

LJETNI ISPITNI ROK

DUBOKO UČENJE 1

04.07.2024.

TEORIJA

1. Prednost AdaGrad nad osnovnim SGD:
 - a. skaliranje komponenata gradijenata s akumulacijskom normom
 - b. skaliranje gradijenata eksponencijalno pomičnim prosjekom norme
 - c. dodavanje zaleta u procjenu trenutnog gradijenta
 - d. linearno skaliranje stope učenja
2. Korištenje mini grupe umjesto čitavog skupa uzoraka opravdano je jer:
 - a. s povećanjem uzoraka preciznost gradijenata raste ispodlinearno
 - b. preciznost ne ovisi o broju uzoraka mini grupe
 - c. preciznost određivanje gradijenata nema utjecaj na algoritam
 - d. preciznost raste kvadratno s brojem uzoraka
3. Koji je gradijent gubitka po ulazu u softmax:
 $\text{Softmax}([\ln(2) \ 0 \ -\ln(2) \ -\ln(4) \ -\ln(4)])$
Podatak pripada razredu s indeksom 0.
 - a. $[-0.25, 0, 0.25, 0.5, 0.5]$
 - b. -0.55
 - c. Undefined
 - d. $[-0.5, 0.25, 0.125, 0.0625, 0.0625]$
4. Učenje modula na skupu uzoraka umjesto nad distribucijom koja generira podatke za posljedicu ima:
 - a. nema lokalnog optimuma kod učenja
 - b. brzo učenje, dobra generalizacija
 - c. ne trebamo Hessovu matricu
 - d. sklonost prenaučivosti
5. 2 ista konvolucijska sloja bez pomaka K1 i K2. K1 se primjenjuje na grayscale slike, a K2 na RGB slike. Koji je odnos parametara između slojeva K1 i K2.
 - a. K2 ima 3x više parametara
 - b. K2 ima 3x manje parametara
 - c. isti broj parametara

d. ista računska složenost pod uvjetom da su slike istih dimenzija

6. Formula za odziv:

a. $R = TP / (TP + FP)$

b. $R = TP / (TP + FN)$

c. $R = FP / (TP + FP)$

d. $R = FP / (TP + FN)$

7. Ako RNN ima n parametara, koliko parametara ima LSTM?

a. $2n$

b. $4n$

c. n

d. $3n$

8.

9. .

10..

11..

12..

ZADACI

1. .
2. Napišite jednađbe za SGD, SGD s momentom, RMS prop, ADAM. Za svaki od njih memorijske zahtjeve u ovisnosti o broju parametara n .
Raspišite prve dvije iteracije minimizacije $f(x) = x^2 - 4x + 2$ algoritmom ADAM počevši od $x = 0$. Hiperparametri su $\rho_1 = \rho_2 = 0.9$, $\delta = 0$ s korakom $\varepsilon = 0.1$.
3. Imamo model sa jednim potpuno povezanim slojem kojem formula glasi $\mathbf{h} = \mathbf{W}\mathbf{x} + \mathbf{b}$, gradijent $\frac{\partial L}{\partial \mathbf{h}}$ je poznat.
 - a) izraz za gradijent gubitka po ulazima sloja
 - b) kako bi izgledali gradijenti gubitka po vektorizirani matrici $\frac{\partial L}{\partial \text{vec}(\mathbf{W})}$
 - c) izraz za gradijent gubitka po j -tom retku matrice $\frac{\partial L}{\partial W_j}$. Pokažite kako računski i memorijski efikasno možemo izračunati gradijent gubitka po svim retcima \mathbf{W}
 - d)
$$\mathbf{x} = [2, 4]^T$$
$$\frac{\partial L}{\partial \mathbf{h}} = [a, b, a + 2b]$$
$$\frac{\partial L}{\partial \mathbf{W}} = \begin{bmatrix} 4 & ? \\ ? & 2 \\ ? & ? \end{bmatrix}$$

Odredite a i b te sve nepoznate vrijednosti u $\frac{\partial L}{\partial \mathbf{h}}$
4. Metrička ugrađivanja s dvodimenzionalnim podacima. Ugrađivanje izvodimo jednim potpuno povezanim slojem sa parametrima $\mathbf{W} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$.
Na ulazu modela imamo sidro $(0,0)$, pozitivni primjer $(1,1)$ i negativni primjer $(0,1)$. Model koristi trojni gubitak.
 - a) Odredite iznos trojnog gubitka za dani primjer i parametre.
 - b) Odredite gradijent gubitka po parametrima modela.
5. Jednostavna povratna arhitektura za strojno prevođenje.
 - a) Ilustrirajte razmotani dijagram unaprijednog prolaza jednostavne povratne arhitekture za strojno prevođenje ako je na ulazu niz od 4 riječi.

- b) Odredite ukupan broj parametara takve arhitekture ako je veličina ulaznog vokabulara 5, veličina izlaznog vokabulara 10, dimenzija skrivenog stanja povratnih modela 4 i ako ulazne riječi reprezentiramo gustim vektorima dimenzija 10.
- c) Proširite prikazanu arhitekturu modulom pažnje pod pretpostavkom da pažnja razlikuje reprezentacije ključeva i vrijednosti te da sličnost modelira skalarnim produktom.
- Napišite jednadžbe pažnje za neki izlazni trenutak (t_{dec}) i odredite općeniti broj parametara takvog modula.