

Določanje elastičnih konstant nematskega tekočega kristala

3. domača naloga pri predmetu *Praktikum strojnega učenja v fiziki*

Domen Vaupotič

1 Nematski tekoči kristali

V nalogi obravnavamo določanje elastičnih konstant nematskih tekočih kristalov. Deformacije direktorskega polja razdelimo v tri različne deformacijske načine (pahljačasti, zvojni in upogibni), tako da prosto energijo opišemo kot:

$$f_{\text{elast.}} = \frac{1}{2}K_1(\nabla \cdot \mathbf{n})^2 + \frac{1}{2}K_2(\mathbf{n} \cdot \nabla \times \mathbf{n})^2 + \frac{1}{2}K_3(\mathbf{n} \times \nabla \times \mathbf{n})^2,$$

pri čemer so K_1, K_2 in K_3 elastične konstante. Če tekoči kristal izpostavimo zadosti velikemu zunanemu električnemu polju, nastopi Frideriksov pojav, ki k energiji doda člen:

$$f_{\text{elektr.}} = -\frac{1}{2}\varepsilon_0\varepsilon_a(\mathbf{n} \cdot \mathbf{E})^2.$$

Elastičnost tekočih kristalov opazujemo posredno prek spremembe optičnih lastnosti. Zaradi dvolomnosti nastopata v kristalu dva lomna količnika, redni n_o in izredni n_e , v splošnem pa lomni količnik odvisen od vpadnega kota svetlobe:

$$n(\theta)^2 = \left[\frac{n_e^2}{\cos^2 \theta} + \frac{n_o^2}{\sin^2 \theta} \right]^{-1}.$$

Svetloba med potovanjem skozi kristal nabira fazno razliko, ki se po celotnem prehodu sešteje v:

$$\Gamma = \frac{2\pi}{\lambda} \int_0^d (n(\theta(z)) - n_o) dz.$$

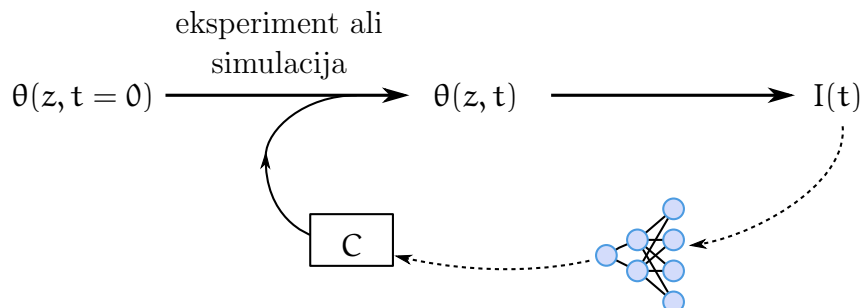
Če modeliramo direktorsko polje kristala v ravnini v približku enakih elastičnih konstant ($K_1 = K_2 = K_3 = K$), pridobimo relaksacijsko (difuzijsko) enačbo:

$$C \frac{\partial^2 \theta(z, t)}{\partial z^2} = \frac{\partial \theta(z, t)}{\partial t},$$

pri čemer je sorazmernostna konstanta $C = K/\gamma$. Časovnega razvoja funkcije $\theta(z, t)$ eksperimentalno ne poznamo, izmerimo pa lahko časovno spreminjanje intenzitete prepuščene svetlobe skozi vzorec:

$$I(t) = \sin^2(\Gamma(t)/2).$$

Osrednja ideja naloge, prikazana na sliki 1, je torej simulirati množico različnih relaksacij direktorskih profilov pri znanih vrednostih konstante C in z njimi naučiti nevronske mrežo, s katero bi lahko nato določili elastično konstanto tudi za eksperimentalne podatke.

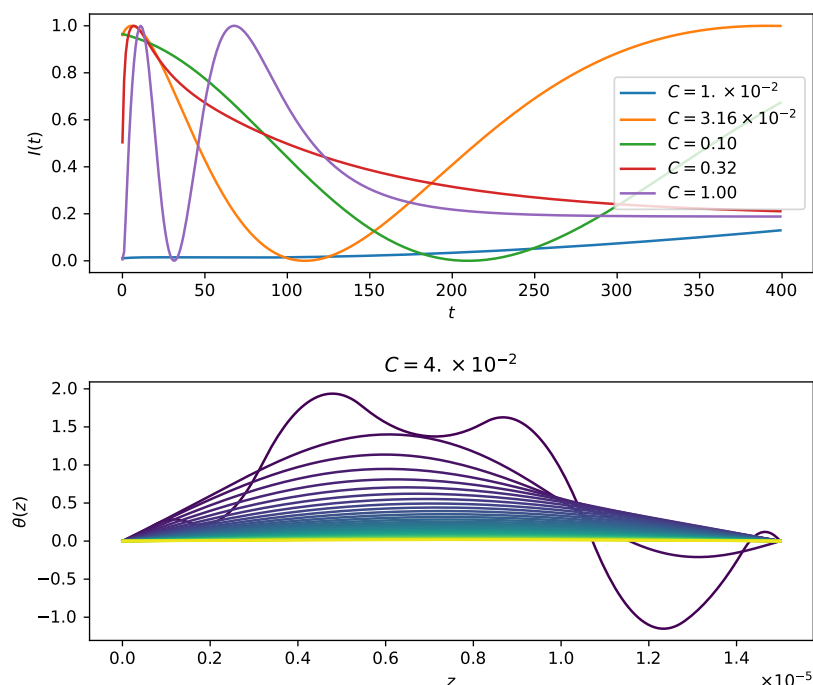


Slika 1: Osrednja ideja naloge.

2 Rezultati

2.1 Časovni profili

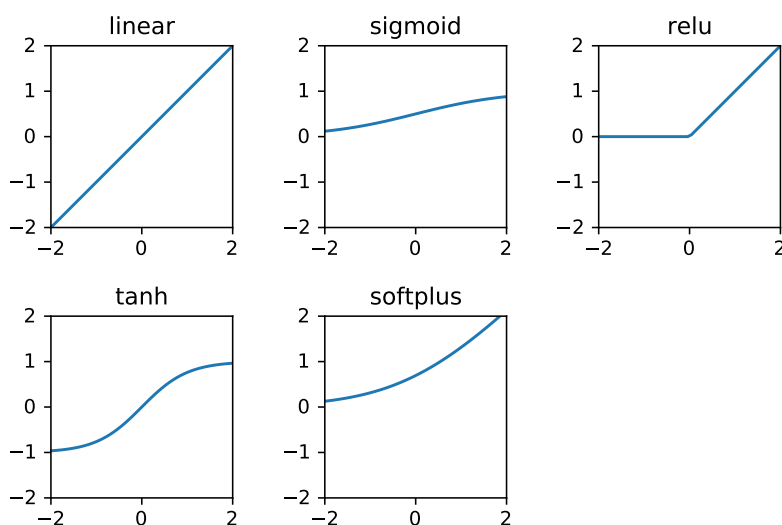
Oglejmo si za začetek nekaj primerov vhodnih podatkov. Na sliki 2 so prikazani intenzitetni profili in časovno spreminjanje direktorskega polja, kot je bilo izračunano s simulacijo.



Slika 2: Primeri intenzitetnih profilov (zgoraj) in časovna relaksacija direktorskega polja (spodaj).

2.2 Nevronska mreža

Preden se posvetimo nevronskim mrežam, si na sliki 3 ogledamo nekaj primerov aktivacijskih funkcij. Opazimo, da se med seboj razlikujejo predvsem po svoji zalogi vrednosti. Linearna aktivacijska funkcija je razmeroma neuporabna, saj ima konstanten odvod in tako ne moremo uporabiti vzratnega učenja. Linearna funkcija prav tako združi vse skrite plasti v eno plast, s čimer zgubimo želeno kompleksnost. Aktivacijske funkcije so zato nelinearne, najpogosteje uporabljene so sigmoidna, katere slabost je izginjanje gradienta pri ekstremnih vrednostih; hiperbolični tangens, ki je centriran in primeren za vhodne podatke, ki so tako močno pozitivni kot močno negativni; relu, ki je računsko učinkovita, a ima v negativnem območju ponovno ničeln odvod, ter njena gladka aproksimacija relu.



Slika 3: Primeri pogosteje uporabljenih aktivacijskih funkcij.

Začeli bomo z najenostavnejšo nevronske mrežo, ki jo sestavlja zgolj 400 vhodnih nevronov in 1 izhodni nevron, med seboj polno povezani, s sigmoidno izhodno aktivacijsko funkcijo. Zanj se odločimo, ker imamo podatke skalirane na območje od 0 do 1, seveda pa ni fizikalnega motiva, da bi konstanto C kakorkoli navzgor omejili, zato bi se lahko poslužili tudi drugačne aktivacijske funkcije.

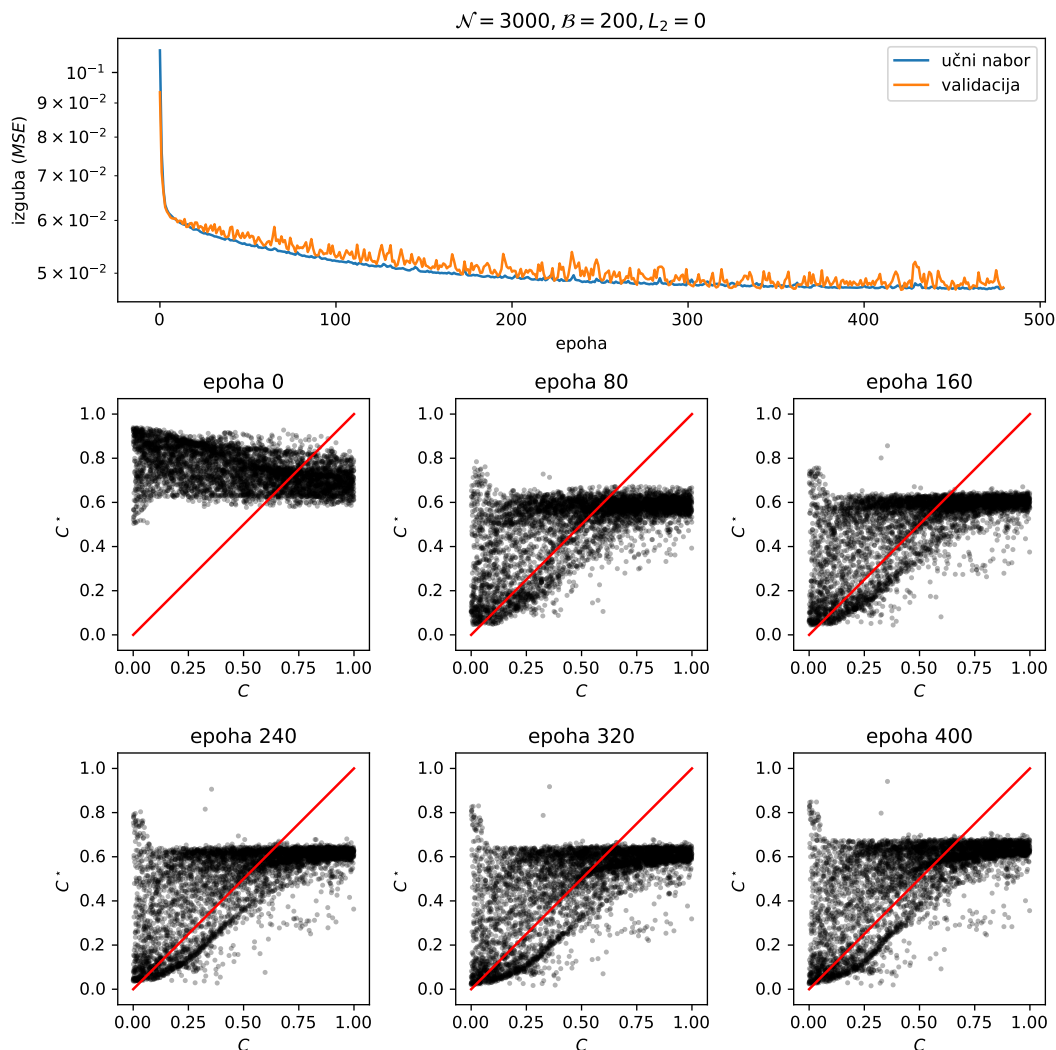
$$400 \xrightarrow{\text{sigmoid}} 1$$

Na sliki 4 je prikazana zmogljivost najenostavnejše mreže. Število učnih podatkov označuje $\mathcal{N}$, velikost paketa pa $\mathcal{B}$. Nevronska mreža je uporabljala privzeti optimizator adam. Prikazana je izguba, izračunana kot povprečni kvadrat napake (MSE), na učnem setu in validacijskem setu (ta predstavlja 10 % celotnega nabora $\mathcal{N}$). Na validacijskih grafih je primerjana prava vrednost C z napovedano vrednostjo C^* .

2.3 Vpliv hiperparametrov

Ker najenostavnejša mreža ni dovolj, dodajmo sedaj skrito plast velikosti L_2 nevronov in z aktivacijsko funkcijo $\mathcal{A}$.

$$400 \xrightarrow{\mathcal{A}} L_2 \xrightarrow{\text{sigmoid}} 1$$



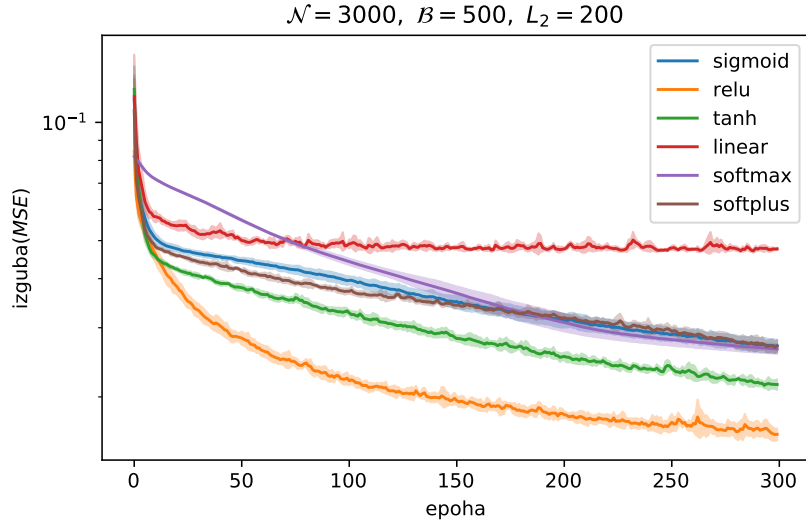
Slika 4: Potek učenja (zgoraj) in validacija na neodvisnih podatkih (spodaj) za najenostavnejšo nevronske mrežo.

Na slikah 5–7 je prikazan vpliv izbire hiperparametrov, in sicer aktivacijske funkcije druge plasti $\mathcal{A}$, števila nevronov v drugi plasti L_2 in velikosti paketa (*batch size*) $\mathcal{B}$. Prikazan je tudi povprečni čas učenja. Izračun je bil izveden s petkratno navzkrižno validacijo (KFold), na grafih je z močno črto prikazano povprečje, okoli njega pa svetleje pobarvana standardna napaka.

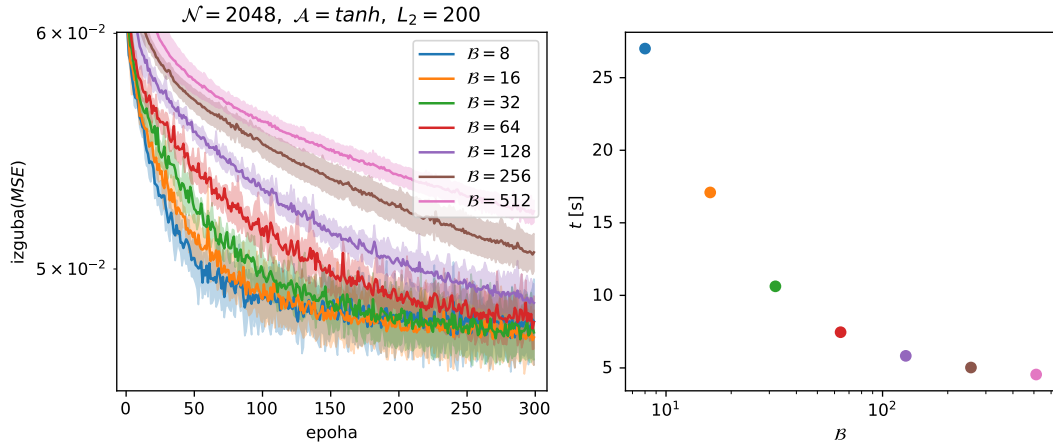
Opazimo, da se linearna aktivacijska funkcija odreže najslabše, medtem ko dasta najboljše rezultate funkciji relu , tanh . Funkcije softmax , softplus in sigmoid so se do izbrane končne epohe primerljivo izkazale.

Pomemben hiperparameter je velikost učnega paketa (*batch size*). Če v nevronske mrežo podajamo po en učni vhod naenkrat, govorimo o *stohastičnem gradientnem spustu*, če pa nevronske mrežo učimo z več učnimi podatki naenkrat, govorimo o *minibatch gradientnemu spustu*. Opazimo, da je stohastični spust veliko bolj zašumljen, a hitreje konvergira, medtem ko večji učni paketi vodijo v počasnejšo, a gladkejšo konvergenco. Na prvi pogled bi torej nevronske mrežo želeli učiti s čim manjšimi učnimi paketi, a to seveda vodi v znatno povečanje računskega časa.

Posvetimo se še velikosti skrite plasti. Opazimo, da že zgolj nekaj nevronov v drugi



Slika 5: Vpliv aktivacijske funkcije na potek učenja nevronske mreže.

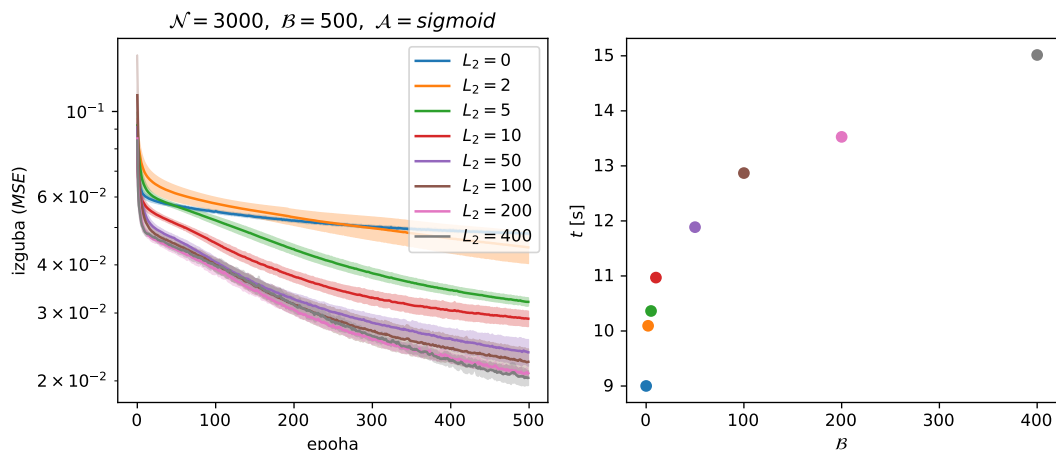


Slika 6: Vpliv velikosti paketa na potek učenja nevronske mreže (levo) in povprečni čas učenja (desno).

plasti (npr. $L_2 = 5$) znatno izboljša napoved nevronske mreže. Pri zadosti velikem številu (npr. $L_2 = 200$) opazimo, da se s povečevanjem števila nevronov mreža ne izboljšuje več znatno. Primerjava je seveda izvedena pri istem številu epoh, čeprav bi bilo bolj smotrno primerjavo izvajati tako, da bi nevronske mrežo učili do njenih končnih zmogljivosti, ko se napaka na validacijskem setu več ne zmanjšuje. Seveda opazimo, da se s povečevanjem števila nevronov povečuje tudi čas učenja.

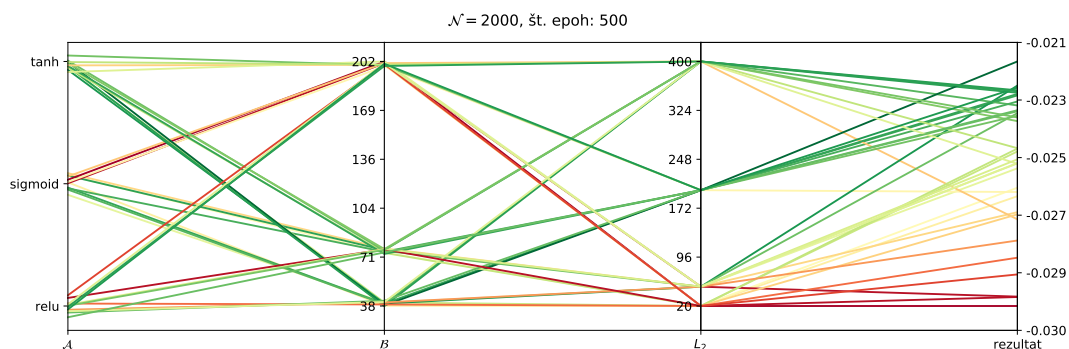
2.4 Optimalni hiperparametri

Da bi izbrali optimalno kombinacijo hiperparametrov, smo izvedli mrežno iskanje. Rezultati za vsako kombinacijo hiperparametrov so bili izračunani s trikratno navzkrižno validacijo in so prikazani na sliki 8, pri čemer vsaka lomljenka predstavlja eno kombinacijo hiperparametrov, njena barva pa ustreza napaki nevronske mreže. Najboljši rezultat je dala kombinacija $\mathcal{A} = \tanh, \mathcal{B} = 40, L_2 = 200$. Sicer podobno kot prej opazimo, da



Slika 7: Vpliv števila nevronov v skriti plasti na potek učenja nevronske mreže (levo) in povprečni čas učenja (desno).

manjše število nevronov v drugi plasti daje slabše rezultate, prav tako pa se je v splošnem funkcija tanh odrezala bolje od sigmoidne in relu.



Slika 8: Rezultat mrežnega iskanja optimalnih hiperparametrov.

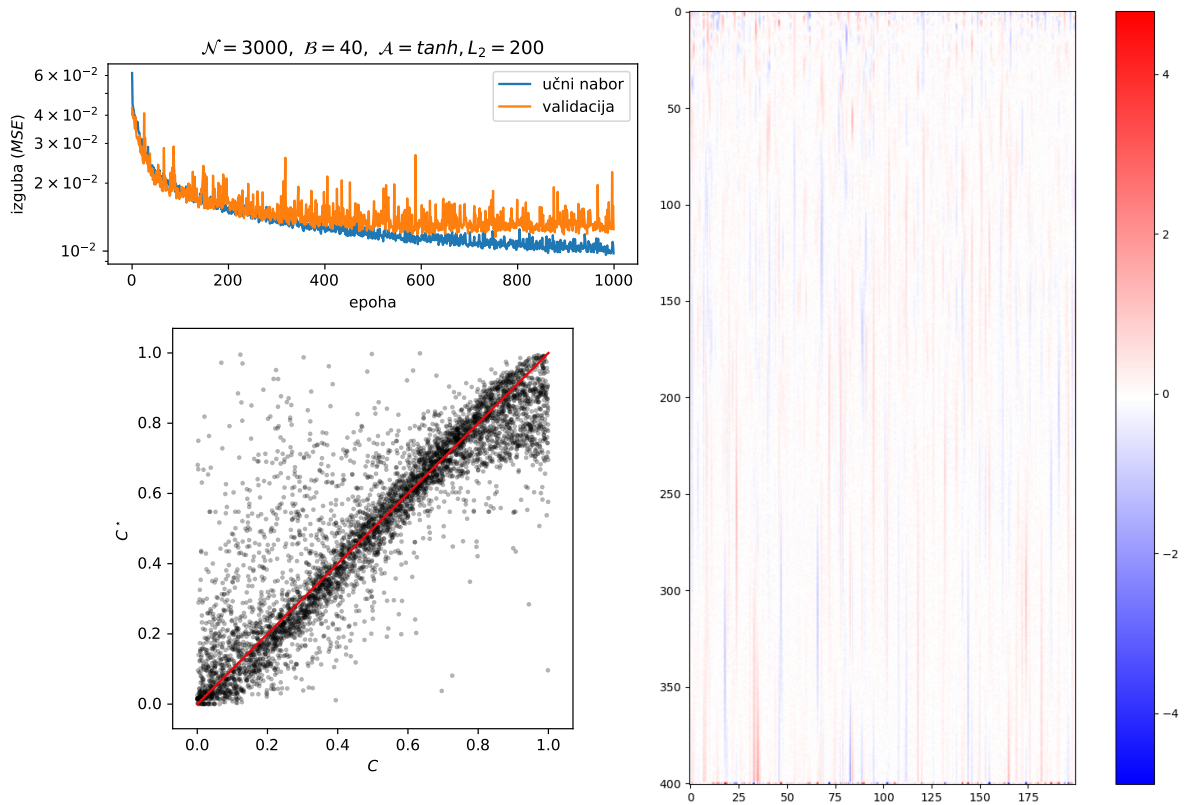
Na sliki 9 je prikazan potek učenja za mrežo z optimalnimi hiperparametri in končne uteži. Opazimo, da je napoved sicer dobra, a se kljub temu pojavljajo še zelo odstopajoči osamelci. Izbrali smo tri najslabše določene intenzitetne profile in jih prikazali na sliki 10 ter jih primerjali s tremi najboljše določenimi. Z enostavno okometrično analizo nismo mogli določiti kakšne kvalitativne razlike med najboljše in najslabše izbranimi profili.

2.5 Določanje dveh elastičnih konstant

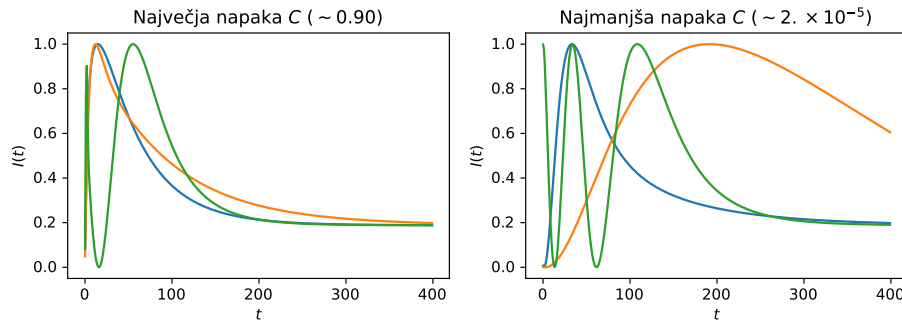
Pri izpeljavi v prvem delu smo uporabili približek enakih elastičnih konstant. Model lahko nekoliko nadgradimo, tako da izenačimo le dve elastični konstanti. Mrežo sedaj želimo naučiti določiti dve elastični konstanti.

$$400 \xrightarrow{\mathcal{A}} L_2 \xrightarrow{\text{sigmoid}} 2$$

Na sliki 11 so prikazani rezultati učenja nevronske mreže. Na grafu poteka učenja lepo vidimo, da se je učenje mreže ustavilo že pri približno 100. epohi, kar pomeni, da je končna mreža bila prenasajena na vhodne podatke (*overfit*). V splošnem je obe konstanti dobro



Slika 9: Potek učenja mreže z optimalnimi hiperparametri (levo zgoraj), validacija na neodvisnih podatkih (levo spodaj) in prikaz uteži (desno).

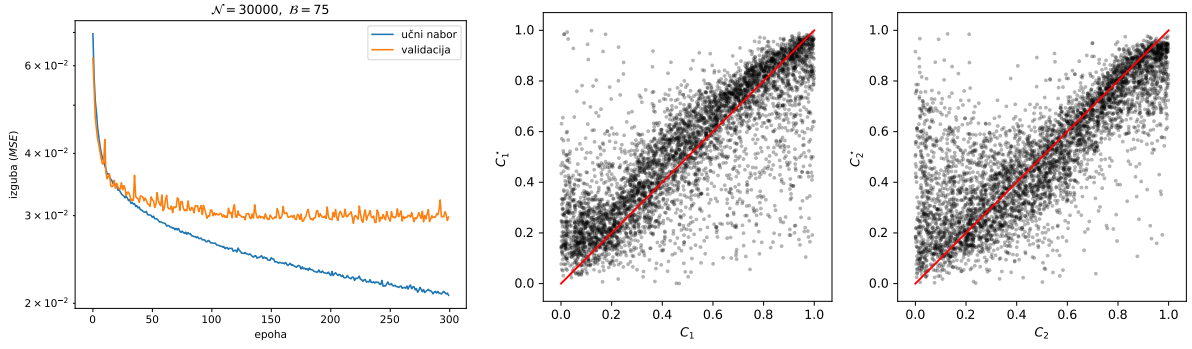


Slika 10: Primer treh najslabše (levo) in najboljše (desno) določenih intenzitetnih profilov pri mreži z optimalnimi hiperparametri.

določila, vendar pri konstanti C_1 opazimo sistematično konkavno ukrivljenost (mreža napoveduje prevelik C_1), pri konstanti C_2 pa večjo razpršenost.

2.6 Večspektralni vhodni podatki

Prepustnost svetlobe skozi vzorec nematskega tekočega kristala lahko merimo pri različnih valovnih dolžinah. V ta namen smo preverili, ali se napoved nevronske mreže izboljša, če kot vhodne podatke vstavimo meritve pri različnih valovnih dolžinah. Da bi bila primerjava smiselna, smo ohranili število vhodnih nevronov (400), kar pomeni, da smo podatke pri posamezni valovni dolžini vzorčili redkeje, nato pa vse podatke zlepili v en vektor.



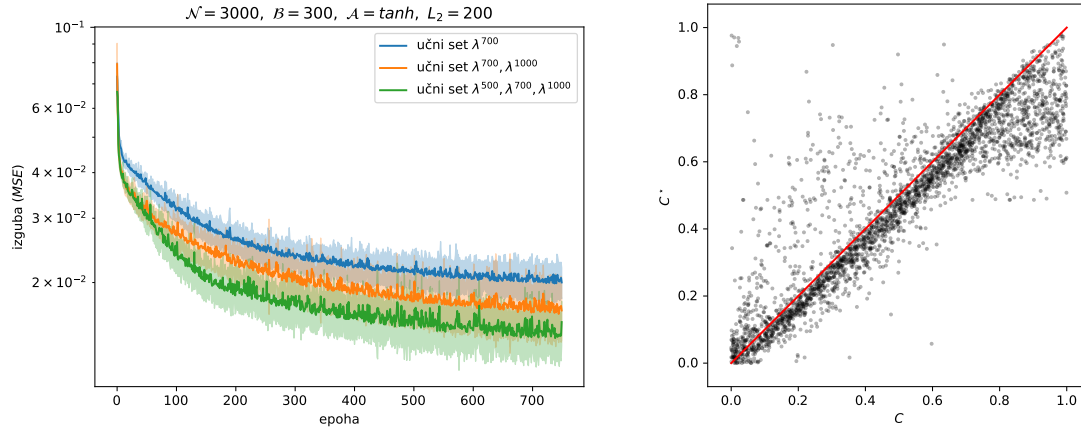
Slika 11: Potek učenja (levo) in validacija (desno) nevronske mreže za določanje dveh elastičnih konstant.

$$400_{\lambda^{700}} \xrightarrow{\mathcal{A}} L_2 \xrightarrow{\text{sigmoid}} 1$$

$$200_{\lambda^{700}} + 200_{\lambda^{1000}} \xrightarrow{\mathcal{A}} L_2 \xrightarrow{\text{sigmoid}} 1$$

$$133_{\lambda^{500}} + 133_{\lambda^{700}} + 134_{\lambda^{1000}} \xrightarrow{\mathcal{A}} L_2 \xrightarrow{\text{sigmoid}} 1$$

Rezultati učenja so prikazani na sliki 12. Zares, opazimo, da z večspektralnimi vhodnimi podatki nevronska mreža daje boljše napovedi.



Slika 12: Potek učenja (levo) in validacija na neodvisnem vzorcu (desno) za večspektralne vhodne podatke.

3 Zaključek

V nalogi smo sestavili različne enostavne nevronske mreže z gosto povezanimi plastmi, ki smo jih uporabili za napovedovanje elastične konstante nematskih tekočih kristalov. Določili smo nekaj optimalnih hiperparametrov, ki dajejo najboljše rezultate, prav tako pa smo ugotovili, da nevronska mreža napoveduje bolje, če uporabimo večspektralne vhodne podatke. V nalogi nismo obravnavali podatkov s simuliranim šumom, prav tako smo preverjali zgolj mreže z eno skrito plastjo ali preverjali vpliva različnih optimizatorjev.