

**ANALIZA GEOMETRIJE MANIFOLDA U AUTOENKODERIMA
MANIFOLD GEOMETRY ANALYSIS IN AUTOENCODERS**Vuko Jovičić, *Fakultet tehničkih nauka, Novi Sad***Oblast – RAČUNARSTVO I AUTOMATIKA**

Kratak sadržaj – Ovaj rad posvećen je pokušaju da se razumeju i rasvetle neki od principa u radu dubokih neuronskih mreža. Konkretno, kako širina središnjeg sloja u autoenkoderu, intrinzična dimenzionalnost ulaznog skupa podataka i dužina obučavanja mreže utiču na geometriju formiranog manifolda.

Ključne reči: Autoenkoderi, mašinsko učenje, manifoldi, interpolacija

Abstract – This paper attempts to help in understanding and clarifying some of the principles in the work of deep neural networks. Specifically, how the width of the bottleneck layer in the autoencoder, the intrinsic dimensionality of the input data and the training duration affect the geometry of the formed manifold.

Keywords: Autoencoders, machine learning, manifolds, interpolation

1. UVOD

Neuronske mreže i duboko učenje u današnje vreme predstavljaju standardnu tehniku za rešavanje problema mašinskog učenja. Ipak, čak i uz svakodnevnu pojavu novih modela i arhitektura, kao i sve bolje performanse koje oni pružaju, neuronske mreže u mnogome i dalje ostaju crna kutija a mnogobrojni detalji u njihovom funkcionisanju ostaju nepoznanica.

Neka od pitanja na koji istraživači pokušavaju da daju odgovor tiču se geometrije interne reprezentacije podataka u skrivenim slojevima mreže. Bolje razumevanje određenih topoloških i geometrijskih struktura unutar neuronske mreže kao i načina na koji se formiraju dovelo bi do mogućnosti modelovanja efikasnijih arhitektura, interpretabilnijih sistema i predvidivijih izlaza.

Podešavanje parametara mreže je jedan od osnovnih koraka u obučavanju svakog modela a svi oni na ovaj ili onaj način utiču na formiranje interne reprezentacije podataka. Iako se ova oblast sve više istražuje i dolazi se do novih saznanja, u praksi se i dalje odabir arhitekture za rešavanje određenog problema kao i podešavanje parametara u najvećoj meri zasniva na iskustvu i empirijskom pristupu.

NAPOMENA:

Ovaj rad proistekao je iz master rada čiji mentor je bio dr Milan Segedinac, vanr. prof.

Različita istraživanja i eksperimenti vršeni su u cilju boljeg razumevanja uticaja interne reprezentacije podataka na funkcionisanje neuronskih mreža.

Rad [1] posvećen je upravo razumevanju geometrije dubokih neuronskih mreža. Potvrđuje se hipoteza da glavni uspeh neuronskih mreža leži u činjenici da prirodni podaci visoke dimenzionalnosti zapravo leže u niže-dimenzionalnom prostoru (hipoteza o manifoldu). Rad [2] se bavio topološkim i geometrijskim osobinama manifolda i modelovanju istih u cilju što boljeg predstavljanja određenog skupa podataka sa određenom intrinzičnom topološkom strukturom. U [3], cilj istraživanja bio je ispratiti intrinzičnu dimenzionalnost interne reprezentacije slike kroz različite slojeve duboke neuronske mreže.

Jedan od glavnih motiva za izučavanje geometrije interne reprezentacije objekata jeste i interpolacija unutar prostora reprezentacija. Rad [4] se zasniva na ideji da dublji slojevi neuronske mreže predstavljaju apstraktnije karakteristike podataka od ranih slojeva i da interpolacija između tačaka u tom prostoru ima više semantičkog značaja nego u početnim slojevima mreže.

Cilj ovog rada bio je da kroz eksperimente pokuša dati svoj doprinos u otkrivanju određenih zakonitosti između različitih parametara mreže i rezultujuće geometrije manifolda.

2. TEORIJSKE OSNOVE

Za razumevanje metodologije i rezultata izvršenih eksperimenata potrebno je poznavati određene teorijske pojmove:

2.1. Manifold

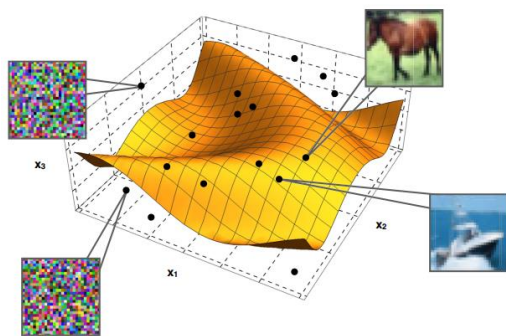
U matematičkoj teoriji, manifold predstavlja topološki prostor koji lokalno liči na euklidski. Preciznije, n -dimenzionalni manifold ili skraćeno n -manifold je topološki prostor sa osobinom da svaka tačka unutar njega ima okolinu koja je homeomorfna sa otvorenom jediničnom loptom u R^n .

2.2. Hipoteza o manifoldu

Hipoteza o manifoldu kaže da se mnogi visoko-dimenzionalni skupovi podataka koji se mogu naći u stvarnom svetu tj. prirodi zapravo koncentrišu u blizini niže-dimenzionalnih manifolda unutar tih visoko-dimenzionalnih prostora.

Opšte je prihvaćeno da hipoteza o manifoldu objašnjava efikasnost algoritama mašinskog učenja u opisivanju visoko-dimenzionalnih skupova podataka malim brojem

parametara tj. u rešavanju problema nelinearne redukcije dimenzionalnosti skupa podataka, kao što je u slučaju autoenkodera.



Slika 1. Pojednostavljeni prikaz hipoteze o manifoldu [5]

2.3. Autoenkodori

Autoenkodori predstavljaju posebnu vrstu veštačkih neuronskih mreža koji za cilj imaju da nauče reprezentativne karakteristike - kodove neoznačenih skupova podataka.

Kod standardnog autonekodera, središnji sloj u kome se čuva reprezentativni kod ulaznog skupa podataka manje je dimenzionalnosti od ulaznog sloja, pa je tako mreža prinuđena da odbaci neke od manje važnih karakteristika skupa (šum) a zadrži samo one najekspresivnije, stvarajući tako kompresovani oblik ulaznog skupa koji je u rešavanju mnogih problema korisniji od originalne reprezentacije. Prostor u kome se nalazi kompresovana reprezentacija podataka naziva se i latentni prostor.

2.4. Intrinzična dimenzionalnost

Intrinzična dimenzionalnost - ID predstavlja minimalni broj dimenzija neophodnih da se opišu elementi skupa bez značajnog gubitka informacija.

Problem pronalazjenja ID određenog skupa je vrlo zahtevan i postoje mnogi algoritmi i metodi u njegovom rešavanju. Neki od estimatora ID su: DANCo [6], KNN, PCA, MADA [7], TwoNN [8] itd.

Za većinu realnih skupova podataka iz prirode, moguće je napraviti razliku između pojmova globalna intrinzična dimenzionalnost GID i lokalna intrinzična dimenzionalnost LID. Ovi koncepti usko su povezani sa konceptom manifolda. LID predstavlja dimenzionalnost manifolda u okolini tačke koja se posmatra, dok GID sagledava čitav skup tačaka i određuje dimenzionalnost konveksnog omotača manifolda.

2.5. Prokletstvo dimenzionalnosti

Ovaj pojam se odnosi na različite fenomene do kojih dolazi pri radu sa podacima u visoko-dimenzionalnim prostorima, za razliku od nisko-dimenzionalnih sa kojima se svakodnevno susrećemo.

Osnovna pojava do koje dolazi je da zapremina prostora raste eksponencijalno sa uvećavanjem broja dimenzija, pa tako i broj tačaka koji je potreban da bi se zadržala određena gustina mora rasti eksponencijalno.

Takođe, kako dimenzionalnost prostora raste, tačke tog prostora međusobno postaju sve udaljenije i razuđenije,

sve do stepena da praktično postaju jednako udaljene jedne od drugih. Ova osobina, između ostalog, onemogućava rad mnogim algoritmima koji se oslanjaju na prepoznavanje određenih grupa tačaka ili celina koje dele određenu zajedničku osobinu.

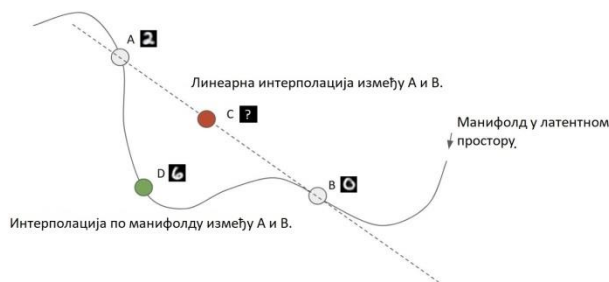
Eksperimentalni deo ovog rada baviće se i pokušajem da se donesu zaključci o promeni prirode skupa podataka pri drastičnom povećanju dimenzionalnosti.

2.6. Interpolacija i ekstrapolacija

Pojmovi interpolacije i ekstrapolacije su fundamentalni kako u mašinskom učenju tako i u drugim oblastima oblastima tehničkih nauka i matematike.

Do interpolacije dolazi, posmatrajući tačku x , kada god ona pripada konveksnom omotaču skupa tačaka $X = \{x_1, x_2, \dots, x_n\}$, u suprotnom dolazi do ekstrapolacije.

U kontekstu mašinskog učenja i dubokih neuronskih mreža, interpolacija se može povezati sa pojmom manifolda. Kao što se može videti na slici 2, linearna interpolacija kroz latentni prostor reprezentacija rezultovace nepredvidivim slikama, dok će interpolacija po manifoldu rezultirati poznatima primerima iz skupa podataka ili njihovim kombinacijama na čemu se u konačnici zasnivaju generativni modeli.

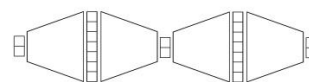


Slika 2. Prikaz linearne interpolacije i interpolacije po manifoldu u latentnom prostoru mreže [9]

3. EKSPERIMENTI

3.1. Eksperimenti sa niskodimenzionalnim podacima

Prvi eksperimenti vršeni su sa najprostijim skupovima podataka – tačkama, ravnomerno raspoređenim u 2D i 3D prostoru, koje čine mrežu. Kako je za ove skupove i početna i intrinzična dimenzionalnost dva odnosno tri, nemoguće je u stilu standardnog autoenkodera kompresovati podatke u prostor niže dimenzije. Iz ovog razloga autoenkoder je imao specifičnu strukturu, kao što se vidi na slici 3.

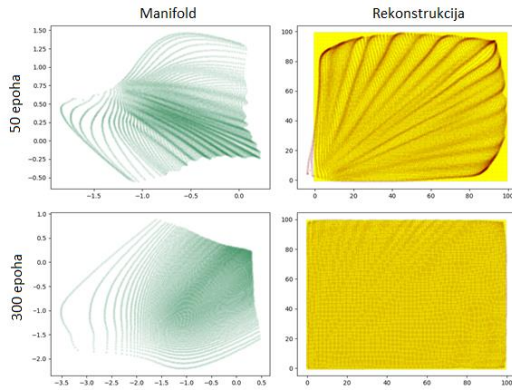


Slika 3. Arhitektura autoenkodera u inicijalnim eksperimentima

Dimenzionalnost skupa je veštački povećana da bi naknadno bila smanjena.

Rezultati eksperimenta pokazali su da duže obučavanje mreže „pegla” tj. ravna manifold, što dovodi do bolje rekonstrukcije, makar za visoko-strukturirane skupove podataka – slika 4. Takođe, do ovog efekta dovodi i pove-

ćavanje širine skrivenih slojeva u mreži tj. projektovanje podataka u više-dimenzionalne prostore pre kompresije.



Slika 4. Primer manifolda koji predstavlja mrežu tačaka u 2D i njegovu rekonstrukciju, u zavisnosti od dužine obučavanja autoenkodera.

3.2. Interpolacija i ekstrapolacija

Da bi se prevazišao problem veštačkog uvećavanja dimenzionalnosti, 2D tačke su u narednim eksperimentima prikazane kao bele tačke poluprečnika 4 piksela na crnoj slici veličine 28x28 piksela. Na ovaj način neosporno je održana $ID = 2$, ali je zato početna tj. stvarna dimenzionalnost skupa bila $28 \times 28 = 784D$, čime je autonekoderu data mogućnost da efikasno kompresuje podatke i iskaže određene karakteristike.

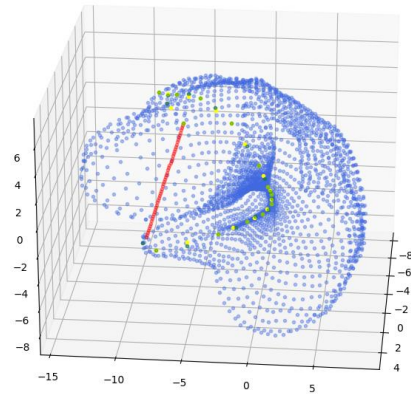
Slično prethodnom primeru, skup podataka predstavljao je skup 2D tačaka koje zauzimaju svaku od 784 moguće pozicije na slici. Mreža je trenirana 50 epoha, koristeći tanh aktivacionu funkciju, brzinu obučavanja od 0.001 i širinu središnjeg sloja (dimenzionalnost latentnog prostora LD) koji je varirao od 1 do 300.

S obzirom na ID od 2, obistinilo se očekivanje da će mreža sa LD 1, imati problem sa rekonstrukcijom primera iz skupa podataka. Za $LD > 1$ rekonstrukcija je bila uspešna, a već kod $LD=5$, konvergirala je do apsolutno tačne. Uspešna rekonstrukcija bila je početni uslov za dalje eksperimentisanje.

Za svaki od 300 eksperimenata, vršeno je poređenje između linearne interpolacije i interpolacije po manifoldu. Kao početna i krajnja tačka interpolacije uzimane su prva i poslednja tačka skupa odnosno tačke u gornjem-levom i donjem-desnom uglu slike. Kod linearne, uzrokovane su tačke sa konstantnom međusobnom udaljenošću najkraćom mogućom putanjom između prve i početne tačke u euklidskom prostoru. Interpolacija po manifoldu sa druge strane rađena je na sledeći način: Odabrane su tačke dijagonale koje predstavljaju „ground truth” u interpolaciji između zadatih tačaka. Nakon toga izostavljena je svaka druga tačka i aproksimirana sa središnjom tačkom linearne interpolacije između njoj prethodne i naredne tačke. Potom je merena prosečna udaljenost aproksimiranih tačaka od njihovih pravih vrednosti.

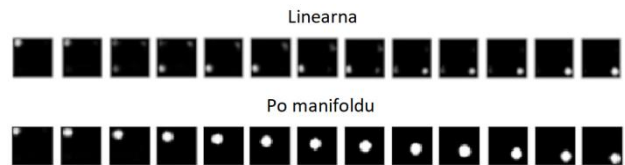
Na slici 5 može se videti primer linearne interpolacije (crvena trajektorija) i interpolacije po manifoldu (žuta trajektorija) u 3D prostoru, kao i odstupanje aproksimirane trajektorije od „ground truth” trajektorije – $mean\ dist = 0.280$.

MEAN DIST: 0.2802013



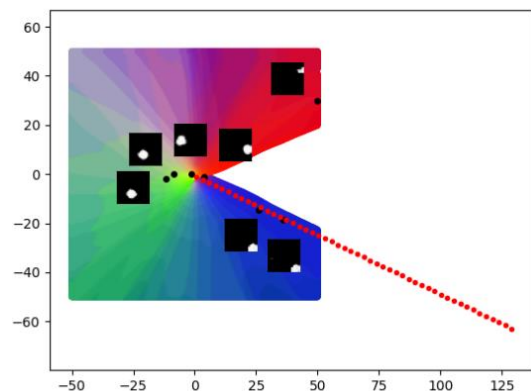
Slika 5. 2-manifold u 3D prostoru

Kao što se može videti na slici 6, linearna interpolacija je rezultirala tačkama koje sadrže deljene karakteristike svojih najbližih primera dok je interpolacija po manifoldu kreirala očekivanu translaciju tačke po dijagonali.



Slika 6. Rekonstrukcija nakon linearne i interpolacije po manifoldu

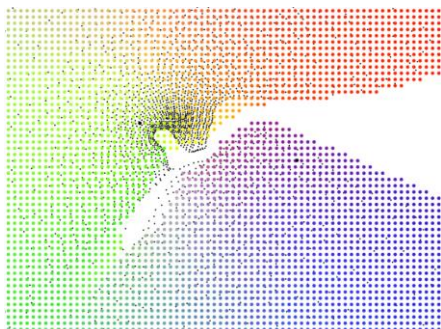
U cilju boljeg razumevanja dobijenih rezultata i izvođenja određenih zaključaka izvršen je dodatni eksperiment: uzrokovan je kompletan 2D latentni prostor u određenim granicama i svi uzorci dekodovani su u slike. Dobijene rekonstrukcije procesirane su tako da su slike čistih tačaka prikazane kao pikseli određenih boja koje koreliraju sa 2D prostorom, dok su slike sa artefaktima ili više od jedne tačke ostavljene u beloj boji. Rezultat se može videti na slici 7.



Slika 7. Uzrokovanje čitavog 2D prostora i predstavljanje rekonstruisanih slika bojama

Ovaj rezultat protumačen je na sledeći način: Informacije o određenom primeru iz skupa podataka najpreciznije su sadržane u tački koji je taj primer zauzeo na manifoldu. Ipak, čak i daleko od manifolda, te informacije ostaju zadržane i moguće je rekonstruisati primer iz udaljene tačke, dok god je ta tačka najbliža originalnoj tački. Ovo

viđenje razlikuje se od očekivanog ponašanja po kome bi uzrokovanje tačke udaljene od manifolda trebalo da rezultuje šumom. Takođe, sa slika 7 i 8 može se zaključiti da su bele, neodređene, regije prostora nastale tamo gde je manifold bio presavijen.



Slika 8. Crnom bojom prikazane su tačke manifolda

Ovaj zaključak potvrđen je i testiranjem ekstrapolacije iz određene tačke na manifoldu. Kao što se vidi na slici 7, crvenom trajektorijom prikazana je ekstrapolacija od tačke iz središta prostora u proizvoljnom pravcu. Čak i na velikoj udaljenosti, rekonstrukcija je ostala konstantna tj. zadržala je formu slike bele tačke u donjem desnom uglu.

Ovakva pojava nastavila se sve do $LD > 100$. Sa povećanjem broja dimenzija u latentnom prostoru, pojavile su se slike sa artefaktima u ekstrapolativnom režimu. Ovaj rezultat slaže se sa teorijom o visokodimenzionalnim prostorima i tezom da u takvim prostorima tačke postaju međusobno jednako udaljene pa tako poprimaju karakteristike tj. informacije svih ostalih tačaka.

Merenjem udaljenosti tačaka manifolda od koordinatnog početka kroz različite LD, ustanovljeno je da standardna devijacija raspodela tih udaljenosti postaje sve manja, tj. da se tačke koncentrišu oko određene hiper sfere ili makar u nekim njenim delovima, sve dok ne dostignu određeni limit tj. do konvergencije. Ta konvergencija se u ovom eksperimentu dešava oko dvadesete dimenzije, što se poklapa sa dimenzijom u kojoj konvergira i odstupanje aproksimirane od „ground truth” trajektorije.

5. ZAKLJUČAK

Upotrebom geometrijski visoko-strukturiranog skupa podataka sa poznatom ID i interpretabilnim rezultatima interpolacije, istraživana je unutrašnjost latentnog prostora neuronske mreže tj. autoenkodera i doneti su određeni zaključci:

- Dužina obučavanja mreže i dimenzionalnost središnjeg sloja utiču na stepen uvijenosti formiranog manifolda.
- Linearna interpolacija između tačaka u najvećem broju slučajeva, a pogotovo u prostorima visoke dimenzionalnosti, ne prolazi kroz tačke manifolda pa samim tim ne rezultira primerima iz skupa podataka.
- Neuronska mreža smešta informacije o enkodovanim primerima po celom latentnom prostoru, a ne samo u blizini manifolda. Ipak, daleko od manifolda a pogotovo kod velike LD, informacije se prepliću i nepredvidive su.

- Sa povećanjem LD, tačke manifolda koncentrišu se u blizini delova hiper-sfere, tj. razlika u razdaljinama tačaka manifolda od koordinatnog početka prostora postaje sve manja.

Ovaj rad samo je zagrebao u neistraženu prirodu strukture latentnih prostora i ostavlja mnoštvo pitanja za dalje istraživanje: Da li je moguće kretati se geodezično (sferično) po manifoldu pri dovoljno velikom LD? Kako bi uvođenje dodatnog skupa podataka različite strukture i ID uticalo na manifold? Kako se odnos GID i LID menja kroz različite LD i u kakvoj je vezi sa uvijenošću manifolda?

6. LITERATURA

- [1] Lei, Na, et al. "A geometric understanding of deep learning." *Engineering* 6.3 (2020): 361-374.
- [2] Rey, Luis A. Pérez, Vlado Menkovski, and Jacobus W. Portegies. "Diffusion variational autoencoders." *arXiv preprint arXiv:1901.08991* (2019).
- [3] Ansuini, Alessio, et al. "Intrinsic dimension of data representations in deep neural networks." *Advances in Neural Information Processing Systems* 32 (2019).
- [4] Ozair, Sherjil, and Yoshua Bengio. "Deep directed generative autoencoders." *arXiv preprint arXiv:1410.0630* (2014).
- [5] Goldt, Sebastian, et al. "Modeling the influence of data structure on learning in neural networks: The hidden manifold model." *Physical Review X* 10.4 (2020): 041044.
- [6] Ceruti, Claudio, et al. "DANCo: dimensionality from angle and norm concentration." *arXiv preprint arXiv:1206.3881* (2012).
- [7] Farahmand, Amir Massoud, Csaba Szepesvári, and Jean-Yves Audibert. "Manifold-adaptive dimension estimation." *Proceedings of the 24th international conference on Machine learning*. 2007.
- [8] Facco, Elena, et al. "Estimating the intrinsic dimension of datasets by a minimal neighborhood information." *Scientific reports* 7.1 (2017): 1-8.
- [9] Chollet, Francois. *Deep Learning with Python*. Manning Publications, 2017.

Kratka biografija:



Vuko Jovičić rođen je 1994. god. u Novom Sadu. Završio je gimnaziju „Svetozar Marković“ u Novom Sadu 2013. godine. Fakultet tehničkih nauka je upisao iste godine. Odbranio je diplomski rad 2019. godine na temu „Primena metoda mašinskog učenja u semantičkoj analizi stripa“, nakon čega nastavlja sa master studijama na istom fakultetu.
kontakt: jovicicvuko@gmail.com