Rast onesnaženosti z ogljikovim dioksidom

Emil Žagar

14. april 2021

V praksi se pogosto srečamo s problemom, ko je treba neko množico podatkov aproksimirati s predvidenim modelom funkcije, vendar imamo veliko več podatkov, kot prostih parametrov v modelu. Najpreprostejši primer je na primer aproksimacija velikega števila točk s polinomom nizke stopnje. Seveda pa ni nujno, da v modelu nastopajo samo polinomi, lahko kakršnekoli funkcije, pomembno je le, da neznani parametri nastopajo linearno.
Tak problem je na primer aproksimacija podatkov o onesnaževanju ozračja z ogljikovim dioksidom. Observatorij Mauna Loa na Hawaiih vsaj od leta 1959 naprej meril vsebovanost CO_2 v ozračju. Znanstveniki so domnevali, da onesnaženost narašča linearno, treba pa je upoštevati letne periode zaradi različnih dejanikov, denimo letnih časov, razmer v industriji preko enega leta,.... Tako so menili, da lahko onesnaženost ozračja ob določenem času približno določimo s funkcijo oblike

$$F(a, b, c, d, t) = a\,t + b + c\,\sin(2\,\pi\,t) + d\,\cos(2\,\pi\,t).$$

Parametre a, b, c in d moramo določiti na podlagi ogromnega števila meritev, da se bo funkcija F čim bolj prilegala izmerjenim vrednostim F_i. Torej želimo, da je

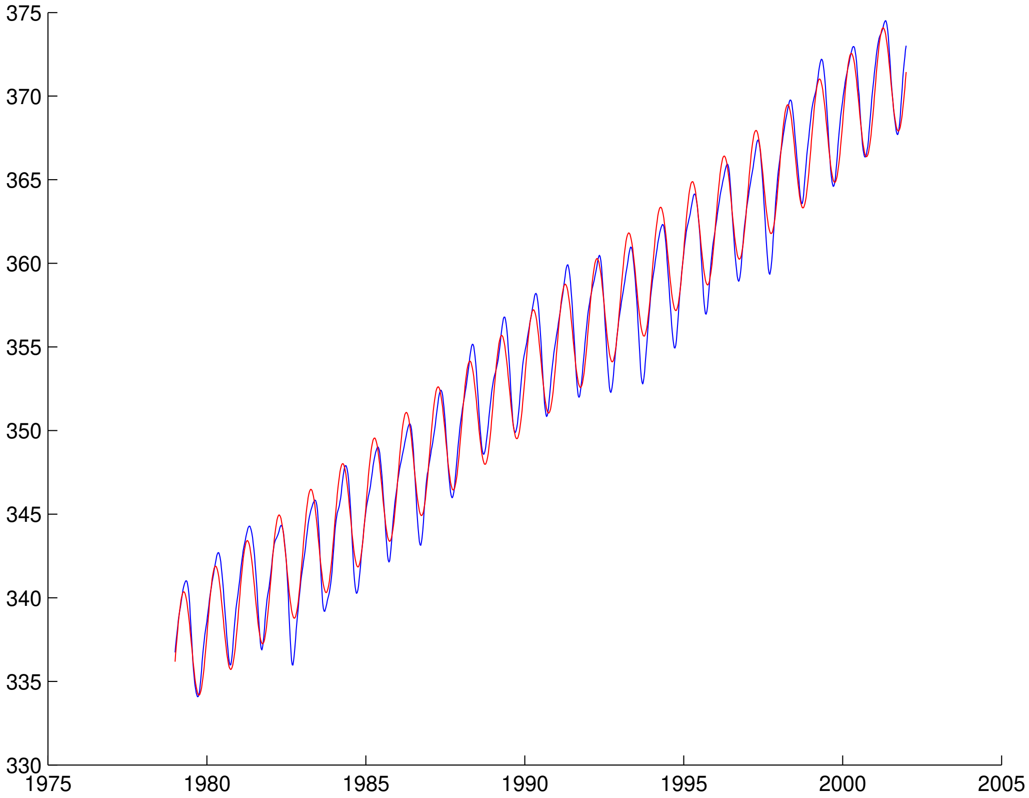
$$\sum_{i=1}^{n} \|F(a, b, c, d, t_i) - F_i\|_2$$

minimalno (pri tem je n število meritev). To je klasičen problem najmanjših kvadratov, ki ga lahko rešujemo na več načinov. Začnemo takole. Najprej napišemo predoločen sistem enačb (več enačb kot neznank)

$$F(a, b, c, d, t_i) = F_i, \quad i = 1, 2, \ldots, n,$$

za neznane parametre a, b, c in d. Lahko ga zapišemo v matrični obliki

$$A\boldsymbol{x} = \boldsymbol{b},$$



Slika 1: Rdeča črta je graf aproksimacijske funkcije, modra pa graf izmerjenih podatkov (vsebovanosti CO_2 v milijoninkah).

kjer je $A \in \mathbb{R}^{n\times 4}$, $\boldsymbol{x} \in \mathbb{R}^{4\times 1}$ in $\boldsymbol{b} = [a, b, c, d]^T$. Sistem je predoločen, zato ga rešimo tako, da ga pomnožimo z A^T, torej $A^T A\boldsymbol{x} = A^T \boldsymbol{b}$. Ker je sedaj kvadraten dimenzije 4×4 (in neizrojen, če je A ranga 4, kar je spet res, če so funkcije v modelu neodvisne), ga zlahka rešimo z Matlabom. Lahko pa kar na originalnem sistemu $A\,\boldsymbol{x} = \boldsymbol{b}$ uporabimo Matlabov ukaz `x=A\b`, saj je znano, da Matlab predoločene sisteme rešuje po metodi najmanjših kvadratov (vendar ne tako, kot smo opisali zgoraj, temveč uporabi t.i. QR razcep s pivotiranjem po stolpcih).
Na sliki imamo primer aproksimacije meritev vsebovanosti CO_2 na podlagi meritev imenovanega observatorija v letih 1979 do 2002.[1]

[1]Podatki so na voljo na strani
`https://www.esrl.noaa.gov/gmd/ccgg/trends/data.html`