

Rekonstrukcija hamiltonijana i određivanje faznih prijelaza Isingovog modela strojnim učenjem

Marko Bandalo*

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

Strojno učenje jedno je od najuzbudljivijih i najdinamičnijih područja modernog istraživanja i primjena. Cilj ovog rada je predstaviti neke modele iz područja strojnog učenja na Isingov model: Matematički model feromagnetizma u statističkoj mehanici. Proučavamo jednostavne modele poput logističke regresije te se prebacujemo na konvolucijsku neuronsku mrežu koja je klasa neuronskih mreža. Za kraj gledamo generativni model Duboki Boltzmanov stroj koji nam je bitan u slučaju kada nam nedostaje informacija o kritičnoj temperaturi T_C Isingova modela.

I. UVOD

Strojno učenje i umjetna inteligencija postali su jedna od najatraktivnijih i najbrže rastućih tehnologija današnjice. Virtualni asistenti, autonomni automobili, pametni zvučnici, chatbotovi, sustavi preporuka i filteri neželjenih e-mailova sve su rašireniji u svakodnevnom životu. Dostupnost velikih količina podataka i računalne snage omogućili su svojevrsnu tehnološku revoluciju pomoću algoritama neuronskih mreža koje sve bolje razumiju svijet koji vidimo i jezik kojim pričamo.

U ovom istraživanju primjeniti ćemo neke poznate modele iz strojnog učenja na poznati problem iz statističke fizike: Isingov model. Baviti ćemo se faznim prijelazima 2D Isingova modela. Klasificirati ćemo stanja Isingova modela te pokušati pomoću strojnog učenja odrediti temperaturu sustava. Za kraj ćemo pogledati kako preko neuronskih mreža možemo dobiti nove konfiguracije te kako se one razlikuju od konfiguracija dobivenih Monte Carlo simulacijom.

II. ISINGOV MODEL

Model koji ćemo koristiti za našu studiju faznih prijelaza na konačnim temperaturama za magnetske sustave zovemo Isingov model. U najjednostavnijem slučaju energija je izražena kao:

$$E = -J \sum_{\langle kl \rangle} N s_k s_l - B \sum_k s_k \quad (1)$$

gdje je $s_k = \pm 1$, N je ukupni broj spinova, J je konstanta vezanja koja izražava jačinu interakcija između susjednih spinova i B je vanjsko magnetsko polje koje interagira sa magnetskim momentom spinova. Simbol $\langle kl \rangle$ ukazuje da sumiramo samo po najbližim susjedima. Treba primjetiti da $J > 0$ upućuje da je energijski povoljnije da se susjedni spinovi usmjere. Ovo dovodi do fenomena, na dovoljno niskim temperaturama kojeg nazivamo spontana magnetizacija, tj. kroz interakcije sa najbližim su-

sjedima, dani magnetski moment može utjecati na usmjerenje spinova koji su udaljeni od danog spina za makroskopsku udaljenost. Ove dugodosežne korelacije između spinova su povezane sa dugodosežnim uređenjem u kojem rešetka ima ukupnu magnetizaciju u odsutnosti magnetskog polja. U daljnoj studiji Isingovog modela ograničavamo se na slučaj $B = 0$. Da bi izračunali izraze poput usrednjene energije $\langle E \rangle$ ili magnetizacije $\langle M \rangle$ u statističkoj fizici na danoj temperaturi, trebamo probabilističku distribuciju¹

$$P_i(\beta) = \frac{e^{-\beta E_i}}{Z} \quad (2)$$

, gdje je $\beta = \frac{1}{kT}$ inverzna temperatura, k je boltzmannova konstanta, E_i je energija stanja i Z je particijska funkcija za kanonski ansambl definirana kao $Z = \sum_{i=1}^M e^{-\beta E_i}$ gdje suma ide duž svih mikrostanja M . P_i izražava vjerojatnost nalaska sustava u danoj konfiguraciji i . Energija za specifičnu konfiguraciju je dana sa: $E_i = -J \sum_{\langle kl \rangle} N s_k s_l$. Magnetizacija je definirana kao:

$$M_i = \sum_{j=1}^N s_j \quad (3)$$

, gdje sumiramo po svim spinovima za danu konfiguraciju i .

II.1. Isingov Model i Fazni prijelazi

Isingov model u dvije dimenzije sa $B = 0$ prolazi kroz fazni prijelaz drugog reda. Pod ovim se podrazumijeva da ispod dane kritične temperature T_C , Isingov model pokazuje spontanu magnetizaciju sa $\langle M \rangle \neq 0$. Iznad T_C prosječna magnetizacija je 0. Usrednjena magnetizacija prilazi nuli na T_C sa beskonačnim nagibom. Takvo ponašanje je primjer onoga što se naziva kritičnim fenomenom. Kritični fenomen je normalno definiran sa jednom ili više termodinamičkih varijabli koje nestaju iznad kritične točke. U našem slučaju to je usrednjena magnetizacija $\langle M \rangle \neq 0$. Takav parametar se uobičajeno naziva parametrom uređenja. Za kvadratnu rešetku kada uzimamo slučaj $J = 1$ i $k_b = 1$ kritična temperatura je $T_C \approx 2.27$ dok je za trokutastu rešetku kritična temperatura $T_C \approx 3.65$.²

* mbandalo.phy@pmf.hr

III. METROPOLIS ALGORITAM I 2D ISINGOV MODEL

Algoritam za rješavanje Isingovog modela je predložio Nicholas Metropolis 1953. Nove konfiguracije su generirane od prethodnih koristeći vjerojatnost tranzicije koja ovisi o razlici energije između početnih i konačnih stanja. U našem slučaju imamo kao Monte Carlo funkciju uzorkovanja (eng. Sampling function) vjerojatnost nalaška sustava u stanju s zadanu sa:

$$P_s = \frac{e^{-\beta E_s}}{Z} \quad (4)$$

sa energijom E_s , $\beta = \frac{1}{kT}$ i Z je normalizacijska konstanta koja definira particijsku funkciju u kanonskom ansamblu. $Z(\beta) = \sum_s e^{-\beta E_s}$ je teško izračunati jer su nam potrebna sva stanja. U izračunu Isingovog modela u dvije dimenzije, broj konfiguracija je dan sa 2^N gdje je $N = L \times L$ broj spinova za rešetku duljine L . Na svu sreću, Metropolis algoritam uzima u obzir samo omjere između vjerojatnosti i ne trebamo uopće računati particijsku funkciju. Algoritam slijedi:³

1. Uspostavimo početno stanje sa energijom E_b tako da se pozicioniramo na nasumičnu konfiguraciju u rešetci.
2. Mijenjamo početnu konfiguraciju tako da okrenemo samo jedan spin. Izračunamo energiju ovog probnog stanja E_t .
3. Izračunamo $\Delta E = E_t - E_b$. Broj vrijednosti ΔE je limitiran na 5 za Isingov model u dvije dimenzije.
4. Ako je $\Delta E \leq 0$ prihvaćamo novu konfiguraciju, što znači da je energija spuštena i nadamo se da se krećemo prema energijskom minimumu za danu temperaturu. Prelazimo na korak 7.
5. Ako je $\Delta E \geq 0$, izračunamo $\omega = e^{-\beta \Delta E}$
6. Uspoređujemo ω sa nasumičnim brojem r . Ako je $r \leq \omega$, tada prihvaćamo novu konfiguraciju, inače zadržavamo staru konfiguraciju.
7. Sljedeći korak je anžurirati različite očekivane vrijednosti.
8. Koraci (2)-(7) se tada ponavljaju da bi dobili dovoljno dobre reprezentacije stanja.
9. Svaki put kada prođemo kroz rešetku, tj. kada smo sumirali po svim spinovima to nam konstituira ono što zovemo Monte Carlo ciklus. Možemo gledati na jedan ovaj ciklus kao mjerenje. Na kraju, možemo podijeliti različite očekivane vrijednosti sa ukupnim brojem ciklusa. Možemo odabrati želimo li podijeliti sa ukupnim brojem spinova ili ne. Ako također podijelimo sa ukupnim brojem spinova, energija je sada energija po spinu.

Ključni korak je izračun energije i promjene u magnetizaciji. Ovaj dio mora biti postavit na što efikasniji mogući način jer se promjena u energiji izračunava mnogo puta. U izračunu razlike u energiji od jedne konfiguracije do druge, ograničavamo promjenu na okretanje samo jednog spina. Za Isingov model u dvije dimenzije to znači da će biti samo ograničen skup vrijednosti za ΔE .

IV. LOGISTIČKA REGRESIJA

Logistička regresija je algoritam strojnog učenja koji se koristi za klasifikacijske probleme. To je algoritam prediktivne analize i baziran je na konceptu vjerojatnosti. Hipoteza logističke regresije teži ograničavanju funkcije gubitka između 0 i 1:

$$0 \leq h_\theta(x) \leq 1 \quad (5)$$

Da bi preslikali predviđene vrijednosti u vjerojatnosti, koristimo sigmoidnu funkciju. Funkcija preslikava bilo koju realnu vrijednost u drugu vrijednost između 0 i 1. Kada koristimo linearnu regresiju koristimo formulu za hipotezu:

$$h\Theta(x) = \beta_0 + \beta_1 X \quad (6)$$

gdje je β_0 odsječak na y-osi te β_1 nagib, slično kao u jednadžbi pravca. Za logističku regresiju malo ćemo to modificirati da dobijemo:

$$\sigma(Z) = \sigma(\beta_0 + \beta_1 X) \quad (7)$$

Očekujemo da naša hipoteza daje vrijednosti između 0 i 1.

$$\begin{aligned} Z &= \beta_0 + \beta_1 X \\ h\Theta(X) &= \text{sigmoid}(Z) \\ h\Theta(X) &= \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}} \end{aligned} \quad (8)$$

Sljedeće kreirajmo funkciju gubitka koju moramo minimizirati da bi dobili što precizniji model sa minimalnom greškom. Za logističku regresiju, funkcija gubitka definirana je ako:

$$Gubitak(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & \text{ako } y = 1 \\ -\log(1 - h_\theta(x)) & \text{ako } y = 0 \end{cases} \quad (9)$$

Te dvije jednadžbe možemo sažeti u jednu:⁴

$$J(\theta) = \frac{1}{m} \sum [y^{(i)} \log(h_\theta(x(i))) + (1 - y^{(i)}) \log(1 - h_\theta(x(i)))] \quad (10)$$

Logistička funkcija gubitka je konveksna. Da bi odredili θ moramo riješiti sljedeći optimizacijski problem:

$$\min_{\theta} J(\theta) \quad (11)$$

Postoji niz metoda koje se mogu koristiti za rješavanje ovog optimizacijskog problema poput metode prvog reda

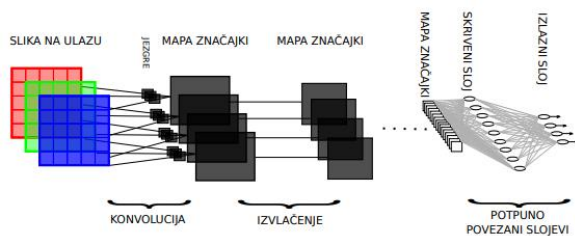
gradijentnog spuštanja za koju je potreban gradijent logističke funkcije gubitka ili metoda drugog reda poput Newtonove metode za koju je potreban gradijent i Hessian logističke funkcije gubitka. U slučaju gradijentnog spuštanja smjer u kojem tražimo minimum je negativna parcijalna derivacija logističke funkcije gubitka s obzirom na parametar θ :

$$\frac{\partial J(\theta)}{\partial \theta} = \frac{1}{m}(h_{\theta}(x) - y)x \quad (12)$$

U svojoj osnovnoj formi, gradijentno spuštanje će se nastaviti duž negativnog gradijentnog smjera od θ dok ne dosegne konvergenciju.

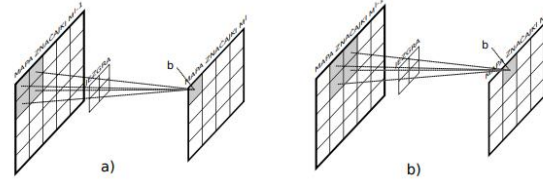
V. KONVOLUCIJSKE NEURONSKE MREŽE

Konvolucijska neuronska mreža se sastoji od nekoliko građevnih elemenata koje nazivamo konvolucijski slojevi. Na ulazu CNN-a nalazi se monokromatska slika ili slika u boji, nakon čega se izmjenjuju konvolucijski slojevi i slojevi sažimanja. Zatim slijedi nekoliko potpuno povezanih jednodimenzionalnih slojeva te izlazni sloj (Slika 1.).



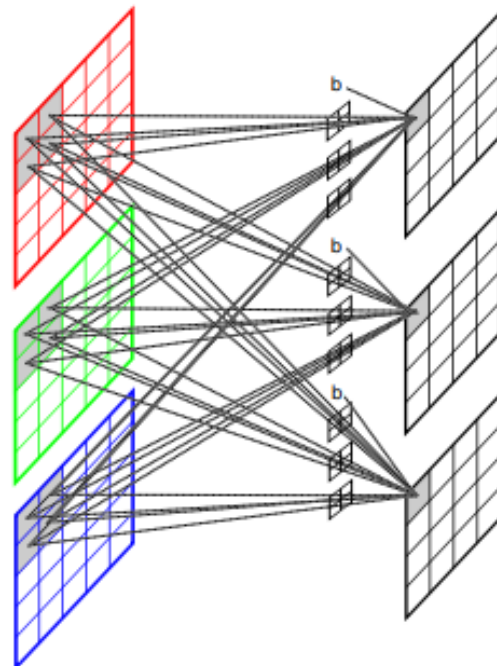
Slika 1: Opća struktura konvolucijskih neuronskih mreža⁵

Konvolucijski slojevi i slojevi sažimanja imaju dvodimenzionalne neurone koje nazivamo mapama značajki. Mape značajki nastaju tako da konvolucijski slojevi uzimaju mapu na ulazu sloja te izvede dvodimenzionalnu konvoluciju s jezgrom. Konvolucija nastaje prolazom kroz ulaznu mapu prozorom istim kao i jezgra, umnoškom ulaza i pripadajućih vrijednosti jezgre, zbrajanjem dobivenih vrijednosti s pragom, izračunom aktivacijske funkcije i u konačnici pohranom u pripadnu lokaciju izlazne mape. Prozor se potom pomiče pa se postupak ponavlja za sljedeći neuron u izlaznoj mapi (Slika 2.). Svaka veza sloja s prethodnim predstavljena je jezgrom. Prema tome, svaka mapa mora imati onoliko jezgri koliko je mapa u prethodnom sloju na koji je spojena.

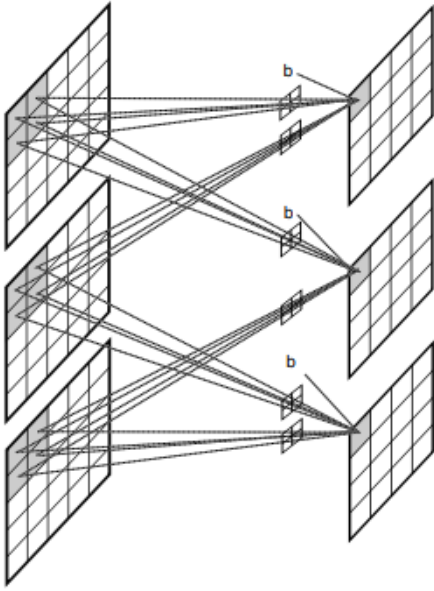


Slika 2: Konvolucija: a) prvi korak; b) korak nakon pomaka receptivnog polja⁵

Većina autora koristi potpuno povezane konvolucijske slojeve u kojima je svaka mapa značajki pojedinog sloja povezana sa svim mapama značajki prethodnog sloja (Slika 3.), no koristi se i raspršena povezanost u kojoj nisu sve mape značajki sloja povezane sa svim ostalim mapama značajki prethodnog sloja (Slika 4.).⁵

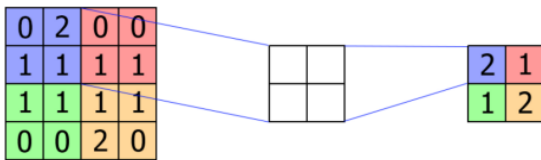


Slika 3: Prvi korak konvolucija (za svaku mapu značajki u sloju) kod potpuno povezanih slojeva⁵



Slika 4: Primjer raspršene povezanosti u kojoj je svaka mapa povezana samo sa dvije mape iz prethodnog sloja⁵

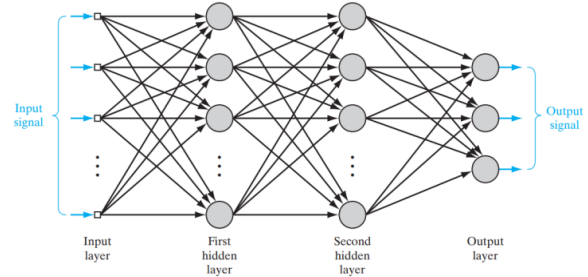
Slojevi sažimanja (engl. pooling) služe za smanjenje dimenzija mapi značajki, te za uklanjanje varijance. U slojevima sažimanja imamo i okvire s kojima prolazimo preko mape značajki. Mapa značajki sažima se na način da se okvir predstavi sa samo jednom vrijednošću. Tako npr., na slici 5. možemo vidjeti kako se okvir veličine 2x2 reprezentira s jednom vrijednošću dobivenom iz 4 vrijednosti unutar okvira čime se mapa značajki smanjuje 4 puta. Pomicanje okvira u navedenom primjeru bio bi jednak 2 u horizontalnom, te 2 u vertikalnom smjeru. Sažimanje je moguće odraditi na dva načina: sažimanje usrednjavanjem, te sažimanje maksimalnom vrijednošću. Sažimanje usrednjavanjem (engl. mean pooling) uzima aritmetičku sredinu vrijednosti koje se nalaze unutar okvira sažimanja. Sažimanje maksimalnom vrijednošću (engl. max pooling) uzima maksimalnu vrijednost unutar okvira sažimanja. Na slici 5. dan je primjer sažimanja maksimalnom vrijednošću.



Slika 5: Primjer sloja sažimanja⁶

Posljednji sloj konvolucijske neuronske mreže koristi se isključivo tijekom procesa učenja kada koristi funkciju gubitka radi procjene kvalitete predviđanja same mreže. Posljedni sloj se naziva „Multilayer Perceptron(MLP)” i sastoji se od tri tipa slojeva: ulazni, skriveni i izlazni

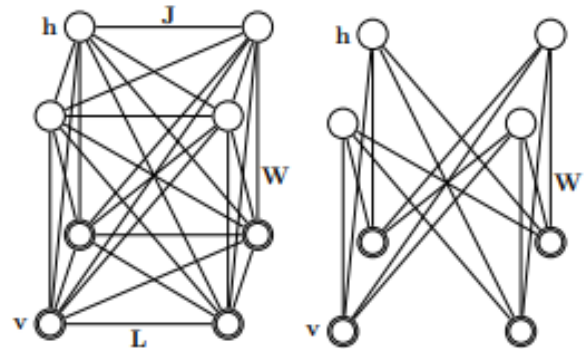
sloj. U ulazni sloj ulaze značajke generirane od strane CNN-a. Skriveni sloj je niz neurona za težinama koje će biti naučene tijekom postupka učenja. MLP se sastoji od jednog ili više skrivenih slojeva. Izlazni sloj je također niz neurona, ali ima različitu aktivacijsku funkciju. Najčešće se koristi softmax funkcija da bi se generirale vjerojatnosti za svaku kategoriju.



Slika 6: „Multilayer Perceptron(MLP)” sa dva skrivena sloja⁷

VI. DUBOKI BOLTZMANOVI STROJEVI

Boltzmanov stroj je mreža simetričkih vezanih stohastičkih binarnih jedinica. Sadrži skup vidljivih jedinica $\mathbf{v} \in \{0, 1\}^D$ i skup skrivenih jedinica $\mathbf{h} \in \{0, 1\}^P$ (Slika 7.).



Slika 7: Lijevo: Općeniti Boltzmanov stroj. Gornji sloj predstavlja vektor stohastičkih binarnih ”skrivenih” značajki dok donji sloj predstavlja vektor stohastičkih binarnih ”vidljivih” varijabli. Desno: Ograničeni Boltzmanov stroj bez skriveno-skriveno povezanosti i bez vidljivo-vidljivo povezanosti⁸

Energija stanja $\{\mathbf{v}, \mathbf{h}\}$ je definirana kao:⁸

$$E(\mathbf{v}, \mathbf{h}; \theta) = -\frac{1}{2}\mathbf{v}^T \mathbf{L} \mathbf{v} - \frac{1}{2}\mathbf{h}^T \mathbf{J} \mathbf{h} - \mathbf{v}^T \mathbf{W} \mathbf{h}, \quad (13)$$

gdje su $\theta = \{\mathbf{W}, \mathbf{L}, \mathbf{J}\}$ su parametri modela: $\mathbf{W}, \mathbf{L}, \mathbf{J}$ predstavljaju vidljivo-skriveno, vidljivo-vidljivo i

skriveno-skriveno simetričke interakcijske članove. Dijaagonalni elementi od \mathbf{L} i \mathbf{J} su jednaki 0. Vjerojatnost da model poprimi stanje zadano s \mathbf{v} je:⁸

$$p(\mathbf{v}; \theta) = \frac{p^*(\mathbf{v}; \theta)}{Z(\theta)} = \frac{1}{Z(\theta)} \sum_{\mathbf{h}} \exp(-E(\mathbf{v}, \mathbf{h}; \theta)), \quad (14)$$

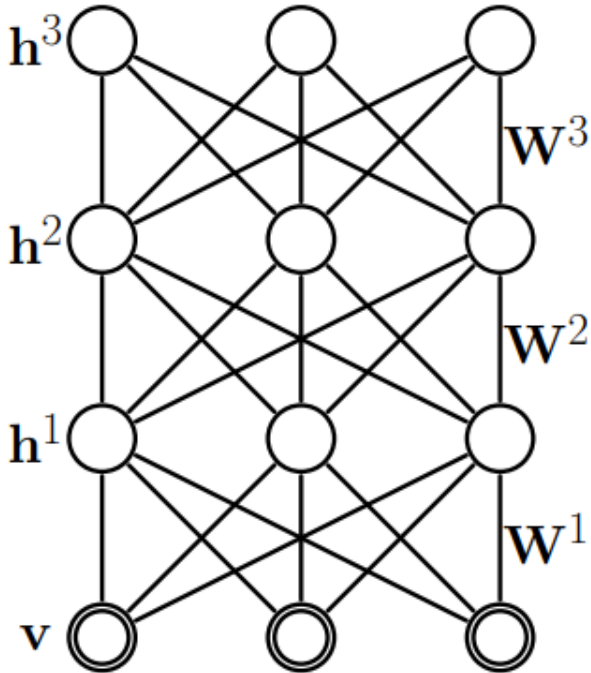
$$Z(\theta) = \sum_{\mathbf{v}} \sum_{\mathbf{h}} \exp(-E(\mathbf{v}, \mathbf{h}; \theta)).$$

gdje p^* označava nenormaliziranu vjerojatnost, a $Z(\theta)$ je particijska funkcija. Rijetko kad bi bili zainteresirani za učenje potpuno povezanog Boltzmanovog stroja. Umjesto toga promotrimo duboki Boltzmanov stroj sa tri skrivena sloja prikazan na slici 8. u kojem svaki sloj bilježi komplicirane korelacije višeg reda između aktivnosti skrivenih značajki u sloju ispod. Energija stanja $\{\mathbf{v}, \mathbf{h}\}$ je definirana kao:⁸

$$E(\mathbf{v}, \mathbf{h}; \theta) = -\mathbf{v}^T \mathbf{W}^1 \mathbf{h}^1 - \mathbf{h}^{1T} \mathbf{W}^2 \mathbf{h}^2 - \mathbf{h}^{2T} \mathbf{W}^3 \mathbf{h}^3, \quad (15)$$

gdje je $\mathbf{h} = \{\mathbf{h}^1, \mathbf{h}^2, \mathbf{h}^3\}$ skup skrivenih jedinica, dok je $\theta = \{\mathbf{W}^1, \mathbf{W}^2, \mathbf{W}^3\}$ su parametri modela koji predstavljaju vidljivo-skriveno i skriveno-skriveno simetričke interakcijske članove. Vjerojatnost da model poprimi stanje zadano s vidljivim vektorom \mathbf{v} je:

$$p(\mathbf{v}; \theta) = \frac{p^*(\mathbf{v}; \theta)}{Z(\theta)} = \frac{1}{Z(\theta)} \sum_{\mathbf{h}} \exp(-E(\mathbf{v}, \mathbf{h}^1, \mathbf{h}^2, \mathbf{h}^3; \theta)) \quad (16)$$



Slika 8: Troslojni Duboki Boltzmanov stroj⁹

Derivacija logaritma vjerojatnosti u odnosu na parametar W^1 jednaka je:⁸

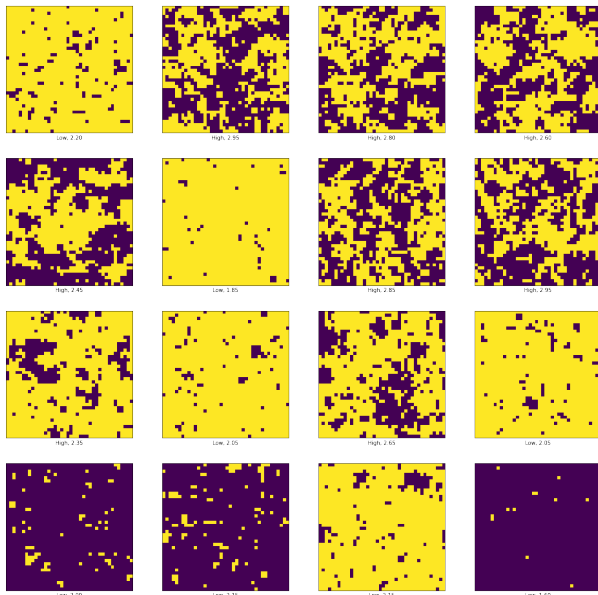
$$\frac{\partial \log P(\mathbf{v}; \theta)}{\partial W^1} = E_{P_{data}}[\mathbf{v} \mathbf{h}^{1T}] - E_{P_{model}}[\mathbf{v} \mathbf{h}^{1T}], \quad (17)$$

gdje $E_{P_{data}}[\cdot]$ predstavlja očekivanost s obzirom na potpunu distribuciju podataka $P_{data}(\mathbf{h}, \mathbf{v}; \theta) = P(\mathbf{h}|\mathbf{v}; \theta)P_{data}(\mathbf{v})$, gdje $P_{data}(\mathbf{v}) = \frac{1}{N} \sum_n \delta(\mathbf{v} - \mathbf{v}_n)$ predstavlja empirijsku distribuciju i $E_{P_{model}}[\cdot]$ je očekivanje s obzirom na distribuciju definiranu modelom.

VII. REZULTATI I ANALIZA

VII.1. Podaci

Podatke za 2D Isingov model generiramo pomoću Monte-Carlo simulacije objašnjene u poglavlju III. Radi jednostavnosti postavljamo $J = 1$. Kod strojnog učenja potrebna su nam 3 skupa podataka: skup za učenje, skup za validaciju i skup za testiranje. Skup za učenje je glavni dio strojnog učenja i koristi se da bi namjestili parametre nekog modela. Skup za validaciju se koristi kako bi namjestili hiperparametre modela. Primjer hiperparametra za neuronsku mrežu je broj skrivenih jedinica u svakom sloju. Skup za testiranje nam pokazuje učinkovitost našeg modela. Za skup za učenje i za skup za validaciju generiramo 2D kvadratnu rešetku dimenzija 20×20 i 40×40 s nasumičnim vrijednostima $\sigma_{x,y} = \pm 1$. Za skup za testiranje generiramo trokutastu rešetku dimenzija 20×20 i 40×40 . Za skup za testiranje nismo odabrali kvadratnu nego trokutastu rešetku kako bi potvrdili da nam model ne upada u problem prenaučnosti (eng. overfitting). Problem prenaučnosti (eng. overfitting) jedan je od najvećih problema kod učenja modela. Problem se sastoji u tome što model počinje previše odgovarati konkretnim primjerima na kojima je treniran čime više ne odgovara generalnim uzorcima već samo tim specifičnima iz skupa za učenje. Ako su izračunate vrijednosti na skupu za učenje i na skupu za testiranje slični tada naš model nema problema prenapućnosti. Sustav generiramo na temperaturama od 1.5 K do 3.2 K s ikrementom od 0.05 K za kvadratnu rešetku, te na temperaturama od 2.5 K do 4.5 K s ikrementom od 0.1 K za trokutastu rešetku. Za svaku temperaturu simulaciju ponavljamo 1000 puta.



Slika 9: 16 nasumično izvučenih konfiguracija iz skupa za učenje s dimenzijama kvadratne rešetke 40×40 . Za svaku konfiguraciju prikazana je odgovarajuća temperatura

Kao što vidimo na slici 9. kod konfiguracije za nisku temperaturu, tj. temperaturu ispod kritične T_C većina spinova je usmjerena ili prema gore (žuta boja) ili prema dolje (ljubičasta boja). Ovo je očekivano jer su usmjereni spinovi u nižem energetsom stanju.

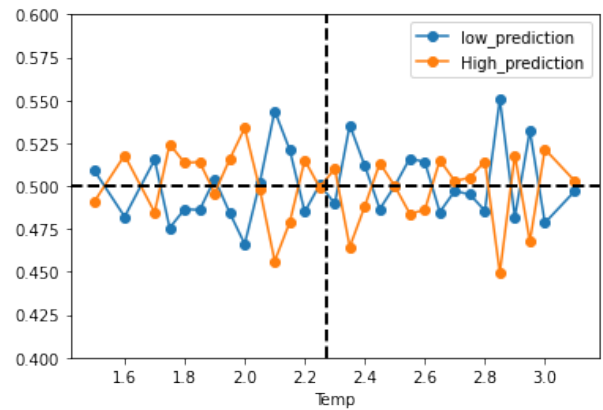
VII.2. Klasifikacija faznih prijelaza

Za klasifikaciju faznih prijelaza krećemo sa binarnom logističkom regresijom koju smo objasnili u poglavlju IV. Postoje samo dvije moguće klase: Niska temperatura i Visoka temperatura. Prvo krećemo sa sustavom 20×20 . Učimo model na kvadratnoj rešetci i prolazimo kroz sustav 15 puta tj. imamo 15 epoha. Nakon toga prelazimo na validacijski skup podataka na kojem možemo namještati hiperparametre te naposljetku prelazimo na skup za testiranje na kojem gledamo preciznost našeg modela. Prag za logističku regresiju smo postavili na 0.5, odnosno 50% Rezultati za 20×20 i 40×40 prikazani su u tablicama I i II.

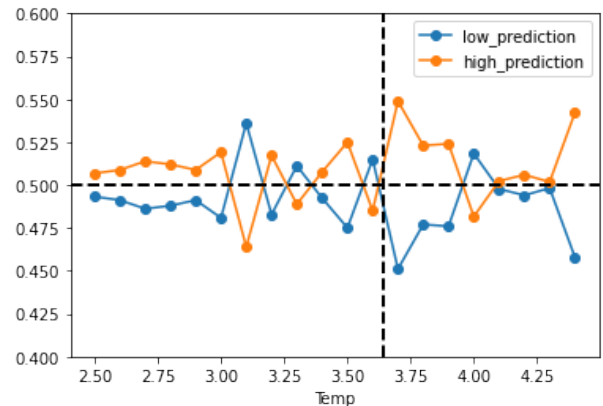
Iz rezultata vidimo da logistička regresija nije baš dobar klasifikator za Isingov model. Trebamo izabrati neki novi model koji bi nam poboljšao preciznost i smanjio gubitak.

Tablica I: Prikaz preciznosti i gubitka za sustav 20×20 za logističku regresiju.

	Set za učenje	Set za validaciju	Set za testiranje
Preciznost	57.8 %	50.2%	51.9 %
Gubitak	0.7180	0.8308	0.8166



(a)



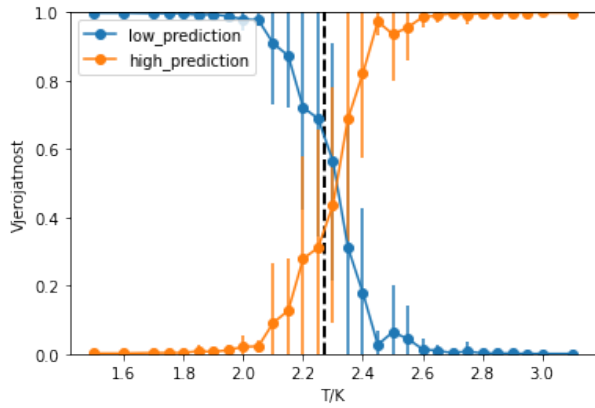
(b)

Slika 10: Prikaz vjerojatnosti za različite temperature za (a) kvadratnu 40×40 rešetku (b) trokutastu 40×40 rešetku

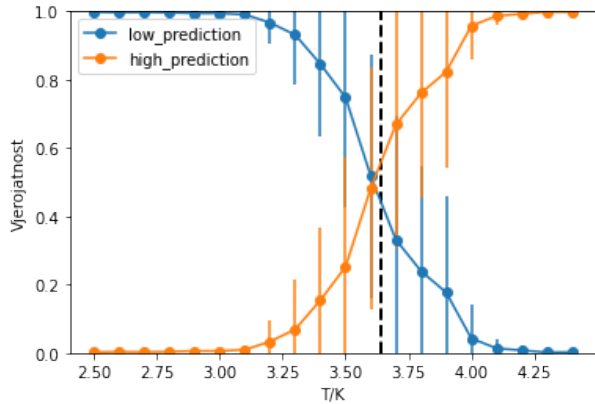
Tu ćemo izabrati konvolucijsku neuronsku mrežu koju smo kratko objasnili u poglavlju V. Konstruirat ćemo mrežu sa dva skrivena sloja. Prvi je 2-D konvolucijski sloj sa 64 filtera kojeg slijedi potpuno povezani sloj u kojem koristimo dva skrivena sloja i kao aktivacijsku funkciju koristimo softmax funkciju. Klasifikator treniramo tako da minimiziramo binarnu unakrsnu entropiju (eng. Binary cross-entropy loss). Da bi pripremili podatke za model potrebno je prilagoditi rubove kako bi se uzeli u obzir periodični rubni uvjeti u Monte Carlo simulaciji. Zbog toga slike koje nam ulaze u simulaciju će biti dimenzija 22×22 za sustav dimenzija 20×20 . Isto tako za sustav 40×40 će nam u simulaciju ulaziti konfigu-

Tablica II: Prikaz preciznosti i gubitka za sustav 40×40 za logističku regresiju.

	Set za učenje	Set za validaciju	Set za testiranje
Preciznost	68.1 %	54.0%	44.8 %
Gubitak	0.6348	0.7813	0.8570



(a)



(b)

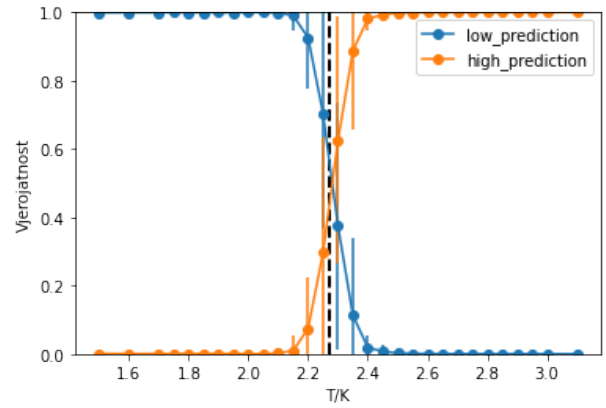
Slika 11: Prikaz vjerojatnosti za različite temperature za (a) kvadratnu 20×20 rešetku (b) trokutastu 20×20 rešetku

racije dimenzija 44×44 . Koristiti ćemo 2×2 filtere u konvolucijskom sloju. Kroz sustav prolazimo 3 puta, tj. imamo 3 epohe. Za hiperparametar od 3 epohe imamo najmanju mogućnost "overfittinga". Rezultati za preciznost i gubitak su prikazani u Tablicama III i IV. dok su odgovarajući grafovi prikazani na slikama 11 i 12.

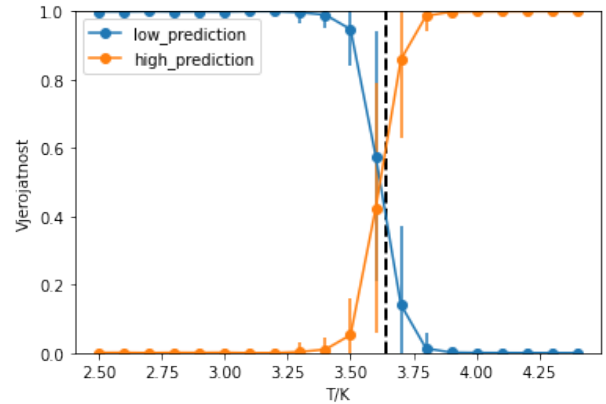
Usporedbom rezultata CNN-a i logističke regresije vidimo da nam je CNN puno bolji model za klasifikaciju faznih prijelaza Isingova modela. Druga stvar koju smo opazili da je preciznost veća, odnosno gubitak manji za sustav dimenzija 40×40 . To je posljedica činjenice da za se obrasci po kojem naš model uči, očitije pojavljuju za veće sustave. Temperature u kojem nam sustav ko-

Tablica III: Prikaz preciznosti i gubitka za sustav 20×20 za konvolucijsku neuronsku mrežu.

	Set za učenje	Set za validaciju	Set za testiranje
Preciznost	95.0 %	94.5%	92.0 %
Gubitak	0.1972	0.1430	0.1889



(a)



(b)

Slika 12: Prikaz vjerojatnosti za različite temperature za (a) kvadratnu 40×40 rešetku (b) trokutastu 40×40 rešetku

nvergira jednake su 2.30 K za kvadratnu 20×20 rešetku te 3.5 K za trokutastu 20×20 rešetku. Za sustav dimenzija 40×40 dobivamo temperature 2.25 K za kvadratnu rešetku te 3.6 K za trokutastu rešetku. Sa ovim konvergiranim temperaturama vidimo kako je izračunata kritična temperatura većeg sustava bliža teorijskoj kritičnoj temperaturi. Za kraj ćemo pogledati generativne modele. U prošlim poglavljima smo koristili znanje kritične temperature za kvadratnu rešetku ($T_C = \approx 2.26$) te trokutaste rešetke ($T_C = \approx 3.65$) da bi označili konfiguracije spina i promatrali problem klasificiranja stanja s obzirom na njihov fazni prijelaz. Problem nastaje u kompliciranijim modelima, gdje točna pozicija T_C nije poznata i tako

Tablica IV: Prikaz preciznosti i gubitka za sustav 40×40 za konvolucijsku neuronsku mrežu.

	Set za učenje	Set za validaciju	Set za testiranje
Preciznost	98.6 %	97.5%	97.4 %
Gubitak	0.0380	0.0624	0.0683

onda uopće ne možemo označiti stanje.

Generativni modeli se mogu koristiti za učenje varijacijske aproksimacije za distribuciju vjerojatnosti koje su generirale podatke. Koristeći samo 2D konfiguracije, sada učimo Duboki Boltzmanov stroj. Ovdje izvodimo analizu na fiksnoj temperaturi T . Zatim možemo primijeniti naš model na tri različite vrijednosti $T = 1.75$ K, 2.25 K i 2.75 u uređenom, gotovo kritičnom i neuređenom području. Definiramo Duboki Boltzmanov stroj sa dva skrivena sloja sa $N_{skriveno}$ i $N_{skriveno}/10$ jedinica i primijenjujemo L_1 regularizaciju na sve težine. Slike 13-15 pokazuju rezultate numeričkog eksperimenta na $T = 1.75$ K, 2.25 K i 2.75 K za DBM sa $N_{skriveno} = 800$. Gledajući rekonstruirane i imaginarnе čestice, vidimo da naš DBM radi dobro u neuređenim i kritičnim područjima. Ali arhitektura slojeva nije optimalna za $T = 1.75$ K u uređenom fazi što je posljedica efekta povezanih sa lomom simetrije.

VIII. ZAKLJUČAK

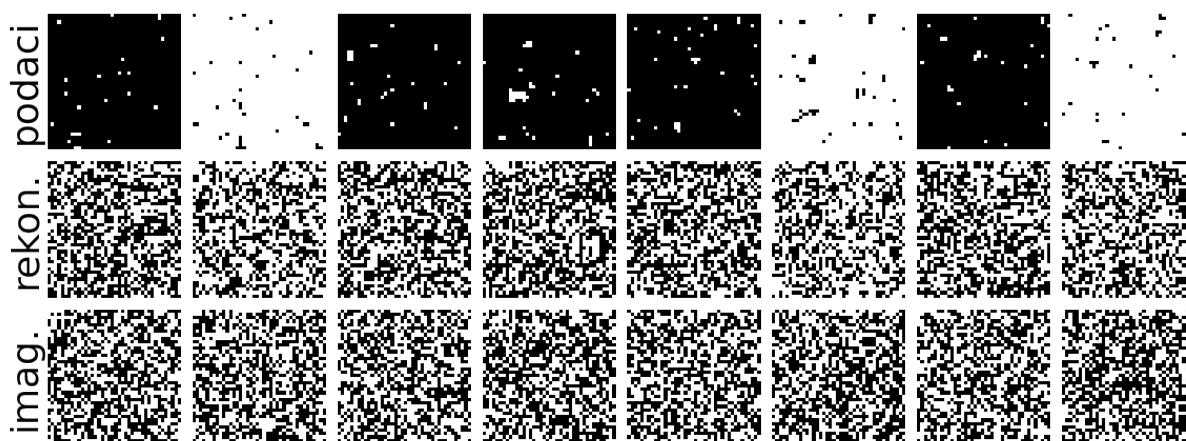
U ovom istraživanju promotrili smo jedan od zanimljivijih problema statističke fizike, tj. Isingov model. Nakon što smo kreirali podatke Monte Carlo simulacijom pomoću Metropolis algoritma klasificirali smo stanja Isingova modela s obzirom na fazna stanja. Proveli smo lo-

gističku regresiju i utvrdili da se preciznost modela kreće oko 50% što znači logistička regresija nije dobar model za klasifikaciju stanja Isingovog modela. Za model koji bi nam davao bolje rezultate izabrali smo Konvolucijsku neuronsku mrežu koja je dala značajno poboljšanje u odnosu na logističku regresiju. Preciznost se značajno poboljšala te prelazi 90% dok se gubitak značajno smanjio. Druga značajna stvar koju smo opazili u našim rezultatima je da veći sustavi, u našem slučaju sustav dimenzija 40×40 daju bolju preciznost i manji gubitak. Ovo je posljedica činjenice što više povećavamo dimenzije sustava više se očitavaju obrasci unutar sustava po kojem naši modeli uče. Zatim smo prešli na generativne modele koji su nam korisni u slučaju da nam točna pozicija kritične temperature nije poznata. Tu smo izvršili simulaciju Dubokog Boltzmanovog stroja na tri temperature u tri ključne regije 2D isingova modela. Model nam je dao prihvatljive rezultate u kritičnom području i neuređenom području dok je u uređenom području podbacio zbog efekta loma simetrije. Svi ovi modeli koje smo objasnili mogu nam dati kritičnu temperaturu sustava kojeg nemožemo teorijski odrediti. Naravno naše modele možemo poboljšati tako da smanjimo inkrement temperature te proizvedemo veći broj podataka u području oko kritične regije, tj. u području oko kritične temperature T_C .

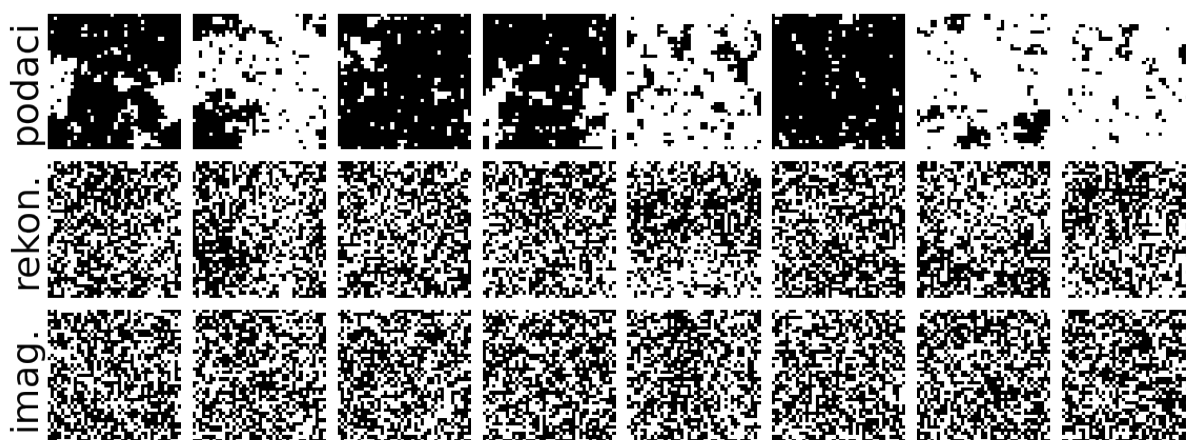
-
- ¹ Satya Pal Singh: The Ising Model: Brief Introduction and Its Application DOI: <http://dx.doi.org/10.5772/intechopen.90875>
- ² Carrasquilla J, Melko R G. Machine learning phases of matter[J]. Nature Physics, 2017.
- ³ Morten Hjorth-Jensen: Computational physics, University of Oslo, Fall 2010
- ⁴ Bishop, C. M. 2006. Pattern recognition and machine learning. Springer.
- ⁵ Božić-Stulić D. Semantička segmentacija slika metodom dubokog učenja (Diplomski rad). Preuzeto sa: [https://data.fesb.unist.hr/public/news/Dunja BožićStulić-a3b4859ad0.pdf](https://data.fesb.unist.hr/public/news/Dunja%20Božić-Stulić-a3b4859ad0.pdf)
- ⁶ Vukotić, V. (2014). Raspoznavanje objekata dubokim neuronskim mrežama (Diplomski rad). Preuzeto s ht-

[tps://urn.nsk.hr/urn:nbn:hr:168:749831](https://urn.nsk.hr/urn:nbn:hr:168:749831)

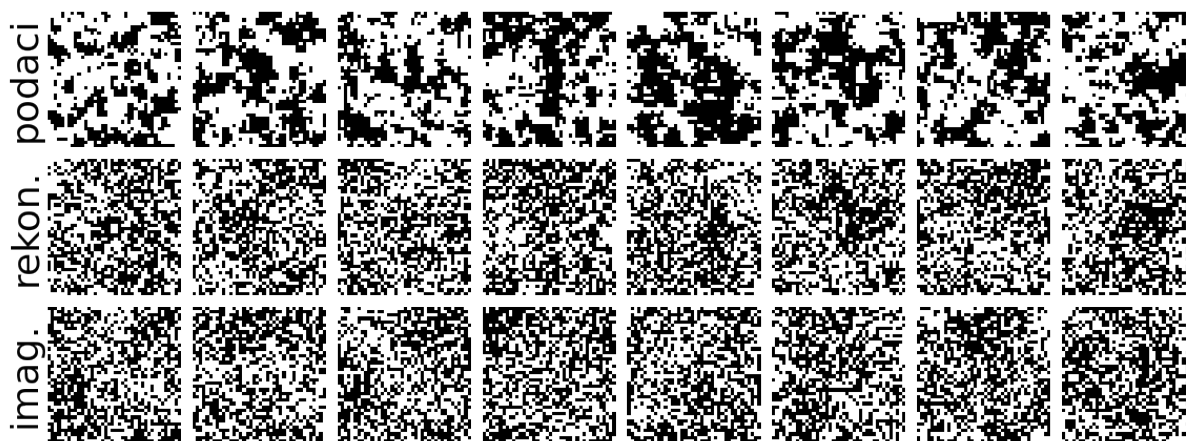
- ⁷ Edgar Amaya, Vicotr Celestino: Different control strategies used in didactic plant PD-3 of SMAR through OPC technology. 19th International Congress of Mechanical Engineering November 5-9, 2007, Brasília, DF
- ⁸ R. R. Salakhutdinov and G. E. Hinton. Deep Boltzmann machines. In Proceedings of the International Conference on Artificial Intelligence and Statistics, volume 12, 2009.
- ⁹ Ruslan Salakhutdinov and Geoff Hinton. A better way to pretrain deep boltzmann machines. In Advances in Neural Information Processing Systems 25, pages 2456–2464. 2012.



Slika 13: Monte Carlo podaci, njihova rekonstrukcija i imaginarne čestice generirane sa Dubokim Boltzmanovim strojem u uređenoj fazi na $T = 1.75$ K.



Slika 14: Monte Carlo podaci, njihova rekonstrukcija i imaginarne čestice generirane sa Dubokim Boltzmanovim strojem u kritičnoj fazi na $T = 2.25$ K.



Slika 15: Monte Carlo podaci, njihova rekonstrukcija i imaginarne čestice generirane sa Dubokim Boltzmanovim strojem u neuređenoj fazi na $T = 2.75$ K.