



УНИВЕРЗИТЕТ У НОВОМ САДУ  
ПРИРОДНО-МАТЕМАТИЧКИ ФАКУЛТЕТ  
ДЕПАРТМАН ЗА МАТЕМАТИКУ И  
ИНФОРМАТИКУ



# Key Features Net i njene primene

Seminarski rad

Miloš Vujašinić

Novi Sad, jul 2020

# Sadržaj

<b>Sadržaj</b>	<b>1</b>
<b>Uvod</b>	<b>2</b>
<b>1 Poglavlje 1</b>	<b>3</b>
1.1 Suština podataka i problem sa tradicionalnim autoenkoderom . . . . .	3
1.2 Treniranje suštinskih autoenkodera . . . . .	4
1.3 Identičnost rezultata klasifikacije pri smanjenju dimenzionalnosti podataka . . . .	7
<b>Zaključak</b>	<b>9</b>
<b>Literatura</b>	<b>10</b>

## Uvod

Autoenkoderi su već godinama zlatni standard u smanjenju dimenzionalnosti podataka. Način na koji rade se pokazao kao veoma efikasan u otklanjanju šuma i dopunjavanju podataka koji su oštećeni. Motivisan datim primerima, ovaj rad pokušava da prikaže novi način gledanja na smanjene dimenzionalnosti podataka mašinskim učenjem. Zatim, primenom iznesenih ideja i nekih od principa koji se nalaze u osnovi autoenkodera se uvodi model neuronske mreže koji ima za cilj da iz podataka koji se prosleđuju modelu izdvoji najbitnije odlike za klasifikaciju. Kroz ovaj postupak se takođe razmatraju novi načini treniranja i evaluacije modela, a na kraju se rezultati datog modela porede sa rezultatima tradicionalnih autoenkodera.

# 1 Poglavlje 1

## 1.1 Suština podataka i problem sa tradicionalnim autoenkoderom

Autoenkoderi su neuronske mreže koje se treniraju da oponašaju identičko preslikavanje, to jest da ulaz preslikava na samog sebe. Ovo samo po sebi ne izgleda veoma korisno, međutim autoenkoderi, pored navedenih, sadrže i osobinu da je broj neurona u najmanjem skrivenom sloju mreže manji od broja neurona ulaznog i izlaznog sloja. Ovo za implikaciju ima da se prilikom prolaska podataka kroz mrežu u datom sloju nalazi reprezentacija podataka za čije je predstavljanje potreban manji broj memorijskih jedinica od originalnih podataka i ovo može da varira od nekoliko jedinica, pa do redova veličina manje. Dati skriveni sloj se naziva *usko grlo* (engl. *bottle neck*) i on deli neuronsku mrežu na dva dela: od ulaza do sebe i od sebe do izlaza. Ovi delovi se redom nazivaju *enkoder* i *dekoder*, a oni se koriste za preslikavanje originalnih podataka u reprezentaciju smanjenih dimenzija i nazad. U nastavku rada će se autoenkoderi definisati na formalniji način, ali za trenutne potrebe dato objašnjenje je dovoljno.

Prilikom treniranja autoenkodera se greška računa kao greška između svakog neurona ulaza i njemu odgovarajućeg neurona izlaza. Ovo je veoma intuitivno rešenje ako želimo da rezultati na izlazu budu naizgled što sličniji ulaznim podacima. Međutim dato preslikavanje je u praksi gotovo nemoguće napraviti tako da bude savršeno. Kao posledica dolazi do izvesne greške između originalnih podataka i rezultata mreže koji možda ne izgledaju značajno ili su nam gotovo neprimetni, ali, teoretski, mogu značajno da utiču na rezultate klasifikacije ako bi se rezultati autoenkodera pustili kroz klasifikator i uporedili sa rezultatima klasifikacije originalnih podataka. Na primer, ako imamo klasifikator koji traga za odlikom koja je sadržana u veoma malom broju memorijskih jedinica u poređenju sa veličinom jedne instance podataka, a autoenkoder zaključi da bi uvrštavanje date odlike samo povećalo grešku, jer bi njenim uvrštavanjem bilo onemogućeno uvrštavanje neke druge odlike koja je iz pogleda greške autoenkodera važnija. Kao posledica, redukcija podataka ovim postupkom je ne potencijalno samo beskorisna, nego i čini podatke neupotrebljivim ako želimo da ih klasifikujemo ili, još gore, koristimo za treniranje klasifikatora.

Diskutovani problem svakako postavlja pitanje kako se može prevazići, ali pre davanja odgovora na njega, potrebno je da definišemo podelu koja će da omogući uočavanje ovog problema jasnijim i lakšim. Iz prikazanog rezonovanja se jasno može uvideti da se autoenkoderi mogu trenirati ili u svrhu očuvanja izgleda podataka ili u svrhu očuvanja onoga što podaci ustvari jesu, odnosno njihove *suštine*. Treba imati u vidu da podaci imaju gotovo bezbroj različitih suština u zavisnosti od toga iz kog se ugla posmatraju, to jest da se u zavisnosti od pitanja koje se postavlja o njima njihova suština menja i odatle možemo da spojimo ideju o suštini sa problemom klasifikacije tako što ćemo reći da pitanje koje postavlja klasifikacioni problem određuje suštinu podataka o kojoj u datom trenutku pričamo. Ove ideje se koriste za podelu definisanu u definiciji 1. Iz ove definicije je jasno da su tradicionalni autoenkoderi ustvari izgledni autoenkoderi, dok autoenkoderi kojima se teži u svrhu prevazilaženja diskutovanog problema su suštinski autoenkoderi, mada se i za njih, u zavisnosti od implementacije, postavlja pitanje koliko dobro mogu da očuvaju odlike bitne za klasifikaciju.

**Definicija 1 (Podela autoenkodera po načinu treniranja)** *Autoenkoder za čije se treniranje koristi znanje o samim podacima se naziva izgledni autoenkoder, dok autoenkoder za čije se treniranje koristi znanje o klasama kojima podaci pripadaju se naziva suštinski autoenkoder. Treba imati u vidu da date osobine nisu međusobno disjunktne, pa postoji i izgledno-suštinski autoenkoder.*

## 1.2 Treniranje suštinskih autoenkodera

Da bismo razmatrali kako se klase kojima podaci pripadaju mogu uključiti u treniranje autoenkodera definišimo prvo autoenkodere na formalniji način. Za ovo ćemo iskoristiti definiciju 2 preuzetu iz [1, Poglavlje 2].

**Definicija 2 (Opšti autoenkoder *framework* [1])**  $n/p/n$  autoenkoder je definisan kao  $t$ -torka  $n, p, m, \mathbb{F}, \mathbb{G}, \mathcal{A}, \mathcal{B}, \mathcal{X}, \Delta$  gde važi:

1.  $\mathbb{F}$  i  $\mathbb{G}$  su skupovi.
2.  $n$  i  $p$  su pozitivni celi brojevi. (Autor razmatra slučaj kada je  $0 < p < n$  što se podrazumeva prilikom pravljenja autoenkodera)
3.  $\mathcal{A}$  je klasa funkcija koje preslikavaju ulaz iz  $\mathbb{G}^p$  na  $\mathbb{F}^n$
4.  $\mathcal{B}$  je klasa funkcija koje preslikavaju ulaz iz  $\mathbb{F}^n$  na  $\mathbb{G}^p$
5.  $\mathcal{X} = \{x_1, \dots, x_m\}$  je skup od  $m$  (trening) vektora u  $\mathbb{F}^n$ . Ako se radi o spoljašnjem skupu ciljeva u  $\mathbb{F}^n$  označavamo ga sa  $\mathcal{Y} = \{y_1, \dots, y_n\}$ .
6.  $\Delta$  je funkcija različitosti ili distorcije definisana nad  $\mathbb{F}^n$

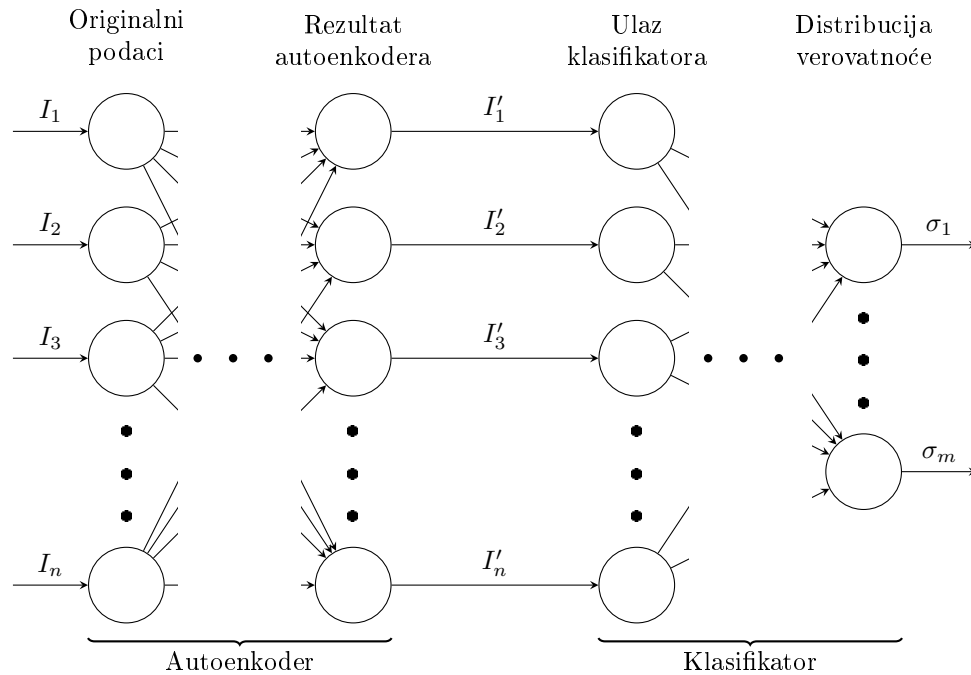
Da bismo bolje razumeli datu definiciju i bliže se upoznali sa terminologijom koja će se koristiti u nastavku prođimo kroz sve elemente date  $t$ -torke koja se koristi za predstavljanje autoenkodera. Kao što smo ranije pričali ideja iza autoenkodera je da nauči da slika ulaz na samog sebe, a da pri tome postoji skriveni sloj koji je manjih dimenzija od samog ulaza i na ovaj način se postiže smanjenje dimenzionalnosti podataka.  $n$  je ništa manje nego broj dimenzija ulaza kao i izlaza, dok je  $p$  broj dimenzija *uskog grla*<sup>1</sup>, pošto se cilja da je broj dimenzija uskog grla manji od broja dimenzija ulaza zato se i pretpostavlja da je  $0 < p < n$ . Svaka memorijska jedinica u ulaznom i izlaznom sloju, odnosno uskog grlu sadrži podatke koji redom pripadaju skupovima  $\mathbb{F}$  i  $\mathbb{G}$ .  $\forall B \in \mathcal{B}$  je funkcija koja preslikava ulazne podatke u reprezentaciju smanjenih dimenzija i ona se naziva *enkoder*.  $\forall A \in \mathcal{A}$  radi obrnut proces od funkcija iz  $\mathcal{B}$  i preslikava reprezentaciju smanjenih dimenzija na izlaz i ova funkcija se naziva *dekoder*. Treba primetiti da se ovde delovi neuronske mreže posmatraju kao funkcije, jer oni to i jesu, i ova notacija će se u nastavku koristiti kada se govori o neuronskim mrežama i njenim delovima. Za treniranje autoenkodera su potrebni podaci i skup trening podataka je u datoj definiciji označen sa  $\mathcal{X}$ . Međutim, podaci se ne koriste samo za treniranje i možemo imati podatke koji se koriste za evaluaciju ili podatke koji se puštaju kroz autoenkoder u cilju smanjenja dimenzionalnosti. Ovakvi skupovi podataka su u definiciji označeni sa  $\mathcal{Y}$ . Na kraju, ostala je još da se objasni funkcija različitosti ili distorcije  $\Delta$  koja igra važnu ulogu u pprocesu treniranja autoenkodera. Cilj treninga autoenkodera je pronalazak funkcija  $A \in \mathcal{A}$  i  $B \in \mathcal{B}$  za koje je vrednost funkcije distorcije minimalna nad celim skupom trening vektora  $\mathcal{X}$ . Funkcija distorcije  $\Delta$  je primer onoga što kod neuronskih mreža znamo kao funkcija greške ili *loss* funkcija.

Funkcija greške je jedna od glavnih karakteristika treninga neuronskih mreža i ona govori u kom smeru će se treniranje kretati. Kao takva ona je očigledna opcija za uključanje klase kojima podaci pripadaju u trening i na taj način dobijanje suštinskog autoenkodera. Međutim, ovo nije tako lako izvesti kao što zvuči. Dok se kod tradicionalnih, izglednih, autoenkodera srednja kvadratna greška nudi kao idealna funkcija greške, kod suštinskih autoenkodera stvari su malo komplikovanije i zahtevaju se određeni trikovi u cilju dobijanja željenog efekta. Glavni problem je što kod treniranja izglednih autoenkodera je vrednost koju mreža treba da proizvede poznata,

<sup>1</sup>sloj u kome se prilikom prolaza podataka kroz autoenkoder nalazi reprezentacija podataka smanjenih dimenzija

dok kod suštinskih autoenkodera je poznata samo vrednost koja treba da se dobije kada se traženi rezultati puste kroz određenu funkciju. Na osnovu ovoga definicija 3 uvodi dva načina treniranja na koja se model mašinskog učenja može učiti.

**Definicija 3 (Podela načina treniranja modela mašinskog učenja)** *Ako je prilikom treniranja modela mašinskog učenja poznata vrednost koja se treba dobiti kao proizvod rada modela onda se dati proces treniranja naziva direktno treniranje. Nasuprot tome, treniranje u kome je poznata samo vrednost koja se treba dobiti nakon što se rezultati rada modela puste kroz određenu funkciju se naziva indirektno treniranje.*



Slika 1: Model klasifikatora dobijen spajanjem autoenkodera i manjeg klasifikatora

Iako su suštinski autoenkodera dati kao primer indirektnog treniranja, uz pomoć malog trika, se može napraviti da njihovo treniranje bude deo direktnog procesa treniranja. Pošto znamo da rezultat suštinskog autoenkodera kada se pusti kroz klasifikator treba da da nama poznatu klasu, možemo ovo svesti na istovremeno treniranje autoenkodera i klasifikatora. Ideja je da imamo strukturu autoenkodera i klasifikatora i da od njih napravimo novu neuronsku mrežu tako što ćemo staviti da ulaz prvo prolazi kroz autoenkoder, potom da izlaz autoenkodera bude ulaz u dati klasifikator, a sve zajedno treniramo kao da je jedan veliki klasifikator. Pošto znamo da je izlaz autoenkodera istih dimenzija kao i ulaz, koji je ujedno istih dimenzija kao i ulaz klasifikatora, nema problema prilikom prosleđivanja rezultata autoenkodera klasifikatoru, jer su dimenzije odgovarajućih slojeva jednake. Grafik na kojem je prikazan jedan ovakav model se može videti na slici 1. Ova ideja iako zvuči moguće ipak sa sobom nosi dva problema. Prvi je što imamo veliki model što za posledicu ima duže vreme treniranja i može da naškodi kvalitetu istreniranog modela. Drugi problem je ono što čini dati pristup beskorisnim, ali da bismo njega shvatili moramo se vratiti na koncept suštine podataka i njene korelacije sa originalnim podacima. Suština podataka je rezultat razmišljanja kako iz podataka može da se ukloni što je više moguće osobina

tako da rezultat klasifikacije podataka koji sadrže preostale osobine bude isti kao rezultat klasifikacije originalnih podataka. Ovo bi značilo da ako imamo savršen suštinski autoenkoder i znamo da se ulazni podaci klasifikuju u određenu klasu da bi onda i rezultat autoenkodera za date podatke morao da se klasifikuje u istu klasu, kao i obrnuto. Rezultat prethodno opisanog procesa treniranja bi bili autoenkoder i klasifikator i iako bi rezultat klasifikacije kroz dati klasifikator podataka propuštenih kroz dati autoenkoder težio ka klasi ka kojoj je treniran, u datom procesu treniranja ne postoji mehanizam koji obezbeđuje da će originalni podaci biti dobro klasifikovani kada se puste kroz dati klasifikator. Ovo kao posledicu ima da je dobijeni klasifikator neispravan, što dalje uzrokuje da se dobijeni autoenkoder ne može koristiti, jer je treniran na neispravnom klasifikatoru.

Prethodno diskutovani proces treniranja, kao što smo videli, sadrži mane i probleme koji ga čine neupotrebljivim, međutim on postavlja dobru osnovu u načinu i smeru u kojem treba da se razmišlja o ovom problemu. Ono što smo uvideli iz prethodne diskusije je da je za pravljenje ispravnog suštinskog autoenkodera potreban ispravan klasifikator, a mašinsko učenje i neuronske mreže se bave temom klasifikatora već dugo vremena i znamo kako se on može napraviti. Kada imamo istreniran klasifikator možemo da iskoristimo ideju da spojimo autoenkoder i klasifikator i treniramo novi model kao jedan veliki klasifikator, međutim moramo da zamrznemo sve parametre već istreniranog klasifikatora, jer samo tako možemo da budemo sigurni da će istreniran klasifikator biti ispravan tokom celog treninga i da će se izbeći problem iz prethodnog primera. Realizacija ovoga je po prilici veoma prosta, i dalje se pušta nazadna propagacija<sup>2</sup> kroz celu mrežu, ali uz izmenu da menja samo parametre autoenkodera, dok parametri istreniranog klasifikatora preskače kada na njih dođe red da se izmene. Dati autoenkoder se na ovaj način trenira da preslikava ulaz na rezultat koji se klasifikacijom kroz dati istrenirani autoenkoder svrstava u traženu klasu.

Iako je prethodna ideja veoma blizu rešenju koje problem zahteva, ona ipak ima jednu manu. Ako se prisetimo priče o suštini podataka, rekli smo da se originalni podaci i rezultat autoenkodera trebaju klasifikovati u istu klasu, a nema garancije da će istrenirani klasifikator ispravno klasifikovati originalne podatke, dok mi treniramo autoenkoder kao da je ovo slučaj. Iz ovog razloga treba da napravimo izmenu onoga čemu veliki klasifikator, koji se sastoji od autoenkodera i istreniranog klasifikatora, cilja. Umesto da vodi trening ka klasi kojom su originalni podaci labelovani treba da cilja na distribuciju verovatnoće<sup>3</sup> koja se dobije kada se originalni podaci puste kroz istrenirani klasifikator.

Diskutovano rešenje je, kao što vidimo, ispravno, ali je računski veoma zahtevno i bilo bi idealno ako bi mogao da se nađe način da se dati proces ubrza. Ono što može da se uradi po ovom pitanju je da popustimo uslov i ne prolazimo kroz ceo istrenirani klasifikator tokom treninga, nego da idemo samo do određenog sloja datog klasifikatora i treniramo autoenkoder tako da vrednosti u datom sloju kada se rezultati puste kroz klasifikator budu iste kao i u slučaju originalnih podataka. Ono što je interesantno je da je način treniranja tradicionalnih autoenkodera i prethodno diskutovano rešenje ustvari specijalan slučaj ovog pristupa: ako se uzme samo ulazni sloj mreže, dobija se tradicionalno rešenje, dok ako se uzme ceo autoenkoder dobija se originalno rešenje izneseno u ovom radu. Realizacija ove ideje je analogna originalnom rešenju, međutim prilikom izbora ovog pristupa treba biti svestan da on sa sobom nosi manu da u opštem slučaju što se uzme manji broj slojeva da će rezultat klasifikacije podataka dobijenih autoenkoderom biti dalje od rezultata klasifikacije originalnih podataka.

Svi autoenkoderi o kojima smo do sada pričali ili su bili čisto izgledni ili čisto suštinski autoenkoderi, a ako se prisetimo definicije 1, u njoj smo uveli i treću vrstu koja kombinuje ove

---

<sup>2</sup>algoritam koji se koristi za treniranje neuronskih mreža

<sup>3</sup>klasifikator preslikava ulazne podatke na distribuciju verovatnoće koja govori sa kojom verovatnoćom podaci pripadaju svakoj od mogućih klasa

dve diskutovane ideje. Kombinovanje ove dve ideje se može izvesti na veoma jednostavan način: ako imamo funkciju distorcije  $\Delta_1 : \mathbb{A} \rightarrow \mathbb{B}_1$  specifičnu za treniranje izglednog autoenkodera, funkciju distorcije  $\Delta_2 : \mathbb{A} \rightarrow \mathbb{B}_2$  specifičnu za treniranje suštinskog autoenkodera, kao i binarni operator  $\oplus : \mathbb{B}_1 \times \mathbb{B}_2 \rightarrow \mathbb{B}$  možemo definisati funkciju distorcije  $\Delta' : \mathbb{A} \rightarrow \mathbb{B} : x \rightarrow \Delta_1(x) \oplus \Delta_2(x)$  koja je specifična za treniranje izgledno-suštinskog autoenkodera. Primere funkcija distorcije  $\Delta_1$  i  $\Delta_2$  već znamo ili smo ih diskutovali u ovom radu, dok za operator  $\oplus$  možemo izabrati neki od operatora koji se najčešće koriste, na primer operator sabiranja  $+$  ili operator množenja  $\cdot$ . Ono o čemu treba voditi računa prilikom primene biblioteka za rad sa neuronskim mrežama je da neke od njih rezultat funkcije greške dobijaju kao višedimenzionalne tenzore koji ne moraju strogo biti istih dimenzija i zato se treba primeniti funkcija koja će date vrednosti da preslika na rezultate koji pripadaju skupovima na koje se može primeniti odabrani operator.

### 1.3 Identičnost rezultata klasifikacije pri smanjenju dimenzionalnosti podataka

Diskutovali smo suštinske autoenkodere i načine na koje se oni mogu trenirati, sledeće logično pitanje je kako možemo da ocenimo kvalitet suštinskog autoenkodera i uporedimo dva ovakva autoenkodera da odredimo koji je bolji. Ideja suštinskih autoenkodera se fundamentalno razlikuje od gotovo svega do sada viđenog u oblasti neurosnkih mreža i mašinskog učenja, s toga poznate standardne metrike nisu dobar pokazatelj kvaliteta modela ovog tipa autoenkodera. Jedna ideja je da ako imamo skup podataka za evaluaciju  $\mathcal{Y} = \{y_1, \dots, y_k\}$  da pustimo podatke prvo kroz autoenkoder, a zatim proverimo koliki je procenat njih svrstan u klasu u koju su labelovane. Ovo izgleda kao dobra metrika, ali ona nije u potpunosti u skladu sa motivacijom iza suštine podataka. Ako se setimo cilj iza suštine podataka je da se podaci u što većoj meri klasifikuju u istu klasu u koju i rezultati kada se ti isti podaci puste kroz suštinski autoenkoder. Možemo pogledati definiciju 4 koja definiše pojam identičnost rezultata klasifikacije pri smanjenju dimenzionalnosti podataka, u nastavku samo identičnost rezultata, koja je upravo mera ovoga do sada diskutovanog. Ova mera predstavlja procenat trening vektora iz skupa  $\mathcal{Y}$  koji su klasifikovani klasifikatorom  $g$  u istu klasu u koju su istim klasifikatorom klasifikovani i rezultati kada se isti ti trening vektori puste kroz autoenkoder  $g$ .

#### Definicija 4 (Identičnost rezultata klasifikacije pri smanjenju dimenzionalnosti)

Za autoenkoder  $f : \mathbb{A}^n \rightarrow \mathbb{A}^n$  i klasifikator  $g : \mathbb{A}^n \rightarrow \mathbb{B}$  (relacija = mora biti definisana nad  $\mathbb{B}$ ) koji vraća klasu u kojoj se ulaz sa najvećom verovatnoćom nalazi definišemo meru identičnost rezultata klasifikacije pri smanjenju dimenzionalnosti podataka  $\Upsilon$  nad skupom  $\mathcal{Y} = \{y_1, \dots, y_k\}, \mathcal{Y} \subset \mathbb{A}^n$  kao:

$$\Upsilon(f, g; \mathcal{Y}) = \frac{|\{y \mid y \in \mathcal{Y} \wedge g(y) = g(f(y))\}|}{|\mathcal{Y}|}.$$

Problem kod identičnosti rezultata je pitanje izbora klasifikatora koji ćemo da koristimo u datoj metrici i kao logičan odgovor na ovo se daje klasifikator koji je korišćen tokom treniranja autoenkodera. Međutim, ovde se lako može primetiti da se tako može dobiti da je identičnost rezultata na datom klasifikatoru visoka, a na ostalim klasifikatorima koji se koriste za isti klasifikacioni problem veoma mala, to jest određeni vid *overfitting*-a. Ovo budi mnogo dublja pitanja. Ako se prisetimo, mi do sada još nismo formalno definisali suštinu podataka i najbliže što smo rekli o njoj je da ako se iz podataka uklone sve osobine koje nisu bitne za klasifikaciju da ono što ostaje je ustvari suština podataka. Ono što je ovde problem je ako postoje dva klasifikatora  $g_1$  i  $g_2$  koji redom tragaju za skupovima osobina  $\mathcal{G}_1$  i  $\mathcal{G}_2$  nad određenim podacima u svrhu pronalaženja odgovora na isto klasifikaciono pitanje i na to sve još važi da je  $\mathcal{G}_1 \cap \mathcal{G}_2 \neq \emptyset$ . Preneseno rečima,



ovo znači da dati klasifikatori tragaju za različitim osobinama u svrhu klasifikacije podataka, što bi značilo da suština podataka nije jedinstvena između klasifikatora koji rešavaju isti klasifikacioni problem. Ovaj problem izlazi iz okvira ovoga rada i s toga ostaje otvoreno pitanje, a ako je odgovor potvrđan, što je verovatno i slučaj, treba razmisliti o načinu da se suština iz ugla više klasifikatora može spojiti u jednu celinu. Do trenutka dok se ne odgovori na ova pitanja svako korišćenje suštinskih autoenkodera sa klasifikatorima sa kojima dati autoenkoder nije treniran je nepreporučljivo.

## **Zaključak**

## Literatura

- [1] Pierre Baldi. Autoencoders, unsupervised learning, and deep architectures. In Isabelle Guyon, Gideon Dror, Vincent Lemaire, Graham Taylor, and Daniel Silver, editors, *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, volume 27 of *Proceedings of Machine Learning Research*, pages 37–49, Bellevue, Washington, USA, 02 Jul 2012. PMLR.