

Duboko Učenje 1- Pokolj na zi 26.6.2025

1)

Vještica Morgana bacila je tri kletve na svakog viteza Okruglog stola. Informacije o vitezima spremljene su u matricu X dimenzija $(N, 5)$, gdje svaki redak predstavlja jednog viteza:

x_1, x_2, x_3 – količine svake od tri različite kletve

x_4, x_5 – x i y koordinate viteza u ravnini (udaljenost od središta koordinatnog sustava)

Nakon što su kletve bačene, dolazi do prijenosa kletvi među vitezima. Novi intenziteti kletvi za svakog viteza izračunavaju se pomoću težinskog prosjeka svih vitezova, uključujući njega samoga. Nove težine ovise o kutnoj udaljenosti između vitezova – manji kut znači jači utjecaj (veća težina).

Napisati funkciju `curse_transmission(X)`, koristeći self attention s jednom glavom koristeći numpy. izracunati nove težine. napisati broj parametara.

2)

Koristeći slijed-u-slijed RNN, napisati dekodera. Dekoder već vraća vektor veličine 10. $x_0 = 1$ ako je mačka, $x_1 = 1$ ako je pas, $x_2 = 1$ ako je mali, $x_3 = 1$ ako je veliki. Ostale vrijednosti su nula. Vektorska reprezentacija služi kao h_0 , početno skriveno stanje dekodera koji generira opis slike. Svaki slijed započinje s `<START>` i završava s `<END>`. Model uvijek generira opis u obliku: an image of a (small/big) (cat/dog).

Rječnik (stoi) je definiran kao: cat:0, dog:1, small:2, big:3, image:4, an:5, of:6, a:7, `<START>`:8, `<END>`:9. Aktivacijska funkcija je identitet, a izlaz se dobiva primjenom softmax funkcije na trenutno skriveno stanje h . Napisati sve parametre modela i ukupan broj parametara.

3)

Dana je matrica E dimenzija $(6, 3)$ Prva dva elementa svakog retka su ugradbeni vektor, a treći element je klasa.

Koristimo kvadratni kontrastni gubitak sa $m=3$

$E = [$

[0, 0, 1],

[1, 0, 1],

[4, 0, 2],

[3, 1, 2],

[2, 0, 2],

[1, 1, 3]

$]$

Promatramo parove $(x_1, x_2), (x_1, x_3), (x_1, x_4), (x_1, x_5), (x_1, x_6)$.

a) Definiraj kvadratni kontrastni gubitak za e_i, e_j .

b) Izračunaj gubitak za parove. Koji neće doprinjeti učenju? zasto?

- c) Koji je najteži negativ za x_1 . Kakav mu je utjecaj na učenje. Što ako ga maknemo
d) $z=2,0$ usporedba s centroidima

4)

Dana je funkcija diskriminativnog modela f , skup podataka $\{x_i\}$ i pripadajuće kategoričke oznake $\{y_i\}$. Definirati funkciju izglednosti modela te objasniti njenu vezu sa standardnim gubitkom u nadziranom učenju. Zatim, za slučaj kada ciljana slučajna varijabla ima Bernoullijevu odnosno višerazrednu kategoričku razdiobu, napisati izraze za gradijent standardnog log-gubitka po predaktivacijama posljednjeg sloja modela.

5)

Nacrtajte tipični postav za samonadzirano učenje s 2 grane. Napišite jednadžbe gubitka za slučaj kad optimiramo:

- a) kontrast podatkovnih ugrađivanja
- b) distribuciju pripadnosti naučenim prototipima

6)

Napišite program u PyTorchu za provlačenje kvadratnog polinoma kroz zadani skup 2D točaka koristeći gradijentni spust. Procijenite memorijski otisak računanja gradijenta širenjem unatrag (podaci, parametri itd.) pod pretpostavkom da računamo u dvostrukoj preciznosti FP64 i učimo s veličinom grupe 100.

```
x = torch.linspace(-10.0, 10.0, steps=100, dtype=torch.float64)
y = x**2 + torch.randn(100, dtype=torch.float64)
```

PITALICE

1. Kako se tretiraju parametri izlazne projekcije Why kroz vremenske korake RNN za guste predikcije?
 - a) parametri se mijenjaju linearno s vremenskim korakom t
 - b) svaki korak ima vlastiti nezavisni Why
 - c) isti parametri se dijele preko svih vremenskih koraka
 - d) inicijaliziraju se nasumično na početku svake epohe
2. Koja je glavna prednost metričkih ugrađivanja naspram klasifikacijskih modela?
 - a) manje parametara
 - b) ne zahtijevaju normalizaciju ulaza
 - c) uvijek bolja točnost
 - d) mogu se koristiti za nove razrede koje nismo vidjeli tokom učenja
3. Pod kojim uvjetom rangiranje podataka temeljeno na skalarnom produktu daje isti poredak kao rangiranje temeljeno na euklidskoj udaljenosti?
 - a) kad su vektori normirani na jediničnu duljinu
 - b) uvijek
 - c) kad se koristi Mahalanobisova metrika
 - d) ako su podaci binarni vektor

4. Imamo 1 konvolucijski sloj (jezgra 3x3, korak 1, bez paddinga) i 1 sloj sažimanja (jezgra 2x2, korak 2).
Koliko je receptivno polje jednog neurona izlaza sloja sažimanja?
A: 4x4
5. Koji je primarni razlog korištenja mehanizma pozornosti kod strojnog prevođenja za duge ulazne slijedove?
A: mogućnost selekcije važnih reprezentacija ulaza kada je potrebno
6. Kako se obično obrađuju skrivena stanja kod dvosmjernih običnih povratnih modela?
A: konkatenuiraju se
7. Koji su ulazi povratne ćelije višeslojnog povratnog modela za koju vrijedi $n > 1$?
A: $h_n(t-1)$, $h_{n-1}(t)$
8. Kod BPTT (eng. backpropagation through time), kako se računa gradijent parametra W_{hh} (ili neki drugi, nemam pojma)?
A: sumiraju se gradijenti tog parametra po svim vremenskim koracima
9. Koji je kontrastni gubitak ekvivalentan standardnoj unakrsnoj entropiji?
A: gubitak N parova
10. Kolika je površina ispod krivulje preciznosti i odziva za binarnu klasifikaciju dvaju podataka ako su oznake $Y=[0,1]$, a predikcije $P(Y=1|x)=[0.9,0.8]$?
A: 0.5
11. Vaswanijev transformer, čega nema u enkoderu (cross attention, rezidual, potpuno povezani sloj, normalizacijski sloj (??))
A: cross attention