

Identifikacija faznih prijelaza strojnim učenjem i metodom konfuzije

Ivan Emanuel Pavlov

Mentor: dr.sc. Vinko Zlatić

Fizički odsjek, Prirodoslovno-matematički fakultet, Bijenička 32, Zagreb

26.1.2020

Sažetak

Proučena su dva modela na 2d rešetki: Isingov model te XY model. Korištene su neuronske mreže za lokaciju faznih prijelaza. Korištena je metoda (metoda konfuzije) za označavanje skupa podataka čime je omogućeno proučavanje faznog prijelaza kao da je točna temperatura prijelaza nepoznata. Takva metoda omogućava daljnje proučavanje modela gdje fazni prijelazi nisu poznati.

1 Uvod

Zadnjih godina u polju strojnog i dubokog učenja su polučeni brojni uspjesi. Strojno učenje se uspješno primjenjuje u raznim poljima fizike.⁸ Jedno interesantnih područja za primjenu strojnog učenja je fizika čvrstog stanja, partikularno faznih prijelaza. Mali broj teoretskih modela je analitički rješiv no tome se može uskočiti numeričkim metodama. Često je teško pronaći relevantne stupnjeve slobode koji dobro opisuju sustav. Algoritmi strojnog učenja takve stupnjeve slobode pronalaze te ih nazivamo značajkama ("features") modela. Strojno učenje omogućava i validaciju prijašnjih pristupa lokacijom već poznatih faznih prijelaza. No, u ovom radu obrađuje se metoda⁴ kojom je moguće pronaći fazne prijelaze čak i u slučajevima gdje se ne zna gdje se oni nalaze bez detaljnog poznavanja fizike modela.

2 Fizikalni modeli

2.1 2d Ising model¹²

Hamiltonijan 2d Ising modela bez vanskog magnetskog polja te s jednakom jačinom interakcije susjednih spinova definiran na $L \times L$ rešetci je

$$H = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j$$

σ varijable označavaju spinove u partikularnim ćelijama rešetke dok je J konstanta vezanja susjednih spinova i u ovom slučaju je odabrana kao jednaka za sva međudjelovanja spinova. Svaka ćelija međudjeluje samo sa susjednim ćelijama. Svaka ćelija ima 4 susjeda tako da će energetske doprinose svake ćelije biti četverostruk. Spinovi mogu poprimiti vrijednosti 1 ili -1, što znači da susjedi mogu biti samo paralelni ili antiparalelni.

Vjerojatnost da se sustav nađe u stanju μ

$$P_\mu = \frac{1}{Z} e^{-\frac{1}{k_B T} E_\mu}$$

gdje je Z particijska funkcija $Z = \sum_\mu e^{-\frac{1}{k_B T} E_\mu}$.

Isingov model je zanimljiv jer je analitički rješiv za 1d i 2d slučaj. U 2d slučaju postoji dugodosežno uređenje na temperaturama ispod temperature faznog prijelaza. Krična temperatura modela je

$$T_c = 2.269 \frac{J}{k_B}$$

Kako Isingov model ima 2 faze, za njih vrijedi uobičajeno ponašanje u blizini faznog prijelaza. Očekuju se široke domene usmjerenih spinova (čime sustav teži minimalizirati energiju sustava) na niskim temperaturama popraćenima s velikim fluktuacijama kako se temperatura bliži faznom prijelazu. Nakon T_{krit} sustav postaje neuređen jer je korelacijska duljina konačna pa je onemogućeno je dalekosežno spinsko uređenje. Sustav teži maksimizirati entropiju sustava.

2.2 2d XY model,¹²

XY model je specijalni slučaj Heisenbergovog modela spinskog međudjelovanja u kojemu su spinovi usmjereni u 2d ravni (x-y). XY hamiltonijan možemo zapisati kao

$$H = -J \sum_{\langle i,j \rangle} \vec{s}_i \vec{s}_j = -J \sum_{\langle i,j \rangle} \cos(\theta_i - \theta_j)$$

Kao i u slučaju 2d Ising modela, J konstanta je konstanta vezanja. Za razliku od 2d Ising modela, u slučaju XY modela spinovi nisu paralelni ili anti-paralelni već

mogu biti proizvoljno usmjereni u X-Y ravnini. Varijable koje onda definiraju energetske doprinose sustavu jednog čvora rešetke npr. i, j su razlike kutova susjednih čvorova rešetke, partikularno kosinus razlike kuta na poziciji i, j i jednog od četiri susjeda na pozicijama $(i+1, j)$, $(i-1, j)$, $(i, j+1)$, $(i, j-1)$. Naravno, energetske doprinose čvora je suma sva četiri doprinosa interakcije sa susjedima. U 2d XY modelu ne postoji dalekosežno uređenje za $T > 0$ no postoji prijelaz na konačnoj temperaturi. 2d XY model posjeduje poznati Kosterlitz-Thouless fazni prijelaz za koji vrijedi

$$\frac{T_c}{J} = 0.893 \pm 0.002K$$

Taj fazni prijelaz spada u klasu topoloških faznih prijelaza. Mermin-Wagnerov teorem pokazuje da u 2d ne može biti spontanog slamanja kontinuirane simetrije na bilo kojoj konačnoj temperaturi. Teorem vrijedi samo za kontinuirane simetrije a ne diskretne pa zato Isingov model može imati uređenu fazu ispod T_{krit} . No, iako je sustav nema uređenu fazu, ispod kritične temperature sustav se ponaša kvazi-uređeno. Kritična temperatura je temperatura (kao i u slučaju Isingovog modela npr.) na kojoj korelacijska duljina divergira (u smjeru nižih temperatura) te ostaje beskonačna za temperature ispod kritične. To znači da korelacije među spinovima opadaju eksponencijalno (kao i u Isingovom modelu) s udaljenosti za temperature iznad kritične

$$G(r) \propto e^{-\frac{r}{\zeta}}, r \gg \zeta$$

gdje je ζ korelacijska duljina.

Ispod kritične temperature korelacijska funkcija u slučaju XY modela ponaša kao

$$G(r) \propto \frac{1}{r^\eta}$$

gdje je η neki kritični eksponent.

Spinovi u rešetci se slažu u spinske virove. Oni su tip topološkog defekta. Susjedni spinovi se okreću diskretni broj puta (svaki okret za neki diskretni kut) u prostoru. Takvi virovi su energetske skupi jer spinovi nisu u istom smjeru. Na niskim temperaturama se virovi poništavaju s antivirovima te nema ukupne virovitosti u sustavu pa tvore kvazi uređen sustav, dok na temperaturama iznad faznog prijelaza virovi tvore plazmu virova jer postoje slobodni neponišteni virovi.

3 Metode

3.1 Metoda konfuzije

Da bi ilustrirali metodu konfuzije⁴ poslužiti ćemo se 2D Ising modelom. Uzimimo neku temperaturu T'_c za koju pretpostavljamo da je kritična temperatura

T_c . Tada ćemo sve konfiguracije na temperaturama manjima od T'_c označiti kao faza 1 dok one na temperaturama većima od T'_c kao fazu 2.

Naravno ovo će biti sasvim točno samo ako $T'_c = T_c$ dok će za međuvrijednosti ovakvo označavanje dati dio krivih oznaka faze sustava.

U 2d Isingovom modelu znamo da je faza 1 uređena feromagnetska faza, dok je faza 2 paramagnetska neuređena faza, no za funkcionalnost metode je ključno da postoje barem 2 faze koje imaju jasna obilježja.

Ideja je locirati kritičnu temperaturu, kao da ju ne znamo iz teoretskih razmatranja.

Neuronska mreža će kao ulaz primiti konfiguracije iz Monte Carlo metode te će uz njih kao dio ulaza dobiti i oznaku faze kako je gore opisano. Pretpostavimo sada da je

$$0 < T'_c < T_c$$

Pretpostavljena kritična temperatura je niža od one stvarne. Gore opisanom metodom označavanja faza sve konfiguracije na temperaturi manjoj od T'_c će dobiti oznaku faze 1, što odgovara stvarnosti, dok će one veće od T'_c dobiti oznaku faze 2. Ovdje vidimo da to nije posve točno. Naime, u intervalu $[T'_c, T_c]$ podaci bi trebali dobiti oznaku faze 1 dok su u ovom slučaju dobili oznaku faze 2. Algoritam strojnog učenja, npr neuronska mreža, će u slučaju da se iz podataka mogu izlučiti dobre značajke koje mogu razlučiti razlike fazama biti sposoban uočiti gore opisanu nepravilnost. Naime, možemo reći da će algoritam biti "zbunjen". Jer oznake će mu govoriti da se radi od jednoj fazi, dok će značajke modela ukazivati da se radi o drugoj fazi. Pretpostavka je da dakle će algoritam ipak dobro naučiti klasificirati faze unatoč "zbunjenosti" tj. krivom označavanju skupa podataka.

Da bi vidjeli zašto to ima smisla, pretpostavimo da je $T=2.0K$. Sve konfiguracije u temperaturnom intervalu $[2.0K, 2.26...K]$ će biti krivo označene s fazom 2. No, ipak je puno više konfiguracija točno označeno (sve manje od $2.0K$) pa je logično da će algoritam ipak na osnovu dobro označenih podataka naučiti dobro klasificirati faze ako su značajke modela dovoljno dobre.

3.2 Monte Carlo metoda¹

Algoritam kao ulaz koristi klasične spinske konfiguracije generirane pomoću Monte Carlo simulacije. U slučaju spinskih sustava, Monte Carlo simulacije služe za generiranje konfiguracija sustava prema Boltzmanovoj distribuciji. U ovom radu se koristi Metropolis metoda. Opisati ćemo ju na partikularnim primjerima.

3.2.1 Isingov model

Prvo generatorom slučajnih brojeva generiramo inicijalnu konfiguraciju modela za neku partikularnu temperaturu. Vrijednosti veće od 0.5 promijenimo u 1, dok one manje od 0.5 u -1. U slučaju Isingovog 2d modela inicijalna konfiguracija je mreža veličine $L \times L$, a vrijednosti u čvorovima su 1 ili -1, što su naravno

vrijednosti spina u čvorovima mreže. Za generaciju ravnotežne spinske konfiguracije korišten je Metropolis algoritam koji funkcionira na sljedeći način.

Izaberemo jedan čvor $S[i, j]$ te izračunamo energetske doprinose tog čvora sustavu

$$dE(i, j) = -J \cdot S(i, j) \cdot (S(i+1, j) + S(i-1, j) + S(i, j+1) + S(i, j-1))$$

Okrećemo spin na poziciji i, j

$$S(i, j) = -S(i, j) \Rightarrow dE'(i, j) = -dE(i, j)$$

To znači da je razlika energija prije i poslije

$$\Delta E(i, j) = -2 \cdot dE(i, j)$$

Generiramo slučajni broj p_{random} za koji vrijedi $0 \leq p_{random} \leq 1$.

$$\begin{aligned} \text{Ako } \Delta E(i, j) < 0 &\Rightarrow S(i, j) = -S(i, j) \\ \text{Ako } \Delta E(i, j) > 0 &\Rightarrow \begin{cases} \text{Ako } e^{\frac{-\Delta E}{k_B T}} > p_{random} &\Rightarrow S(i, j) = -S(i, j) \\ \text{Inače} &\Rightarrow S(i, j) = S(i, j) \end{cases} \end{aligned}$$

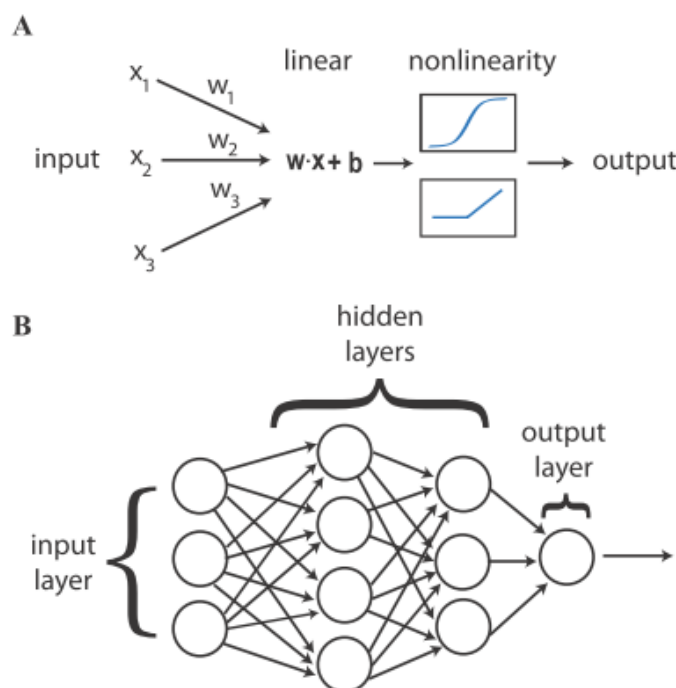
Taj proces ponavljamo velik broj puta (20000 u ovom radu). Konačnu konfiguraciju koristimo kao ulazni podatak neuronske mreže. Isti proces ponavljamo odabran broj puta (u radu 2000) tako da bi generirali dovoljan broj trening primjera na svakoj temperaturi koji se koriste kao ulazni podaci neuronske mreže. Algoritam se provodi za više odabranih temperatura u temperaturnom rasponu (u ovom radu 20 temperatura u rasponu od $[0.01, 2T_{krit}]$).

3.2.2 XY model

Procedura je slična onoj za Isingov model. Generiramo slučajne brojeve u rasponu $[-2\pi, 2\pi]$ kao inicijalnu konfiguraciju sustava s na rešetci veličine $L \times L$. Odabiremo jedan čvor rešetke te računamo energijski doprinos tog čvora cijelom sustavu. Zatim generiramo slučajni kut te ga postavimo na mjesto originalnog kuta te izračunamo novi energijski doprinos tog čvora sustavu. Zatim računamo razliku energija nakon i prije promjene kuta. Ostatak procedure je identičan onomu u slučaju Isingovog modela. Konačnu konfiguraciju koristimo da bi napravili novu listu vrijednosti za svaki čvor, gdje je u svakom čvoru pridružena vrijednost kosinusa i sinusa kuta. To su ulazni podaci neuronske mreže.

3.3 Neuronska mreža⁷⁹

Neuronska mreža je model strojnog (dubokog) učenja koji je sposoban iz dovoljne količine podataka naučiti netrivialna i često nelinearna svojstva sustava čiji input primaju. U principu neuronska mreža sa samo jednim skrivenim slojem može aproksimirati bilo koju nelinearnu kontinuiranu funkciju ("universal approximation theorem"⁷).

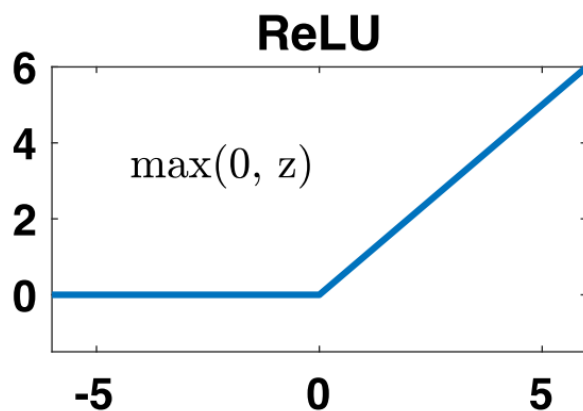


Slika 1: Skica jedne čelije neuronske mreže te obične potpuno povezane neuronske mreže

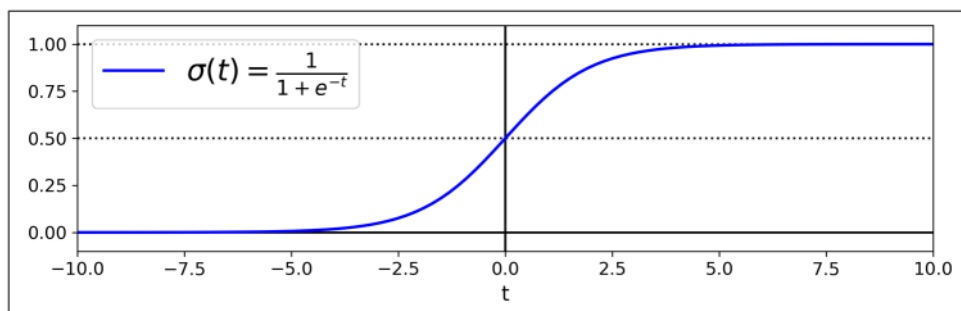
Na slici 1 vidimo kako izgleda jedna tipična čelija neuronske mreže⁸. Čelija u skrivenom sloju će dobiti višestruki ulaz te će ih utežiti vrijednostima w_i ($i = 1, 2, \dots, N_{broj\ ulaza}$) te zatim zbrojiti. Na takvu još uvijek linearnu funkciju se primjenjuje tzv. aktivacijska funkcija koja je nelinearna. U ovom radu se u skrivenom sloju mreže koristi ReLu aktivacijska funkcija čije djelovanje je vidljivo sa slike 2.⁸ Ako je vrijednost ulaza manja od 0 izlaz čelije je 0, dok u slučaju da je ulaz veći od 0 vrijednost izlaza čelije linearno ovisi o ulazu.

Druga aktivacijska bitna aktivacijska funkcija koja se koristi kao aktivacijska funkcija izlaznih čelija u klasifikacijskim problemima kao što je riječ u ovom radu je sigmoidna funkcija na slici 3.⁹

Na slici 1 vidimo arhitekturu jedne potpuno povezane neuronske mreže s



Slika 2: ReLu aktivacijska funkcija



Slika 3: Logistička (sigmoidna) funkcija

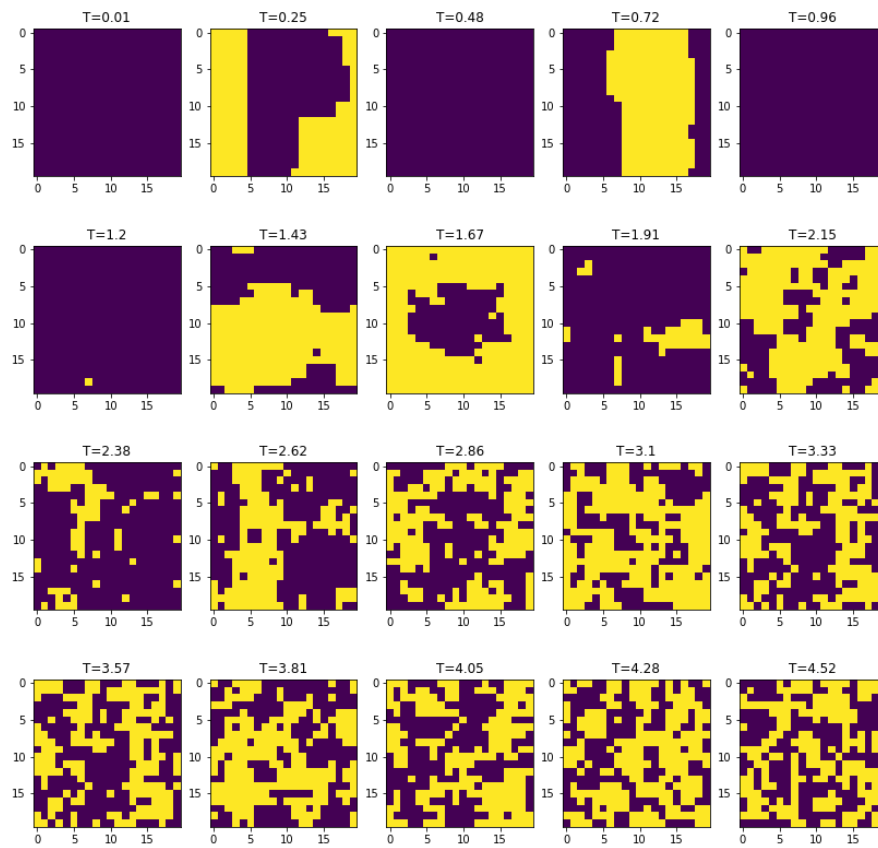
višestrukim ulazom, jednim skrivenim slojem i jednom izlaznom čelijom. Ovakav model bi se mogao koristiti za predviđanje neke kontinuirane vrijednosti (regresijski model). U ovom radu se neuronska mreža koristi za klasificiranje sustava u 2 faze. U takvoj mreži izlaz su 2 čelije sa sigmoidnom aktivacijskom funkcijom u kojoj svaka igra ulogu predviđanja vjerojatnosti pripadnosti ulaza u mrežu (jedne konfiguracije) nekoj od 2 faze.

Mreža uči na velikom broju primjera te kroz trening u smjeru od ulaza do izlaza minimalizira funkciju mjere greške modela ("loss function"). Nakon računanja greške mreža ide unazad te korigira w_i težine u čelijama kroz mrežu koristeći tzv. "backpropagation" algoritam. Kad je trening gotov, koristi se validacijski skup podataka (npr. spinske konfiguracije odvojene od trening skupa prije treninga) za računanje neke metrike modela (kao preciznost npr.) na primjerima koje model nije vidio. Dodatno se može koristiti i testni skup ako su se kroz trening više puta mijenjali hiperparametri modela te je validacijski skup

”zagađen” analogno trening skupu.

4 Rezultati

4.1 2d Ising model



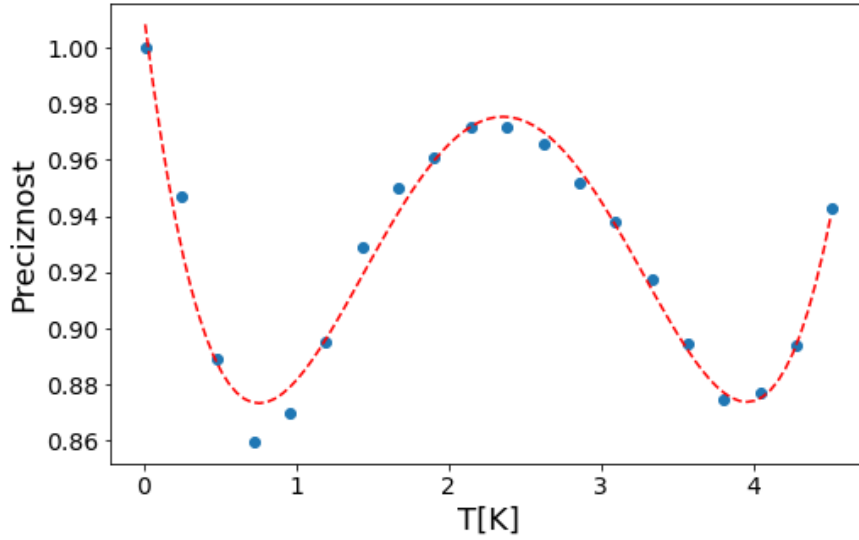
Slika 4: Po jedna odabrana konfiguracija za 20 temperatura

Parametri modela su radi jednostavnosti i bez smanjenja općenitosti postav-

ljeni na

$$J = 1, k_B = 1$$

Pomoću Monte Carlo metode je generirano 2000 konfiguracija za 20 različitih temperatura. Na slici 4 je prikazana po jedna konfiguracija za svaku temperaturu. Neuronsku mrežu čini $L \times L$ ulaza (broj čvorova rešetke), skriveni sloj (100 čelija) te izlazni sloj s 2 čelije sa sigmoidnom aktivacijskom funkcijom koji predviđa vjerojatnost pripadnosti jednoj od faza. Vidimo da konfiguracije možemo podijeliti u one uređene (otprilike do oko 2K), one u kritičnom području te one neuređene (otprilike nakon 2.5K).



Slika 5: Preciznost 20 istreniranih neuronskih mreža za svaki od 20 odabira označavanja skupa podataka

Korištenjem metode konfuzije istrenirano je 20 različitih neuronskih mreža. Svaka mreža odgovara jednom odabiru označavanja skupa podataka opisanom u metodi konfuzije ranije. Dakle, za neku odabranu temperaturu za koju se pretpostavlja da je kritična, sve konfiguracije na temperaturi manjoj od te su dobile oznaku (0,1) dok su one veće dobile oznaku (1,0). Očekivali smo W oblik krivulje i u slučaju Isingovog modela to je približno tako (slika 5). Nelinearni fit na podatke je napravljen radi vizualizacije. Naravno teško je očekivati da će krivulja biti idealnog W oblika. Fazni prijelaz očekujemo u drugom vrhu W krivulje (na sredini). Vrh krivulje je između vrijednosti na $T=2.15$ i $T=2.38$. Sredina između te 2 vrijednosti je na $T=2.265$ što i jest temperatura faznog prijelaza. Da bi dobili precizniji interval unutar kojega se nalazi fazni prijelaz trebalo bi napraviti finiju raspodjelu temperatura, s idealno jednom točkom

točno na temperaturi faznog prijelaza. No, i ovako možemo vidjeti da metoda konfuzije uspješno predviđa fazni prijelaz u rasponu u kojemu se on i nalazi.

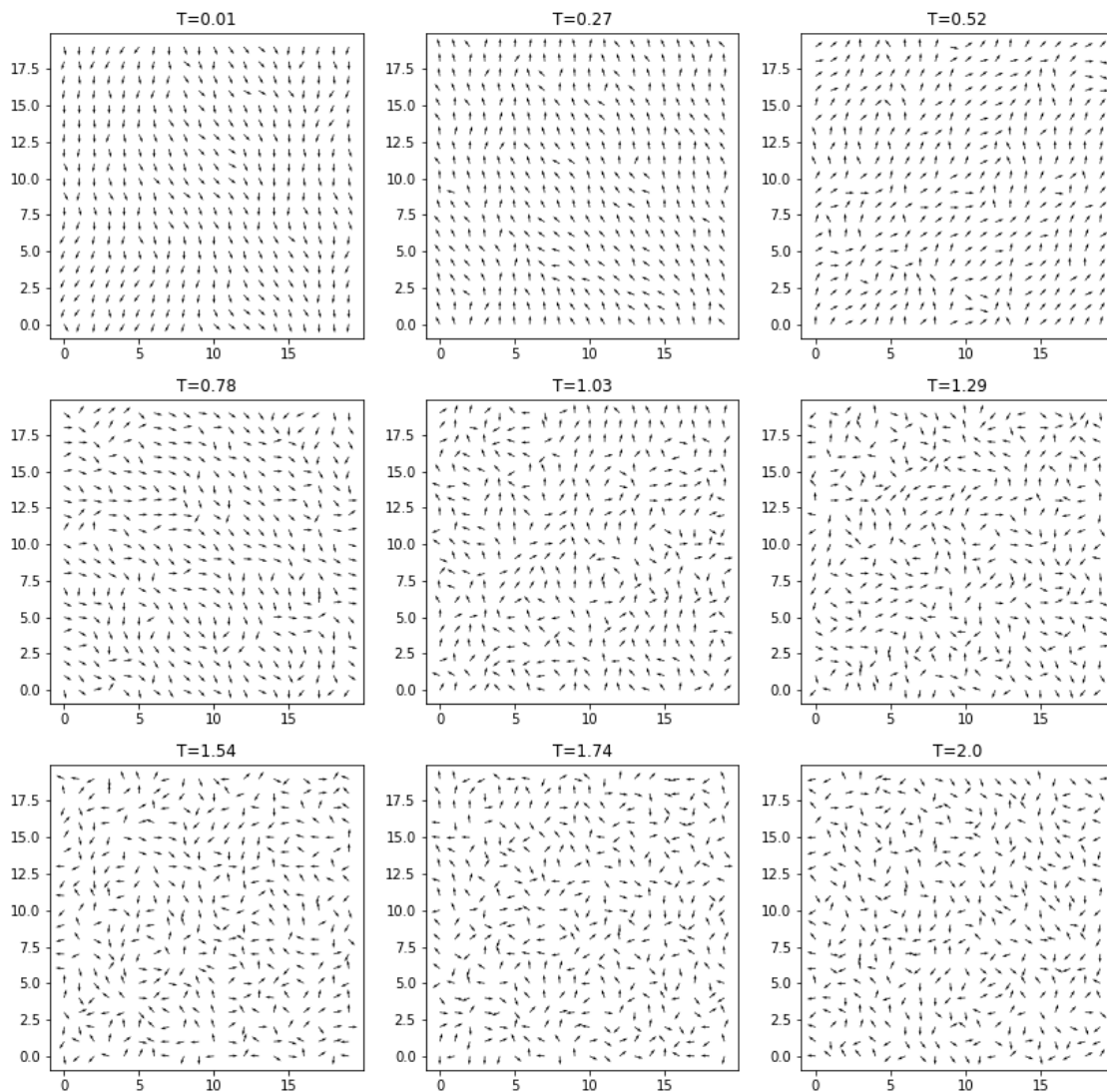
4.2 2d XY model

Parametri modela su radi jednostavnosti i bez smanjenja općenitosti postavljeni na

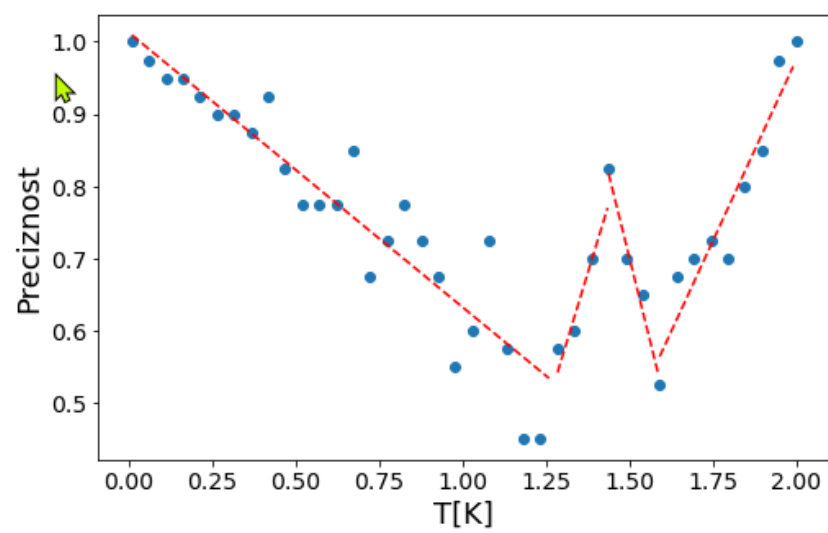
$$J = 1, \quad k_B = 1$$

Monte Carlo metodom su generirane konfiguracije za 40 različitih temperatura u rasponu $[0, 2]$ K. Neuronska mreža je analogna onoj za Isingov model s korekcijom za duplo veći broj ulaza (sinusi i kosinusi kuteva). Na slici ?? vidimo 9 odabranih konfiguracija na 9 odabranih temperatura. Vidimo da su blizu $T=0$ spinovi kvazi uređeni. Kako temperatura raste tako vidimo sve češće pojavljivanje spinskih vrtloga koji se ne poništavaju nužno sa susjednim vrlozima obratnog "predznaka". Fazni prijelaz nije lako uočiti okom. Na svakoj temperaturi generirano je 2000 konfiguracija koje su korištene kao ulazni podaci neuronske mreže kao i u slučaju Isingovog modela. Procedura označavanja podataka je ista kao i u Isingovom modelu.

Na slici 7 vidimo rezultat treninga 40 neuronskih mreža na 40 različitih odabira označavanja podataka metodom konfuzije. Na slici su i parcijalni fitovi radi vizualizacije W oblika krivulje. Fazni prijelaz očekujemo na 0.893 K no najbolji signal smo dobili za vrh na $T=1.43$. Lokaciranje faznog prijelaza iznad 1K je poznat problem u literaturi⁵ te je skoro nemoguće precizno locirati fazni prijelaz u rešetki konačnih dimenzija radi jako uskog područja kritikalnosti. Vidimo i zamjetne oscilacije u preciznosti u području u blizini faznog prijelaza. Objašnjenje bi se moglo tražiti u tome što su konfiguracije dosta slične za uzastopne temperature a model izgleda ne može pronaći dovoljno dobre značajke kojima bi diskriminirao takve konfiguracije, tj. konfiguracije za npr. 0.5 i 0.525 su modelu skoro iste. Tome se može probati doskočiti daljnim rafiniranjem modela. Mijenjanjem širine skrivenog sloja, i broja slojeva. U ovom radu takve promjene nisu previše pomogle, no vjerojatno postoji neki bolji odabir konfiguracije mreže nego je korišten u ovom radu.



Slika 6: Po jedna izabrana Monte Carlo konfiguracija XY modela na 9 odabranih temperatura u rasponu od 0 do 2 K.



Slika 7: Preciznosti 40 različitih neuronskih mreža za 40 različitih označavanja podataka

5 Zaključak

Metoda konfuzije se pokazala uspješnom u detekciji termalnog faznog prijelaza u Isingovom modelu. Isingov model je detaljno obrađivan u literaturi raznim metodama strojnog učenja tako da nije iznenađujuće da metoda konfuzije daje dobre rezultate. XY model se pokazao puno zahtjevniji. Radi specifičnosti i uskosti faznog prijelaza metoda konfuzije je dala grub rezultat a takvi su i u literaturi. To signalizira da je metoda konfuzije jako dobra za lociranje faznih prijelaza u slučajevima da su ti prijelazi dovoljno široki i klasičnog tipa. Topološki fazni prijelazi se čine zahtjevnijima. No, odabirom bolje konfiguracije neuronske mreže ili mijenjanjem tipa arhitekture neuronske mreže u cijelosti bi se moguće mogli postići bolji rezultati. No, konvolucijske mreže u literaturi ne daju nabolji rezultat, štoviše čak i lošiji⁵ od potpuno povezanih mreža pa stoga nisu korištene u ovom radu. Kako metoda konfuzije ne zahtjeva označeni skup podataka (u smislu da unaprijed znamo gdje je fazni prijelaz i označimo podatke shodno tome), ona se može koristiti u lokaciji nepoznatih faznih prijelaza. Jedan način za istražiti topološke fazne prijelaze bi bio umjesto klasičnih konfiguracija dobivenih pomoću Monte Carlo simulacije koristiti neke bolje ulazne podatke za neuronsku mrežu. U originalnom članku autora koji su osmislili metodu konfuzije⁴ kao ulazni podaci modela (ali ne XY modela) koji sadrži topološki fazni prijelaz (Kitaev lanac) se koristi spektar svojstvenih vrijednosti reduciranih matrica gustoće sustava. Ostaje za vidjeti je li takav pristup moguć u slučaju XY modela. U ovom radu su odabrani 2d Isingov model te 2d XY model no postoje mnoge varijacije na slične probleme za koje bi metoda konfuzije dala potencijalno dobre rezultate kao što je npr. 2d Isingov model na trokutastoj rešetci.

Literatura

- ¹ Anders W. Sandvik. *Computational Studies of Quantum Spin Systems*. AIP Conference Proceedings 1297, 135 (2010)
- ² Sacha Friedli, Yvan Velenik. *Statistical Mechanics of Lattice Systems*, (2017)
- ³ Barry M. McCoy and Tai Tsun Wu, *The Two- Dimensional Ising Model*. Harvard University Press, Cambridge Massachusetts, (1973)
- ⁴ Evert P. L. van Nieuwenburg, Ye-Hua Liu and Sebastian D. Huber. *Learning phase transitions by confusion*. Nature physics, (2017).
- ⁵ Song Sub Lee and Beom Jun Kim. *Confusion scheme in machine learning detects double phase transitions and quasi-long-range order*. Physical Review E, (2019)
- ⁶ P. Olsson and P. Minnhagen, *On the helicity modulus, the critical temperature and Monte Carlo simulations for the two-dimensional XY-model*. Phys. Scr. 43, 203 (1992).

- ⁷ Nielsen, Michael A. *Neural networks and deep learning*. Determination Press, (2015)
- ⁸ Pankaj Mehta, Marin Bukov, Ching-Hao Wang, Alexandre G.R. Day, Clint Richardson, Charles K. Fisher, David J. Schwab. *A high-bias, low-variance introduction to Machine Learning for physicists*. Physics Reports,(2019)
- ⁹ Aurélien Géron.*Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media, (2019)