Univerzitet u Beogradu

Elektrotehnički fakultet



Neuralne mreže

- prvi projektni zadatak -

Mentor : Marija Novičić, asistent

Studenti : Filip Galik 2020/0290

Nikola Zlatanović 2020/0212

februar 2024.

**OPIS PROBLEMA**

Na raspolaganju su nam rentgenske slike pluća (sive slike), i to slike zdravih pluća i zaraženih pneumoniom, dakle dataset koji smo odabrali sadrži 2 klase : **normal** i **pneumonia.**

Cilj ovog projekta je projektovanje konvolucione neuralne mreže koja će biti sposobna da razlikuje te dve klase, odnosno da uspešno klasifikuje odbirke iz tih klasa, kao i da ostvaruje veliku **senzitivnost**.

Ovakva mreža ima ogromu praktičnu vrednost, jer se njen izlaz može upotrebiti kao početna dijagnoza pacijenta, naravno da krajnju dijagnozu donosi ekspert (doktor), dok mu mreža daje **sugestiju .**

Još je bitno napomenuti da ako imamo ogroman broj pacijenata, ekspertu će trebati puno vremena da klasifikuje ručno sve rentgenske slike, zbog toga može upotrebiti neuralnu mrežu koja će prepoznati **karakteristične** **slučajeve** i samo njih klasifikovati (verovatnoća greške je vrlo mala), dok će odbirke za koje nije sigurna poslati ekspertu.

Prikaz određeno izabranih pacijenata :



**OPIS DATASET-A**

Dataset koji koristimo za ovaj projekat sadrži 5563 rentgenske slike pluća koje su labelirane sa **normal** i **pneumonia**. Podaci jesu podeljeni u 3 skupa (train,val,test), mada nama se nije svidela razmera podele tako da smo sami podelili kompletan skup.

Klase nisu balansirane, jer klasa **normal** ima oko 1500 uzoraka dok klasa **pneumonia** ima oko 4000, dakle nemaju približno isti broj odbiraka. Na sledećem grafiku je prikazan histogram dataset-a :



Koristimo metod **class\_weights** koji svakoj klasi pridruži težinu kojom se množi inkrement greške pogrešne klasifikacije odbiraka iz te klase (dakle pogrešnu klasifikaciju odbiraka iz neke klase više penalizujemo, čime povećavamo važnost te klase).

Dataset smo podelili na trening i test, a trening skup delimo za krosvalidaciju. Podela podataka je bitna jer ne možemo nad istim skupom i projektovati klasifikator (neuralnu mrežu) i testirati klasifikator, jer će estimacija tačnosti tog klasifikatora biti pomerena, jer je klasifikator već “video” te podatke. Dakle, dobijena estimacija tačnosti klasifikator će biti **optimistična.**

**Pravilna podela podataka je bitna zbog dobijanja što bolje estimacije performansi klasifikatora.**

Da bi dobili nezavisnu procenu tačnosti klasifikatora moramo testirati klasifikator na skupu podataka koje nije “video”, odnosno skupu nad kojim nije projektovan.

I dalje ova estimacija tačnosti klasifikatora nije savršena, odnosno dosta zavisi od nasumične podele dataset-a na podskupove, zato su izmišljene još bolje tehnike za estimaciju tačnosti koje se grupno nazivaju **krosvalidacija**.

Najčešće primenjivana tehnika iz te grupe metoda je **KFold** krosvalidacija koja podrazumeva da trening skup podelimo na K poskupa i projektujemo klasifikator K puta, pri čemu se u svakoj iteraciji jedan podskup koristi za validaciju, a ostali za projektovanje klasifikatora.

Konkretno, za parametar K smo uzeli vrednost 5 (80% trening skupa koristimo za projektovanje mreže, ostatak za validaciju). Iskustveno je poznato da je ova vrednost za K odlična, jer val skup nije ni preveliki ni premali.

Dakle, mreža se obuči na poskupovima za trening, a performanse se provere na validacionom skupu. Procena tačnosti mreže je srednja vrednost tačnosti na validacionim skupovima.

Bitno je naglasiti da i dalje moramo imati test skup za finalnu proveru tačnosti mreže, jer time, na neki način, simuliramo praktični rad obučene mreže.

Krosvalidacija daje dosta realniju i tačniju procenu tačnosti mreže, njena estimacija tačnosti klasifikatora ne zavisi od slučajne podele odbiraka u podskupove.

**PREDPROCESIRANJE PODATAKA**

Pikseli slika su skalirani na opseg od 0 do 1 (početne slike su uint8 matrice), takođe, vršili smo augumentaciju trening skupa (skaliranje, rotacija, okretanje) da bi dobili više podataka za trening i da bi učinili mrežu robusnijom na neke pojave sa kojima se mreža susreće u praksi.

**ARHITEKTURA I HIPERPARAMETRI MREŽE**

Neuralna mreža koju koristimo za rešavanje ovog problema klasifikacije naziva se konvoluciona neuralna mreža.

Ona je **struktuirana**, što znači da neuroni iz predhodnog sloja utiču samo na neke neurone u sledećem sloju, za razliku od klasične (FC) neuralne mreže kod koje svaki neuron iz predhodnog sloja utiču na sve neurone u sledećem sloju.

Sastoji se iz 2 glavna dela : konvolucioni slojevi i FC slojevi. Konvolucioni slojevi vrše izdvajanje obeležja dok FC slojevi vrše klasifikaciju na osnovu tih obeležja.

Kriterijumska f-ja koju biramo se naziva **SparseCategoricalCrossEntropy**, jer svaki odbirak pripada tačno jednoj klasi (ova kriterijumska f-ja može se primeniti samo pod tim uslovom), takođe, na osnovu iskustva je poznato da su loss f-je tipa kros-entropije vode ka odličnim podešavanjima mreže za problem klasifikacije.

F-ja aktivacije neurona koju smo izabrali je **ReLu**, jer nema saturaciju, ima veliki i konstantni gradijent i računski nije zahtevna .

Metoda kojom optimizujemo kriterijumsku f-ju se naziva **Adam** (bitno je reći da on računa prosečni drugi izvod loss f-je, čime postiže stabilnu konvergenciju i za veće konstante obučavanja)

**Arhitektura modela CNN** : 3 konvoluciona sloja i 2 maxpool sloja između, zatim jedan FC sloj (128 neurona) i na kraju izlazni sloj (softmax). Konvolucioni slojevi sadrže 16, 32, 64 filtra , respektivno, koji imaju 3x3 kernel, maxpool slojevi imaju 2x2 kernel. Čitava mreža sadrži 2 121 122 (oko 2M) parametara.

**ZAŠTITA OD PREOUBUČAVANJA**

**Preobučavanje** se definiše kao pojava da neuralna mreža iz trening skupa nauči i outlajere i statistički šum (što nije poželjno, zbog degradacije performansi mreže).

Značajno umanjujemo preobučavanje primenom tehnike Dropout-a, koja podrazumeva da tokom treninga nasumično izbacujemo neke neurone iz mreže, čime se postiže promenljiva arhitektura mreže tokom treninga. Tokom testiranja koristimo sve neurone.

Takođe, dodali smo i L2 regularizaciju koja, takođe, smanjuje preobučavanje, tako što modifikuje loss funkciju, tako da sprečava ekstremna odsupanja težina mreže.

**PRIKAZ PERFORMANSI MREŽE**

U ovom poglavlju ćemo prikazati performanse i matrice konfuzije projektovane mreže, kao i neke primere dobro i loše klasifikovanih pacijenata.







Kfold metod proizvodi 5 modela, od kojih biramo onaj koji ostvaruje najveću **tačnost** na validacionom splitu. Bitno je napomenuti da je kriterijum za izbor modela **tačnost**, a ne F1 skor, iako su klase nebalansirane, jer u ovom konkretnom zadatku želimo da model ima bajes prema PNEUMONIJA klasi, što povećava osetljivost modela na tu klasu (samo treba paziti da preciznost nije premala, na šta ne moramo brinuti ako je metrika tačnost a ne senzitivnost, jer tačnost uzima u obzir i prvu klasu), što je traženo u zadatku.

Sa konfuzionih matrica vidimo da je senzitivnost modela velika i na trening i na test skupu, ali po cenu manje preciznosti na test skupu.

Dakle, tokom treninga model uči da razlikuje zdrave od bolesnih pacijenata, validacija nam služi da nađemo model koji ima najbolji osetljivost-preciznost tradeoff, zatim test nam daje konačnu procenu performansi izabranog modela.

Performanse mreže su odlične (grafici izgledaju oscilatorno, ali to je zbog male skale po y osi).

Konačno, prikaz dobro i loše klasifikovanih pacijenata :

