



Modelska Analiza 1

6. naloga

Luka Papež
28221030

1 Naloga - Luščenje modelskih parametrov in razdelčni modeli

- V farmakologiji merijo odziv tkiv na različne reagente. Za večino teh pojavov lahko privzamemo, da gre za reakcijo, kjer spremljamo vezavo molekul reagenta X na receptorje Y v tkivu.



V stacionarnem stanju dobimo zvezo

$$y = \frac{y_0 x}{x + a}. \quad (2)$$

Iz merskih podatkov v datoteki `farmakoloski.dat` določi parametra y_0 in a , kjer pomeni y_0 nasičeni odziv tkiva in a koncentracijo, potrebno za odziv, ki je enak polovici nasičenega. Napaka v meritvi odziva je v vsem področju enaka trem enotam. Zvezo lahko lineariziramo. Pazi, kako se pri tem transformirajo napake.

Določi še parametre razširjenega nelinearnega modela

$$y = \frac{y_0 x^p}{x^p + a^p} \quad (3)$$

in statistično upravičenost vpeljave dodatnega parametra p .

- Poišči najboljšo vrednost za čistilnost ledvic iz kliničnih podatkov v datoteki `ledvice.dat` z uporabo enorazdelčnega in dvorazdelčnega modela ter primerjaj rezultate. Pri dvorazdelčnem modelu lahko za začetni približek vzameš eksponentni konstanti v razmerju 1:10. Ali je dodatek aditivne konstante ("ozadje" pri štetju razpadov) statistično upravičen? Poskusиш lahko tudi s funkcijo $e^{-\lambda\sqrt{t}}$, ki jo izvedemo iz bolj zapletenih modelov.

Spremenljivka t v podatkih je čas na sredi vsakega merilnega intervala.

- Za uporabo visokoločljivostnega magnetnega spektromетra potrebujemo preslikavo, ki iz izmerjenih količin rekonstruira parametre trajektorije delcev, potrebne za izračun energije in drugih kinematičnih količin. V datoteki `thtg-xfp-thfp.dat` najdete kalibracijske podatke s stolci ϑ_{tg} (disperzijski kot na tarči v stopinjah) ter x_{fp} (položaj v goriščni ravnini v milimetrih) in ϑ_{fp} (kot iščni ravnini v stopinjah). Natančnosti meritev kotov so velikostnega reda miliradianov, položajev pa okrog milimetra. Sestavi varčni model za preslikavo

$$(x_{fp}, \vartheta_{fp}) \Rightarrow \vartheta_{tg}. \quad (4)$$

Uporabiš lahko na primer najnižje potence x_{fp} in ϑ_{fp} ali pa kakšne druge funkcije teh dveh spremenljivk.

Za izbiro relevantnih členov uporabiš SVD razcep in reducirani χ^2 .

2 Vezava molekul reagenta na receptorje

Reakcijo iz enačbe (1) lahko opišemo s hitrostma reakcije v desno $v_+ = k_+[X][Y]$ in v levo $v_- = k_-[Y^*]$, iz česar sledi diferencialna enačba:

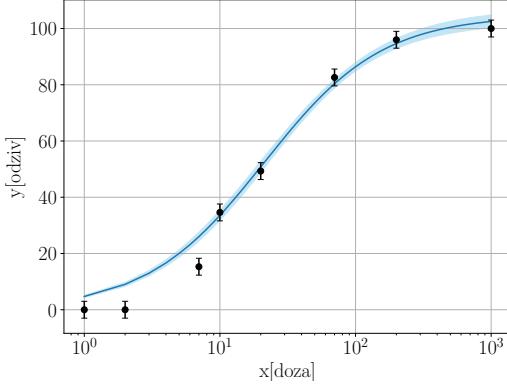
$$\frac{d[Y^*]}{dt} = v_+ - v_- = k_+[X][Y] - k_-[Y^*].$$

V stacionarnem stanju z ohranitvijo reagentov $y_0 = [Y] + [Y]^*$ nato opišemo sprememjanje koncentracije $[Y^*]$ na enačbi (2). Formulo lahko nato lineariziramo v obliko

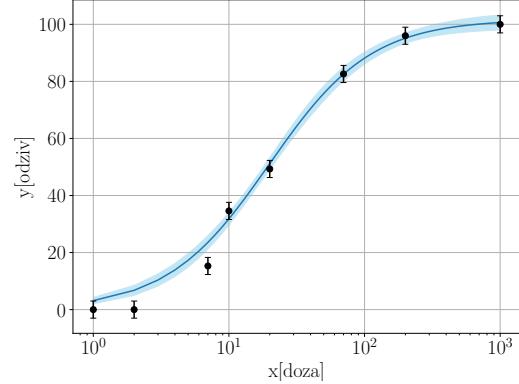
$$\frac{1}{y} = \frac{1}{y_0} + \frac{a}{y_0} \frac{1}{x} = a_0 + a_1 \frac{1}{x},$$

kar nam olajša fitanje podatkov. Za opis stacionarnega stanja vzamemo še nelinearni model (3), ki ga zapišemo v polinomsko obliko

$$\frac{1}{y} = \frac{1}{y_0} + \frac{a^p}{y_0} \frac{1}{x^p} = a_0 + a_1 \left(\frac{1}{x} \right)^p.$$



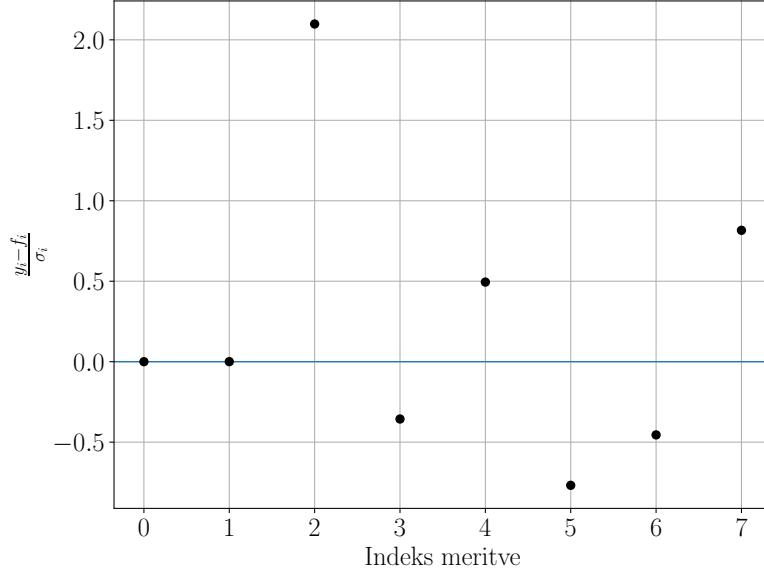
(a) Fit linearnega modela.



(b) Fit nelinearnega modela.

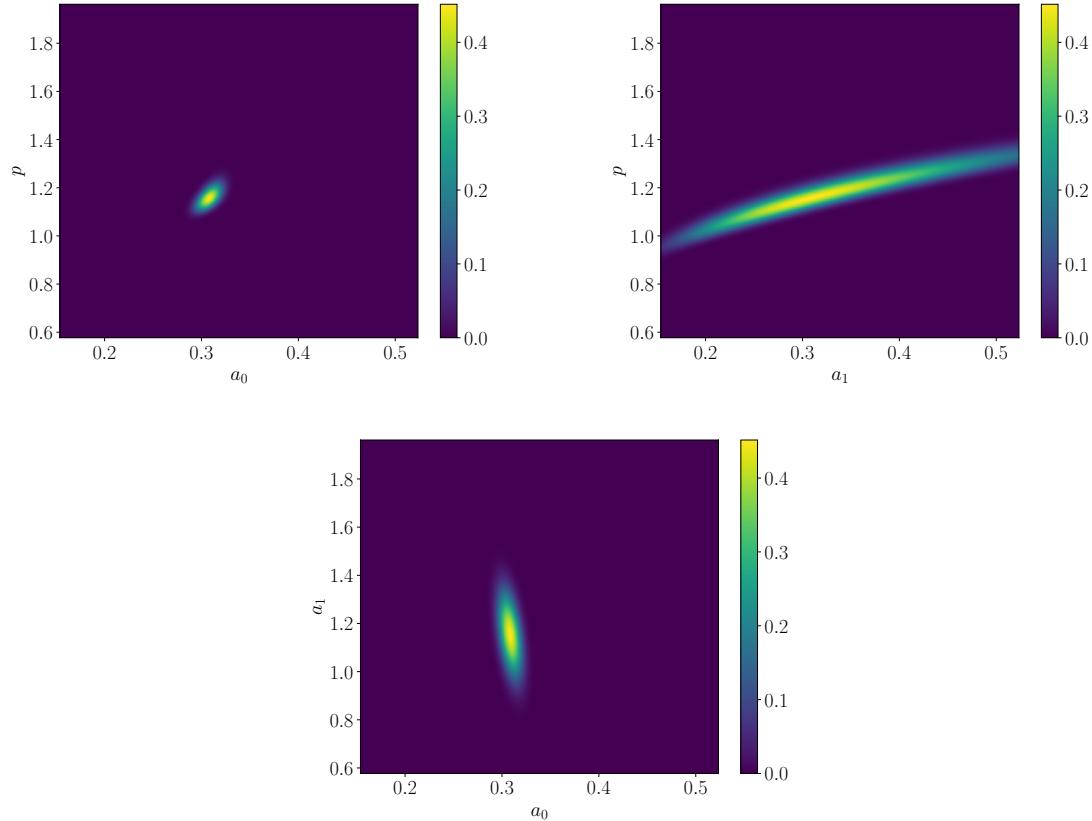
Slika 1: Fit linearnega in nelinearnega modela na podatke iz datoteke `farmakoloski.dat` in napaka fita propagirana iz kovariančne matrike.

Statistično upravičenost modelov lahko upravičimo s χ^2 test s katerim lahko določimo pogostost scenarija, če je uporabljen model upravičen glede na izmerjene podatke. Za linearni model je p vrednost enaka $p = 0.40$ za nelinearni model pa $p = 0.452$. Za delovanje χ^2 testa predpostavimo, da je vsaka meritev upravljena neodvisno po naključni Gaussovi porazdelitvi z znano standardno deviacijo. Tega sicer zaradi majhnega števila meritev ne moremo preveriti, a lahko kljub temu narišemo odstopanje od vrednosti iz fittanega modela.



Slika 2: Odstopanje meritev od linearnega modela fitanega na te meritve.

Zanimiv vpogled kako se obnaša parametrov nam ponudi tudi fiksiranje enega izmed treh parametrov v nelinearnem modelu na optimalno vrednost in opazovanje obnašanja p vrednosti ob spremjanju drugih dveh.



Slika 3: Obnašanje p vrednosti pri enem fiksнем parametru na optimalni vrednosti in spremjanju drugih dveh.

3 Čiščenje ledvic

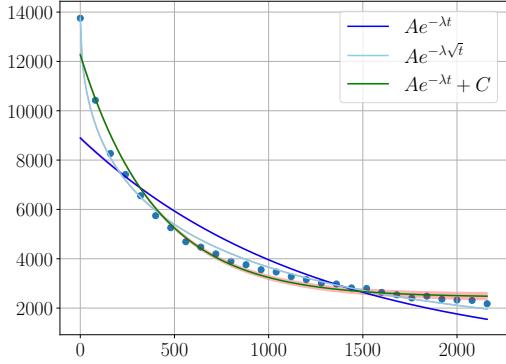
Čiščenje ledvic opišemo z enorazdelčnim $c(t) = A \exp(-\lambda t)$ oziroma z dvorazdelčnim modelom $c(t) = A \exp(-\lambda_1 t) + B \exp(-\lambda_2 t)$. Dodatno si ogledamo še primer z ozadjem in primerom, kjer v modelih uporabimo člene $\exp(-\lambda \sqrt{t})$. Na koncu uporabimo še F-test, da preverimo statistično utemeljenost dodajanja ozadja. Tukaj je pomembna opomba, da F-test deluje le na linearnih gnezdenih modelih. Na nelinearnih kot so naši služijo le kot približek, model brez ozadja sicer lahko lineariziramo, a tega ne moremo storiti na tistem z ozadjem. Modela s korenjenim časom pa na tak način sploh ne moremo primerjati z drugimi saj ni gnezden z ostalimi.

Kot lako opazimo iz fitfov na sliki 4 opazimo, da se v enorazdelčnem modelu najbolje ujema tisti model s korenjenim časom. V dvorazdelčnem modleu pa sta tisti z ozadjem in osnovni dvorazdelčni model boljša od tistega s korenjenim časom. Nato še izračunamo F-test po formuli

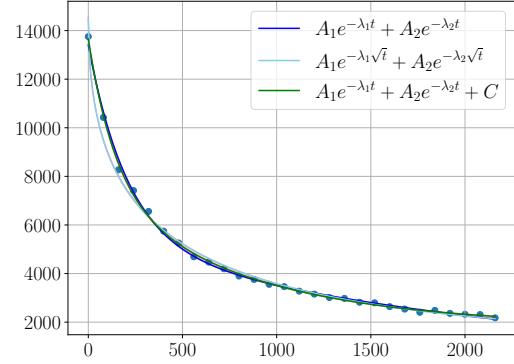
$$F = \frac{(\chi_1^2 - \chi_2^2)/(k_2 - k_1)}{\chi_2^2/(N - k_2)},$$

kjer sta k_1 in k_2 število fitanih parametrov in N število meritev. Za enorazdelčni model dobimo

$p = 1.8 \cdot 10^{-13}$, za dvorazdelčni model pa $p = 2.3 \cdot 10^{-5}$. Kar pomeni, da je dodatek aditivne konstante statistično upravičen.



(a) Enorazdelčni modeli.



(b) Dvorazdelčni modeli.

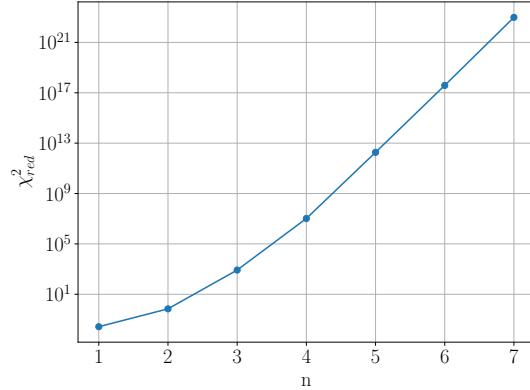
Slika 4: Obnašanje p vrednosti pri enem fiksном parametru na optimalni vrednosti in spremenjanju drugih dveh.

4 Visokoločljivostni magnetni spektrometer

Poskušamo kalibrirati visokoločljivostni magnetni spektrometer tako, da poiščemo transformacijo na enačbi (4). Postavimo model

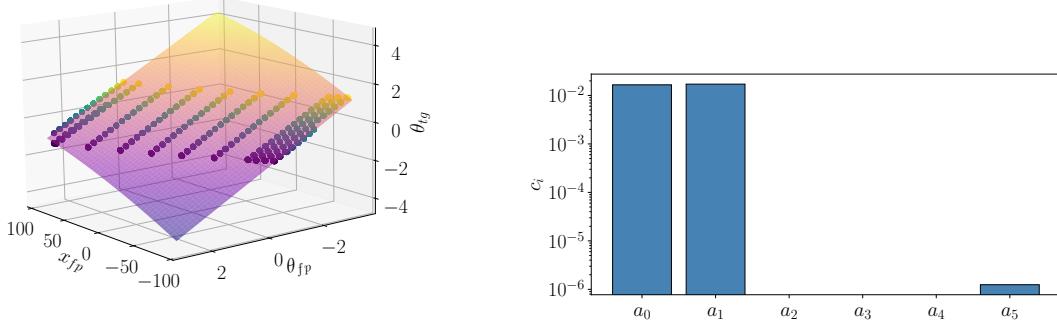
$$\begin{aligned}\theta_{tg}(x_{fp}, \theta_{fp}) &= (a_0 + a_1 x_{fp} + a_2 x_{fp}^2) + (a_3 + a_4 x_{fp} + a_5 x_{fp}^2) \theta_{fp} + \\ &\quad + \dots + (a_6 + a_7 x_{fp} + a_8 x_{fp}^2) \theta_{fp}^n.\end{aligned}$$

Za uspešnost našega modela definiramo $\chi^2_{red} = \frac{1}{N-M} \left(\frac{(\theta_{tg} - \theta_{tg model})}{\sigma} \right)^2$, kjer je N število meritev in M število parametrov. Idealen model ima $\chi^2_{red} \approx 1$, saj je takrat varianca modela približno enaka varianci meritev. Nato konstruiramo matriko A v sistemu $A\mathbf{c} = \theta_{tg}$ in nato še poiščemo vektor \mathbf{c} s pomočjo metode SVD.



Slika 5: Odstopanje meritev od linearnega modela fitanega na te meritve.

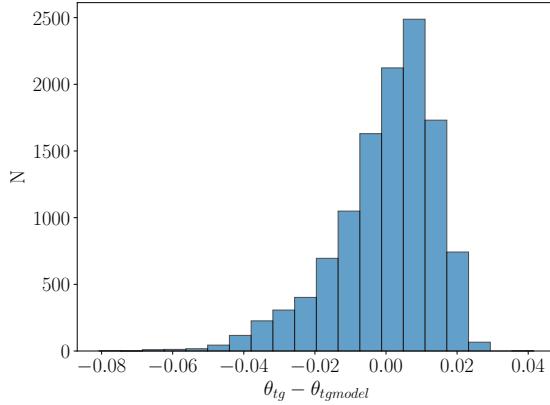
Opazimo, da pri vrednosti $n = 2$ dobimo najmanjo vrednost $\chi^2_{red} = 0.695$, ki je tudi najbližja 1 za prvo potenco je namreč vrednost $\chi^2_{red} = 0.03$ ostale vrednosti pa so več redov prevelike, kljub temu, da je odstopanje modela od meritev pri njih manjše.



(a) Ujemanje modela s kalibracijskimi podatki.

(b) Pomembnost uteži

Za razliko od prvega dela naloge imamo pri tej dovolj podatkov, da lahko bistveno pogledamo porazdelitev odstopanja modela od podatkov. Porazdelitev bolj kot na Gaussovo spominja na Boltzmannovo porazdelitev.



Slika 7: Odstopanje meritev od modela z $n = 2$.

Zanimiva alternativa trenutnemu pristopu pa je simbolična regresija, ki poišče analitične rešitve, ki se najbolje prilegajo podatkom. To storii tako, da generira več naključnih enačb v obliki dreves in nato evolucijsko te enačbe spreminja in rekombinira glede na njihovo prileganje. Na podlagi metod kot je Paretova optimizacija. Knjižnico PySr, ki v Pythonu implementira simbolično regresijo vrne 20 enačb, ki se razporejene po kompleksnosti in prileganju meritvam. Najboljši modeli imajo vrednost $\chi^2_{red} \approx 0.4$, kar je nekoliko slabše od prej najdene rešitve. Najboljši vrednost ima pa sicer model 12 glede na knjižnico pa je najboljše razmerje med kompleksnostjo in prileganje model 17.

Model	0	1	2	3	4	5
χ^2_{red}	∞	1730.89	0.2594	9016.67	0.3798	3385.39
Model	6	7	8	9	10	11
χ^2_{red}	0.3713	0.3824	0.3827	0.3911	0.3916	0.4332
Model	12	13	14	15	16	17
χ^2_{red}	0.4580	0.3903	0.3948	0.3907	0.3863	0.3935
Model	18	19				
χ^2_{red}	0.3936	5.9804				

Model 0: -0.016020572

Model 1: $\theta_{fp} \cdot (-0.24914362)$

Model 2: $(x_{fp} \cdot 0.020328978) - \theta_{fp}$

Model 3: $\sin((x_{fp} \cdot 0.020399055) - \theta_{fp})$

Model 4: $((x_{fp} \cdot 0.020430202) - \theta_{fp}) \cdot 0.8222651$

Model 5: $\sin(((x_{fp} \cdot 0.020228948) - \theta_{fp}) - 0.13380574)$

Model 6: $\frac{(-0.13804644 - (x_{fp} \cdot -0.020217843)) - \theta_{fp}}{1.1899875}$

Model 7: $((0.017186569 - (x_{fp} \cdot 2.9461953 \times 10^{-5})) \cdot x_{fp}) - \frac{\theta_{fp}}{1.1775421}$

Model 8: $\frac{(\sin(0.020238832 - (x_{fp} \cdot 3.471492 \times 10^{-5})) \cdot x_{fp}) - \theta_{fp}}{1.1779437}$

Model 9: $\frac{(x_{fp} \cdot (0.0202491 - (x_{fp} \cdot 3.777832 \times 10^{-5})) - \theta_{fp})}{1.1780322} + 0.016554512$

Model 10: $\frac{(x_{fp} \cdot (0.020234715 - (3.790718 \times 10^{-5} \cdot (-0.33747756 + x_{fp})))) - (\theta_{fp} - 0.01984857)}{1.1781462}$

Model 11: $\frac{((x_{fp} \cdot (0.020234715 - (3.790718 \times 10^{-5} \cdot (\sin(x_{fp}) + x_{fp})))) - \theta_{fp}) + 0.020234715}{1.1781462}$

Model 12: $\frac{(x_{fp} \cdot (0.020234715 - ((x_{fp} + (1.4071028 \sin(x_{fp}))) \cdot 3.790718 \times 10^{-5}))) - (\theta_{fp} - 0.020234715)}{1.1781462}$

Model 13: $\frac{((x_{fp} - (x_{fp} \cdot (x_{fp} \cdot 0.0018827785)))) \cdot 0.020252343) - (\theta_{fp} - 0.025055747)}{1.1766618 - (\theta_{fp} \cdot 0.00651754)}$

Model 14: $((x_{fp} - ((x_{fp} \cdot (x_{fp} - (\sin(\theta_{fp}) / 0.11032114)))) \cdot 0.001999097)) \cdot 0.017209781) - \frac{(\theta_{fp} - 0.017209781)}{1.1744026}$

Model 15: $\frac{((x_{fp} - ((x_{fp} - \theta_{fp}) \cdot (0.0019173961 \cdot x_{fp}))) \cdot 0.020258129) - (\theta_{fp} + -0.025679091)}{1.1759938 - (\theta_{fp} \cdot 0.008254464)}$

Model 16: $\frac{(((x_{fp} - (x_{fp} \cdot (x_{fp} \cdot 0.0018712947)))) \cdot 0.020301536) - \theta_{fp}) - -0.02489745}{1.1736315 - ((\theta_{fp} - \cos(\theta_{fp})) \cdot 0.007410869)}$

Model 17: $\frac{((x_{fp} - (x_{fp} \cdot (0.0018818954 \cdot x_{fp}))) \cdot 0.020245517) - (\theta_{fp} - 0.033733834)}{1.1746142 - ((\theta_{fp} - (x_{fp} \cdot 0.019458424)) \cdot 0.020245517)}$

Model 18: $\frac{(0.020245517 \cdot (x_{fp} - (x_{fp} \cdot (x_{fp} \cdot 0.0018818954)))) - (\theta_{fp} - 0.033733834)}{1.1746142 - ((\theta_{fp} - ((x_{fp} - -0.44894302) \cdot 0.019458424)) \cdot 0.020297563)}$

Model 19: $\frac{((x_{fp} \cdot 0.020190464) - (((\sin(x_{fp}) - \theta_{fp}) / 0.17703694) + x_{fp}) \cdot x_{fp}) \cdot 4.234294 \times 10^{-5}) - (\theta_{fp} - 0.02910507)}{1.1762471 - (\theta_{fp} \cdot 0.015812252)}$

5 Zaključek

V tej nalogi smo se podrobneje spoznali s fitanjem modelskih parameterov in primerjave modelov med seboj. Začeli smo s preprostim modelom kemijskih reakcij in čistilnosti ledvice. Za konec pa smo si pogledali bolj kompleksen primer kalibracije visokoločljivostnega magnetnega spektrometra, kjer smo sami postavili model transformacije disperzijskega kota na tarči. Za konec pa smo poskusili to storiti še s pomočjo simbolične regresije.