

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)

```
1. Koliko ima parametara BatchNormConvReLU, ako je kernel veličine 3,  $c_{in} = c_{out} = 32$  ?  
Pretpostavljamo da BatchNormConvReLU uključuje:  
• Conv2d( $c_{in}=32$ ,  $c_{out}=32$ ,  $kernel\_size=3$ ,  $padding=1$ )  
• BatchNorm2d(32)  
• ReLU()  
Računanje parametara:  
🔗 Conv2d:  
• Parametri:  $c_{out} * c_{in} * kernel\_size * kernel\_size * c_{out}$   
•  $= 32 * 32 * 3 * 3 + 32 = 9216 + 32 = 9248$   
🔗 BatchNorm2d:  
• Ima dvije grupe parametara po kanalu:  $\gamma$  (scale) i  $\beta$  (shift)  $\rightarrow 2 * 32 = 64$   
| (Također ima statistike (mean, var), ali one nisu trenirajući parametri.)  
✅ Ukupno:  
• Conv2d : 9248  
• BatchNorm2d : 64  
📊 Ukupno parametara: 9312
```

2. Kojeg reda je tenzor koji vraća funkcija get_features
(B, emb_size) gdje je B velicina batcha, npr moze bit (64, 32)
3. Koje su veličine emb_size za 2 modela koja smo implementirali. 32 i 784

(Model 1: emb_size = 32 | Model 2 (PCA ili bez obrade): emb_size = 784 ????)

4. Koja je povezanost Module i Sequential (sequential nasljeđuje module)
5. 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 funkcioniра 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 simpleembedding 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.



MEasy 02.06.2024., 21:14

je li odg neko mozda na pitanja za blic

PART1:

pitanja s blica:

- sto rade model.eval i model.train (ODG: to modelu daje do znanja u kojem je modu?)
 - koliko kanala ima ulaz u prvu konvolucijsku jedinicu - 1 je odg
 - 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)
 - koja je povratna vrijednost *_samplepositive funkcije - indeks random izabrane pozitivne slike*
 - *koja je povratna vrijednost one torch.pca.. funkcije (numpy array, tensor, string ili tuple? - Tuple[Tensor, Tensor, Tensor])*
 - koja metoda se koristi za klasifikaciju (logisticka regresija, linearna regresija, najblizi susjedi...) - najblizi susjedi je odg
-
- Kojih je dimenzija tensor koji vraca mean_representation - (10, 32)

PART 2:

odnos razreda nn.module i nn.sequential - nn.module je tata od nn.sequential

get_features u identitymodelu koliki je tensor - 1 x 784 ?

sto sve sprema torch.save() - parametre modela ??

kako se određuje positive i negative za klasu c (tipa positive==c^i; negative != c^i)

kojeg je reda tensor kojeg vraca get_features - 2d

koji tip klasifikacije

dimezija compute_representations return tenzora

koliko kanala ima ulaz u prvi sloj konvolucije

koji je posljednji sloj u simpleMetricEmbedding modelu (edited)

ja san napisao neke odg ali nisan sig jel točno

odnos razreda nn.module i nn.sequential (**sequential nasljeđuje module**)

get_features u identitymodelu koliki je tensor (1 x 784) (mozda je točan 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 (težine 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 tensor 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_representations 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 slike)

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)
- između koliko pozitivnih i negativnih primjera možemo 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: $(U, S, V) = \text{torch.pca_lowrank}(X, q)$ **U:** glavni vektori — **projekcija podataka** na q dimenzija → veličine **(n, q)**, **S:** singularne vrijednosti → veličine **(q,)** **V:** glavne komponente → veličine **(m, q)**)

- koja metoda se koristi za klasifikaciju (logistička regresija, linearna regresija, **najbliži susjedi...**)
- Kojih je dimenzija tensor koji vraća `mean_representation` (**broj_klasa, emb_size**) znaci vj je dimenzija (10,32)



Goliathus 29.05.2023., 17:54

Proučite funkcije za učenje i vrednovanje u modelu `utils.py`. Kako se računaju reprezentacije razreda?

U `compute_representations` (dali su nam ovaj u `utils` datoteci), računamo reprezentacije (embeddings) za podatke. Za svaki razred pronalazimo prosjek reprezentacija, i svaki taj prosjek reprezentacija razreda predstavlja razred, odnosno središte reprezentacije razreda.

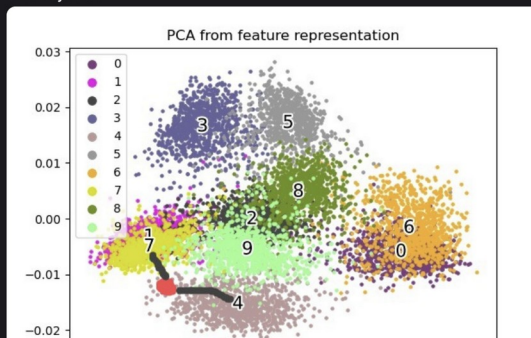
Kako se provodi klasifikacija primjera?

Na novom primjeru računamo njegovu reprezentaciju (u funkciji `evaluate`) i računamo euklidsku L2 udaljenost od središta reprezentacija svakog razreda. Novi primjer se klasificira u onaj razred kojem je najbliži po euklidskoj udaljenosti.

Probajte smisliti alternativne pristupe za klasifikaciju primjera.

Naša implementacija klasifikacije u obzir uzima samo središte svih reprezentacija za svaku klasu. Alternativni način bi bio da se koristi Mahalanobisova udaljenost. Mi smo na našim grafovima vizualizirali samo 2 dimenzija, u stvarnosti reprezentacije imaju 32 dimenzija. Mogli bismo pronaći distribucije reprezentacija svakog razreda pomoću modela Gaussovih mješavina (znači procjenimo distribucije iz kojih dolaze podaci) i tako imamo više informacija iz koje distribucije podatak dolazi. Naš model gleda samo kojem srednjoj reprezentaciji razreda je naša reprezentacija jednog primjera najbliža, dok bi ovakav model bolje mogao iskoristiti informacije o međusobnom položaju distribucija reprezentacija razreda.

Tu je novi crveni primjer bliže klasi 7 i po dobivenoj implementaciji bi se stavio u razred 7, jer je najbliže tom središtu. Ako bismo uzeli u obzir mahalanobisovu udaljenost, vidimo da nas "više iznenađuje" to da stavimo u razred 7 nego u 4, jer je distribucija klase 4 izdužena prema toj crvenoj točki



lab4

Koja je povezanost `Module` i `Sequential`

Ako imamo K klasa i svaka klasa ima M primjera, koliko možemo generirati pozitivnih i negativnih primjera za jedan anchor primjer.

odnos razreda `nn.module` i `nn.sequential`

`get_features` u `identitymodelu` koliki je tenzor

sto sve prema `torch.save()`

kako se određuje `positive` i `negative` za klasu c (tipa `positive==c^i`; `negative != c^i`)

kojeg je reda tenzor kojeg vraća `get_features`

koji tip klasifikacije

dimezija `compute_representations` return tenzora

koliko kanala ima ulaz u prvi sloj konvolucije

koji je posljednji sloj u `simpleMetricEmbedding` modelu





majano3 03.06.2024., 22:32

Blic 4 labos pitanja koje se sjećam

Dimenzionalnost vektora ugrađivanja Identity modela?

- a. 32
- b. 1
- c. **784**
- d. neki broj

model.eval() i model.train()?

- a. **modificiraju normalizacijski sloj**
- b. modificiraju aktivacijski sloj
- c. za treniranje i evaluaciju podataka
- d. sprečavaju racunanje gradijenata

koji tip vraca .sample_positive?

- a. string
- b. list
- c. **int**
- d. neke

Koliko ima parametara batchnormconvrelu ako je kernel veličine 3, c_in=c_out=32?

9312

kak smo kod vizualizacije dosli do dimenzije 2?

- a. potpuno povezani sloj s 2 neurona
- b konvolucijskim slojem
- c. **analizom glavnih komponenti**
- d. sloj sazimanja

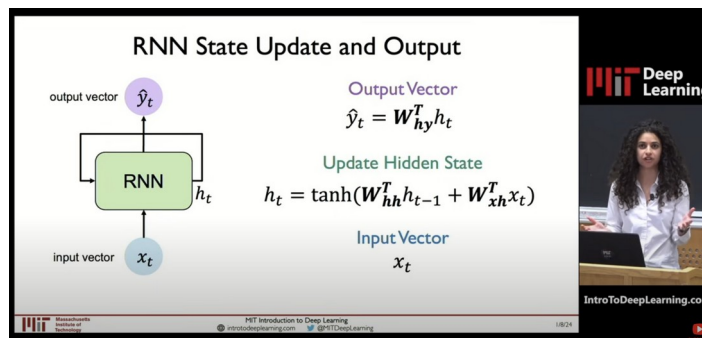
formula za triplet loss?

- a. **$\max(0, d(a,p) - d(a,n) + \text{margin})$**
- b. $\max(0, d(a,p) - d(p,n) + \text{margin})$
- c. $\max(...)$
- d. $\min(...)$

ostala su se ponovila od prošlih godina



LABOS 3 PITANJA



Pitanje za blic koje nisam prije vidjela: koja je vrijednost gradijenta funkcije gubitka, ulaz softmax, 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?

$h[t] = \tanh(W_{hh} * h[t-1] + W_{xh} * x[t] + b_h)$

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 dimenzija 1000. Koja matrica ima najviše parametara?

*a) Whh

b) Why

c) Wxh

d) by

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

ja sam ovo dobila: 1201100 ($W_{hh} + W_{xh} + b_h + W_{hy} + b_y = 1000*1000 + 100*1000 + 1000 + 1000*100 + 100$)

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 $\tanh(W_{hh} \cdot h[t-1] + W_{xh} \cdot x[t])$**)

koja je derivacija tangensa hiperbolnog (**$1 - \tanh^2()$**)

Formula za LSTM: $c(t)=...$

LSTM koristi više "vrata". Najvažnije formule su:

$$\begin{aligned} f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) && \text{(forget gate)} \\ i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) && \text{(input gate)} \\ \tilde{c}_t &= \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) && \text{(candidate state)} \\ c_t &= f_t \odot c_{t-1} + i_t \odot \tilde{c}_t && \text{(cell state update)} \\ o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) && \text{(output gate)} \\ h_t &= o_t \odot \tanh(c_t) && \text{(new hidden state)} \end{aligned}$$

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

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

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

6) Cemu služi 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).)

7) Za što koristimo <PAD> <PAD> (padding token) koristimo za **poravnanje (engl. padding)** nizova različite duljine u **jednake duljine**, kako bi se mogli **obrađivati u batchu**.

8)

Pitala me me po čemu je model GRU različit od običnog 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 između baseline modela i rnn-a. (**baseline je Feedforward, ignorira redoslijed 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 mogući 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 eksplozivajućeg onog sranja i kak bujemo to riješili ŠČim
2. [16:47]
to smo pričali o RNN i LSTM
3. [16:47]
pa kak bu to LSTM riješil

Pitanja s blica:

1. Sto je najveće dimenzije ako je velicina vokabulara i ulazna dimenzija 100 a dim skrivenog sloja 1000:
 - o W_{hh} , W_{hx} , W_{xy} ili b_h ?
2. Od kud do kud ide tangens hiperbolni?
3. Koje smo posebne znakove koristili u labosu?
4. Koju smo funkciju gubitka koristili?
5. Cemu služi 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 najviše elemenata između w_{hh} , w_{hy} , b

1. Aktivacijska funkcija običnog rnn-a? TanH
2. Zbog koje matrice težina dolazi do eksplodirajućeg ili nestajućeg gradijenta? W_{xh}
3. Optimizacijski alg iz labosa?
4. Adam Gubitak iz labosa?
5. Unakrsna entropija
6. Može li gubitak biti 0 u labosu? Mislim da ne jer sigmoida nikad nije 0 ili 1
7. Aha još je trebalo zaokružiti izraz za ažuriranje 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? W_{hh} , W_{hx} , i još neke
9. Jednostavna povratna neuronska mreža, ulaz 100, Skriveni sloj 1000
10. Koliko ima parametara Skriveni i izlazni sloj
11. Formula za ažuriranje lstm ćelije

PART 2:

1. Sto je najveće dimenzije ako je veličina vokabulara i ulazna dimenzija 100 a dim skrivenog sloja 1000: W_{hh} , W_{hx} , W_{xy} ili b_h ?
2. Od kud do kud ide tangens hiperbolni?
3. Koje smo posebne znakove koristili u labosu?
4. Koju smo funkciju gubitka koristili?
5. Čemu služi gradient clipping?
6. Za što koristimo <PAD>
7. Koja je formula za ažuriranje skrivenog stanja? (edited)