LABOS 4 DUBOKO PITANJA

Neka pitanja na blicu kojih se sjećam:

1. Koliko ima parametara batchnormconvrelu ako je kernel veličine 3, c\_in=c\_out=32 (9312)

A screenshot of a computer

Description automatically generated

1. Kojeg reda je tenzor koji vraća funkcija get\_features

(B, emb\_size ) gdje je B velicina batcha, npr moze bit (64, 32)

1. Koje su veličine emv\_size za 2 modela koja smo implementirali. 32 i 784

(Model 1: emb\_size = 32 | Model 2 (PCA ili bez obrade): emb\_size = 784 ???? )

1. Koja je povezanost Module i Sequential ( sequential nasljeduje module )
2. Ako imamo K klasa i svaka klasa ima M primjera, koliko mozemo generirati pozitivnih i negativnih primjera za jedan anchor primjer.

( M -1, (K – 1) \* M)

Neka pitanja asistenta:

1. Objasni kako tripletmarginloss funkcionira i koja je tu uloga margina. Što je veći margin, nekako tražimo da su veće razlike između d(a,p) i d(a,n)
2. Gledajući samo na grafove iz 4 zadatka, koji bi graf pripadao kojem modelu? Onaj koji bolje razdvaja je simpleembeding model, jer on sam po sebi ima veći kapacitet učenja, tj složeniji je.
3. Koja je razlika između anchor, positive i negative primjera i kako bismo dobili jedan skup ta tri primjera.

A screenshot of a computer

Description automatically generated

odnos razreda nn.module i nn.sequential ( sequential nasljeduje module )

get\_features u identitymodelu koliki je tenzor ( 1 x 784 ) ( mozda je tocan odgovor (B, 784) gdje je B velicina batcha, 784 je jer mnist ima slike dimenzije 28x28 i ovo to flattena u 784

sto sve sprema torch.save() parametre modela ( tezine i bias )

kako se određuje positive i negative za klasu c (tipa positive==c^i; negative != c^i) ( za klasu c^i pozitivni je label == c^i, negativni je label != c^i )

kojeg je reda tenzor kojeg vraca get\_features 2d

koji tip klasifikacije ( radi se o: **Metričkom učenju**, koje nije klasična klasifikacija, već **učenje prostora značajki** u kojem su slični uzorci bliže. | Dakle: **k-nn klasifikacija u prostoru značajki (embeddinga)**. ????)

dimezija compute\_reprezentations return tenzora ( broj\_klasa, emb\_size) , konkretno MNIST ima 10 klasa i nek je emb\_size = 32 onda je (10, 32)

koliko kanala ima ulaz u prvi sloj konvolucije (input\_channel = 1 jer kositimo MNIST koji ima 1 kanal jer su crno-bijele sliek )

koji je posljednji sloj u simpleMetricEmbedding modelu ( Global Average Pooling (GAP) )

pitanja s blica:

* sto rade model.eval i model.train ( model.train(): Aktivira **trening način rada,** Omogućava ponašanje slojeva poput Dropout i BatchNorm specifično za treniranje,   
  model.eval():  
  Aktivira **eval (inferencijski) način rada**Dropout i BatchNorm rade deterministički (nema slučajnosti))
* koliko kanala ima ulaz u prvu konvolucijsku jedinicu (Ako se koristi **MNIST**, ulazne slike su 1-kanalne)
* izmedju koliko pozitivnih i negativnih primjera mozemo birati kad odaberemo anchor ako u svakoj od K klasa ima M primjera? (ovdje mislim da je M-1, (K-1)\*M odgovor) ( M – 1 primjera je ostalo iz pozitivne klase jer smo jedan izabrali, a za negativnu imamo K – 1 klasa(ne smijemo pozitivnu klasu) i svaka klasa ima M primjera znaci (K – 1)\*M)
* koja je povratna vrijednost \_sample\_positive funkcije ( **Cijeli broj (eng. int) — indeks slike** iz iste klase kao anchor, ali različit od index.)

koja je povratna vrijednost one torch.pca\_.. funkcije) ( **torch.pca\_lowrank(...) vraća tuple od tri torch.Tensor-a**: (U, S, V) = torch.pca\_lowrank(X, q)  **U:** glavni vektori — **projekcja podataka** na q dimenzija → veličine **(n, q)**, **S**: singularne vrijednosti → veličine **(q, )** **V:** glavne komponente → veličine **(m, q))**

* koja metoda se koristi za klasifikaciju (logisticka regresija, linearna regresija, najblizi susjedi...)
* Kojih je dimenzija tensor koji vraca mean\_representstion ( broj\_klasa, emb\_size) znaci vj je dimenzija (10,32)

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a black screen

Description automatically generated

LABOS 3 PITANJA

A screenshot of a white board

Description automatically generated

Pitanje za blic koje nisam prije vidjela: koja je vrijednost gradijenta funkcije gubitka, ulaz softmaks, jednu iteraciju RNN  
  
1. Nelinearnost u običnoj povratnoj mreži?

tangens hiperbolni (tanh)

2. Funkcija gubitka u labosu?

unakrsna entropija

3. Koji parametar uzrokuje eksplodirajući gradijent?

Whh

4. Optimizator u labosu?

Adam – zadano u zadatku

5. Derivacija tangensa hiperbolnog (tanh)?

1 - np.tanh \* np.tanh

6. Formula za ažuriranje skrivenog stanja?

h[t] = tanh( Whh \* h[t-1] + Wxh \* x[t] + bh)

7. Za model koji nema povratne slojeve, kako bi se promijenila točnost ako nasumično promijenimo poredak riječi?

a) točnost bi se smanjila

b) točnost bi se povećala

\*c) ne bi se promijenila (ako nije povratno onda ne gleda redosljed)

d) ?

8. Veličina vokabulara i dimenzija ulaza su 100, skrivena dimemzija 1000. Koja matrica ima najviše parametara?

\*a) Whh

b) Why

c) Wxh

d) by

9. Veličina vokabulara i dimenzija ulaza su 100, skrivena dimemzija 1000. Koliko parametara ima povratna mreža (uključujući i izlazni sloj)? Mreža klasificira sljedeću riječ.

ja sam ovo dobila: 1201100 (Whh + Wxh + bh + Why + by = 1000\*1000 + 100\*1000 + 1000 + 1000\*100 + 100)

10. Može li se doći do gubitka 0 ako je veličina vokabulara neograničena i pretpostavljamo da neće doći do brojčanog nadljeva i podljeva?

a) ne, nikad

b) da, s dovoljno dugim treniranjem

c) da, s veličinom grupe 1

d) da, s dovoljno velikim skrivenim stanjem  
  
  
Evo pitanja sa Studšoa od prošle godine:

Koji smo optimizer koristili u labosu (adam)

Koji od navedenih tokena se pojavljuju u labosu (pad i unk)

Koju smo funkciju pogreske koristili u labosu (cross entropy)

Kolko parametara ima rnn celija ako nam je poznato bla bla bla

Formula za rnn (to je ono tanh(Whh \* h[t-1] + Wxh \* x[t] ))

koja je derivacija tangensa hiperbolnog ( 1 – tanh^2())

Formula za LSTM: c(t)=…

A math equations on a white background

Description automatically generated

2) koji je raspon slike tangensa hiperbolnog tanh element (-1,1)

4) može li se pri treniranju pojaviti <UNK> simbol (da, ako rijec nije usla u vokabular )

5) računanje broja parametara za RNN, ne sjećam se točnih dimenzija

1. Cemu sluzi gradient clipping? (**Služi za sprječavanje eksplozije gradijenata** kod treniranja RNN-a (posebno pri dugim sekvencama). Ograničava L2 normu gradijenata ispod određene vrijednosti (npr. 1.0).)
2. Za sto koristimo <PAD> <PAD> (padding token) koristimo za **poravnanje (engl. padding)** nizova različite duljine u **jednake duljine**, kako bi se mogli **obrađivati u batchu**.

Pitala me me po cemu je model GRU razlicit od obicnog rnn-a. ( **GRU (Gated Recurrent Unit)** uvodi **gate-ove** koji kontroliraju **protok informacija** i omogućuju **lakše pamćenje dugoročnih odnosa**, dok obični RNN to ne može učinkovito. )

Koja je razlika izmedu basline modela i rnn-a. ( baseline je Feedforward, ignorira redosljed rijeci )

Koji gubitak koristim i kakav je to gubitak tj sto vraca. (  Uspoređuje predikciju modela (vjerojatnosti po klasama) s točnom klasom.Računa koliko su predikcije daleko od stvarnih oznaka. **Skalarna vrijednost (float)**: veći broj = lošija predikcija, manji broj = bolja predikcija. )

Da skiciram kak rnn izgleda.

Koje hiperparametre ima model. ( Learning rate, Broj epoha, Batch size, Dimenzija skrivenog stanja, Dimenzija embeddinga, Dropout, Optimizator, Broj slojeva, Tip RNN-a)

Sto je skriveno stanje h i koji su tu moguci problemi.

Gdje je izlazni sloj rnn i sto on vraca.

1. mene je pital 7. Koja je formula za ažuriranje skrivenog stanja?, gdje bu došlo do eksplodirajučeg onog sranja i kak bujemo to rješili ŠČim
2. [16:47]

to smo pričali o RNN i LSTM

1. [16:47]

pa kak bu to LSTM rješil

Pitanja s blica:

1. Sto je najvece dimenzije ako je velicina vokabulara i ulazna dimenzija 100 a dim skrivenog sloja 1000:
   * Whh. Whx, Wxy ili bh?
2. Od kud do kud ide tangens hiperbolni?
3. Koje smo posebne znakove koristili u labosu?
4. Koju smo funkciju gubitka koristili?
5. Cemu sluzi gradient clipping?
6. Za sto koristimo <PAD>
7. Koja je formula za ažuriranje skrivenog stanja?

neka pitanja:

kakav bi bio loss u 2. zadatku da se promjeni redoslijed rijeci (ostao bi isti) broj parametara u rnn, nisam znao

koja matrica ima najvise elemenata izmedju w\_hh, w\_hy, b

1. Aktivacijska funkcija obicnog rnn-a? TanH
2. Zbog koje matrice tezina dolazi do eksplodirajuceg ili nestajuceg gradijenta? Wxh
3. Optimizacijski alg iz labosa?
4. Adam Gubitak iz labosa?
5. Unakrsna entropija
6. Moze li gubitak biti 0 u labosu? Mislim da ne jer sigmoida nikad nije 0 ili 1
7. Aha jos je trebalo zaokruziti izraz za azuriranje skrivenog stanja rnn-a
8. Derivacija tangensa hiperbolnog

Koja matrica od ovih matrica ima najviše parametara, ako je ulazna dimenzija 100 a Skriveni sloj veličine 1000? Whh, Whx, i još neke

1. Jednostavna povratna neuronska mreza, ulaz 100, Skriveni sloj 1000
2. Koliko ima parametara Skriveni i izlazni sloj
3. Formula za ažuriranje lstm celije

A screenshot of a phone

Description automatically generated