenja [1] i predstavićemo ih kao optimizacioni postupak, što oni svakako jesu, bez ulaska u varijacije algoritma, koji su najčešće posledice prilagođavanja konkretnim inženjerskim problemima.

[1] *machine learning*

Uspostavljanje funkcionalne zavisnosti između dva izmerena obeležja je problem regresije, i ukoliko je ta zavisnost linearna, postupak se očigledno naziva **linearna regresija**. Posmatraćemo dvodimenzionalni uzorak (x_i, y_i) gde je x_i nezavisna ili prediktorska promenljiva, a y_i se naziva zavisna promenljiva ili odziv (ishod). Vezu između obeležja smo pretpostavili u lineranoj formi [2]

[2] Postupci regresije zavise od pretpostavljnog profila veze između nezavisnih i zavisnih promenljivih, mada suštinski u pozadini svih njih se nalaze slični optimizacioni problem, čije rešavanje zavisi od složenosti pretpostavljene funkcionalne veze.

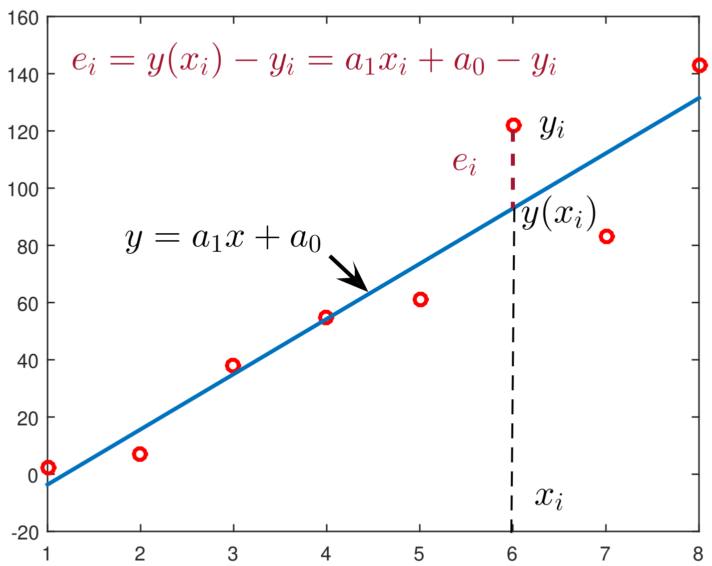
$$y = a_1 x + a_0 \tag{1}$$

gde su a_1 i a_0 nepoznati parametri, koje želimo da odredimo u optimizacionoj proceduri, pri čemu parametar a_1 predstavlja nagib krive, a parametar a_0 je mesto preseka sa ordinatom slika 1 [3].
Pretpostavili smo kriterijum optimalnosti u kvadratnoj formi, gde želimo da minimizujemo grešku predviđanja (reziduala) [4], odnosno

[3] Sigurni smo da su se čitaoci upoznali sa problemom linearne regresije, makar intutitvno, primenjujući Hukov zakon i računajući Jangov modul elastičnosti.

[4] Odatle i ime optimizaciong problema, *metod najmanjih kvadrata*

$$I = \sum_{i=1}^{n} (y - y_i)^2 \,. \tag{2}$$



Slika 1: Ilustracija linearne regresije, greška aprokismacije obležena je sa e_i i izračunava se kao razlika između izračunate vrednosti $y(x_i)$ i ishoda y_i odnosno, $e_i = y(x_i) - y_i$.

Respektujući pretpostavljeni profil funkcionalne veze (1), slika 1, kriterijum optimalnosti (2) sada postaje

$$I\left(a_0, a_1\right) = \sum_{i=1}^{n} \left(a_1 x_i + a_0 - y_i\right)^2 . \tag{3}$$

Kao što smo ranije napomenuli, cilj nam je da odredimo optimalne vrednosti parametara a_0 i a_1 tako da je kriterijum optimalnosti (3) u minimumu. Na osnovu dobro potrebnih uslova ekstrema, sledi

$$\begin{aligned}
\frac{\partial I}{\partial a_0} &= \sum_{i=1}^{n} 2\left(a_1 x_i + a_0 - y_i\right) = 0 \\
\frac{\partial I}{\partial a_1} &= \sum_{i=1}^{n} 2\left(a_1 x_i + a_0 - y_i\right) x_i = \sum_{i=1}^{n} 2\left(a_1 x_i^2 + a_0 x_i - y_i x_i\right) = 0 .
\end{aligned} \tag{4}$$

Linearni sistem jednačina (4) može se napisti i u matričnoj formi

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{bmatrix} \tag{5}$$

Ukoliko smo u stanju da rešimo [5] sistem jednačina (5) i dobijamo optimalne vrednosti parametara a_0 i a_1, formalno smo završili postupak linearne regresije.

U nastvaku ćemo nastaviti sa eksplictnim rešenjem matrične jednačine (2) i njegovom daljom primenom u studiji problema regresije. Do rešenja jednačine (2) lako dolazimo primenom Kramerovog [6] pravila

[5] Rešavanje ovog matričnog sistema jednačina je po pravilu problem sam za sebe i spada u domen linearne algebre.

[6] Gabriel Cramer 1704 – 1752, Švajcarski matematičar, poznat po teroiji linearne algebre, lepo je reći da je doktorirao sa 18 godina.

$$\begin{aligned} a_1^* &= \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i{}^2 - \left(\sum_{i=1}^n x_i\right)^2} \\ a_0^* &= \frac{\sum_{i=1}^n y_i}{n} - a_1^* \frac{\sum_{i=1}^n x_i}{n} \ , \end{aligned} \tag{6}$$

gde su sa a_0^* i a_1^* obeležene optimlane vrednosti parametara a_0 i a_1 u skladu sa pretpostavljenim kriterijumom optimalnosti (3). Optimalnu vrednost kriterijima, obeležićemo sa I_{LSE} i izračunavamo ga na sledeći način

$$I_{LSE} = \sum_{i=1}^{n} (a_1^* x_i + a_0^* - y_i)^2 \ . \tag{7}$$

Ovo praktično znači da će svaka druga prava linija koja daje vezu između (x_i, y_i) imati "lošiju"vrednost kriterijuma optimalnosti tj. $I(a_0, a_1) > I_{LSE}$.
Da bi poredili učinak metode najmanjih kavdrata, uvešćemo i srednju vrednost odziva $\bar{y}_i$,

$$\bar{y}_i = \frac{\sum_{i=1}^n y_i}{n} \ , \tag{8}$$

odnsono sumu kavdrata razlike između y_i i $\bar{y}_i$

$$I_M = \sum_{i=1}^{n} \left(\bar{y} - y_i\right)^2 \ , \tag{9}$$

kao uobičajenu meru odstupanja u ovakvim problemima. Tako na primer ako je razlika $I_M - I_{LSE} > 0$, značilo bi da je linearna aproksimacija bolja od zamene svih odziva srednjom vrednošću. Veza između I_{LSE} i I_M se takođe izražava i relativnom greškom na r^2 na sledeći način

$$r^2 = \frac{I_M - I_{LSE}}{I_M} = 1 - \frac{E_{LSE}}{E_M} \ . \tag{10}$$

pri čemu $r^2 = 1$ znači da je aproksimacija pravom linija šavršena", a ako je $I_M = I_{LSE}$ očigledno neće biti razlike u kvalitetu aproksimacije i tada je $r^2 = 0$. Praktično to znači da je za vrednost $r^2 = 0.9$, linerni model sa 90% uspešnosti daje funkcinalnu zavisnot između vrednosti x_i i odgovarajućih y_i.
Interseantno je uvesti korelacioni keoficijent r, koji se na osnovu (6) i (10) direktno izračunava na sledeći način

$$r = \sqrt{\frac{I_M - I_{LSE}}{I_M}} = \frac{n\sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n x_i{}^2 - \left(\sum_{i=1}^n x_i\right)^2}\sqrt{n \sum_{i=1}^n y_i{}^2 - \left(\sum_{i=1}^n y_i\right)^2}} \tag{11}$$

SLEDI PRIMER IZ MATLABA LSEprvi.m

1.2 *Linearizacija nelinearnih modela*

Logično je pretpostavit da veze između ulaza x_i i izlaza y_i nije uvek linearna, ali postoji široka klasa nelinearnih problema, gde se uz odgovarajuće smene početni problem nelinearne regresije može svesti na linearan. Najlakše je ovu svojevrsnu funkcionalnu linearizaciju ilustrovati na primerima.

1. Eksponencijalni model [7]

$$y(x) = \alpha e^{\beta x} \ .$$

Linearizacija ovog modela je prirodna i dosta jednostavna. Ako nađemo prirodni logaritam prethodnog izraza dobijamo

$$\ln y = \ln \alpha + \beta x \ .$$

sada je očigledno da je novo $Y = \ln y$, a koeficijenti iz jednačine (1) su $a_0 = \ln \alpha$ i $a_1 = \beta$, odnosno

$$Y = a_0 + a_1 x \ .$$

Primenom jednačina (6) i (11) lako određujemo optimalne vrednosti parametara linearizovanog modela, odnosno izračunavamo a_0^* i a_1^*, što konačno daje eksponencijalni regresioni model kao

$$y(x) = \alpha e^{\beta x} = \left(e^{a_0^*}\right) e^{a_1^* x} \ .$$

2. Stepeni model

$$y(x) = \alpha x^{\beta} \ .$$

Linearizacija ovog modela je slična kao i prethodna i grubo je dajemo kroz sledeće korake. Prvo nađemo prirodni logaritam prethdonog izraza

$$\ln y = \ln \alpha + \beta \ln x \ .$$

Smenom $Y = \ln y$ i $X = \ln x$, lako dobijamo formu linearnog modela

$$Y = a_0 + a_1 X \ .$$

Sada je očigledno $\alpha = e^{a_0}$. $\beta = a_1$.

[7] Eksponecijalni model je veoma zastupljen u teoriji sistema i elektrotehnike uošte. Jasno je da se može dobiti kao rešenje linearne diferencijalne jednačine prvog reda

Čiatocima ostavljamo za vežbu da nađu linearnu vezu sledećih funkcija

$y = \alpha \ln x + \beta$, $y = \frac{1}{\alpha x+\beta}$ i $y = \frac{\alpha x}{\beta + x}$.

SLEDI PRIMER IZ MATALABA LSEdrugi.m

1.3 Uoštena linearna regresija i polinomijalni modeli

Pokušaćemo da uopštimo razmatranja iz pretodnog poglavlja i pretpostavimo opštiji profil, koji daje vezu između ulaznih i izlaznih promenljivih x_i i y_i respektivno

$$y = a_m z_m + a_{m-1} z_{m-1} + \ldots + a_1 z_1 + a_0 \qquad (12)$$

gde se z_i predstavljaju bazne funkcije [8] koje zavise x_i. Tako na primer, ako izaberemo $z_i = x^i$, dobijamo da je polinomsku profilnu funkciju m-tog reda

$$y = a_m x^m + a_{m-1} x^{m-1} + \ldots + a_1 x + a_0 \ . \qquad (13)$$

Neki od standradnih linearnih ili profila koji se mogu linearizovati su

$$y = a_0 + a_1 x, \quad y = a_0 + a_1 \cos(x) + a_2 \sin(2x) \quad \text{ili} \quad y = a_0 + a_1 x a_2 e^{-x^2} \ .$$

Intersentna je veza

$$y = a_0 + a_1 e^{a_2 (x-a_3)^2} \ ,$$

koja se naziva radijalna bazna funkcija [9]. Ovu pozitivno definitnu funkciju nije moguće linearizovati u duhu teorije iznete u prethodnom paragrafu, pa se problemi ovog tipa nazivaju nelinearna regresija u njima ćemo se baviti u poglavlju koje sledi.

U nastvaku teksta fokusiraćemo se na problem linearne regresija, ako je veza između ulaza x_i i izlaza y_i data kao polinom m-tog reda

$$y = a_m x^m + a_{m-1} x^{m-1} + \ldots + a_1 x + a_0 \ . \qquad (14)$$

Prepostavimo da za formiranje regresionog modela imamo na raspolaganju n parova (x_i, y_i), odnosno

$\{(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)\}$, a želimo da izračunamo $m+1$ regresioni parametar iz (14), što čini sledeći sistem jednačina

$$\begin{aligned}
a_m x_1^m + a_{m-1} x_1^{m-1} + \ldots a_1 x_1 + a_0 &= y_1 \\
a_m x_2^m + a_{m-1} x_2^{m-1} + \ldots a_1 x_2 + a_0 &= y_2 \\
&\vdots \\
a_m x_n^m + a_{m-1} x_n^{m-1} + \ldots a_1 x_n + a_0 &= y_n \ .
\end{aligned} \qquad (15)$$

[8] Tremin bazne funkcije je direktan prevod engleskog izraza *basis function*, pored ovog izraza koriste se i termini kernel funkcije ili jezgro. U teoriji aproksimacija, koja je vezana za optimalno upravljanje koristi se i pojam *profil*, koji smo kao što ste videli mi rezervisali za nešto uopšteniju funkcionalnz vezu između x_i i y_i.

[9] Prevod sa engleskog *Radial Basis Function* i veoma često je gradivna jedinica vešatčkih neruonskih mreža.

Sistem jednačina (15) može se zapisati i u kompaktnijoj matričnoj formi $\mathbf{Z}\boldsymbol{a} = \mathbf{y}$,

$$\begin{bmatrix} x_1^m & x_1^{m-1} & \dots & x_1 & 1 \\ x_2^m & x_2^{m-1} & \dots & x_2 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{n-1}^m & x_{n-1}^{m-1} & \dots & x_{n-1} & 1 \\ x_n^m & x_n^{m-1} & \dots & x_n & 1 \end{bmatrix} \begin{bmatrix} a_m \\ a_{m-1} \\ \vdots \\ a_1 \\ a_0 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-1} \\ y_n \end{bmatrix}, \tag{16}$$

gde je $\mathbf{Z}$ pravougaona matrica dimenzije $n \times (m+1)$, koja očigledno zavisi od promenljivih x_i, pri čemu važi $z_{ij} = x_i^{m-j+1}$, $i = 1, 2, \dots, n$, $j = 1, 2, \dots, m+1$. Vektor $\boldsymbol{a}$ predstavlja sve nepoznate regersione parametre, vektor kolona $\mathbf{y}$ očigledno predstavlja svih n odziva y_i. Pod realnom pretpostavkom da je broj podataka n veći od broja nepoznatih regresionih keoficijenata $m+1$, naš problem je predefinisan i rešavanje sistema jednačina (15) - (16) u tom duhu nije puno smisleno. Uputno je problem određivanje nepoznatih regresionih koeficijenata postaviti kao problem najmnjih kvadrata i problem rešiti u duhu nelinearnog programiranja.