

1. Prednost algoritma AdaGrad nad osnovnom implementacijom SGD-a je:
 - (a) dodavanje momenta u procjenu trenutnog gradijenta
 - (b) skaliranje komponenata gradijenata njihovom akumuliranom normom
 - (c) skaliranje gradijenata eksponencijalnim pomičnim prosjekom
 - (d) ugradnja momenta u praćenje gradijenta te skaliranja vremenski lokalnom procjenom prosječne norme
2. Povećanjem veličine mini-grupe dobivamo:
 - (a) veću brzinu evaluacije modela na ispitnim podacima
 - (b) manji broj manje preciznih ažuriranja parametara modela
 - (c) veći broj preciznijih ažuriranja parametara modela
 - (d) manji broj preciznijih ažuriranja parametara modela
3. Zašto potpuno povezani modeli s prosječnom unakrsnom entropijom imaju više jednako dobrih lokalnih minimuma?
 - (a) zbog simetrije neurona istog skrivenog sloja
 - (b) zbog simetrije neurona izlaznog sloja
 - (c) zbog doprinosa regularizacije
 - (d) zbog simetrije neurona susjednih skrivenih slojeva
4. Kako biste opisali odnos između gubitka „0-1“ i unakrsne entropije?
 - (a) gubitak „0-1“ je metoda izbora kada imamo malo podataka
 - (b) gubitak „0-1“ je nadomjestak unakrsne entropije
 - (c) unakrsna entropija je derivabilni nadomjestak gubitka „0-1“
 - (d) unakrsna entropija je metoda izbora kad imamo jako puno podataka
5. Kod uporabe lokalne normalizacije odziva (LRN), normaliziraju se:
 - (a) izlazi neurona s obzirom na odzive kada mreža radi ispravno odnosno neispravno
 - (b) izlazi neurona s obzirom na odzive koje generira za različite uzorke za učenje
 - (c) izlazi neurona s obzirom na izlaze njemu susjednih neurona u istom sloju
 - (d) izlazi neurona s obzirom na izlaze njemu susjednih neurona u slojevima prije i nakon njega
6. Predtreniranje je postupak:
 - (a) koji se provodi kada je zadatak učenja neuronske mreže jednostavan
 - (b) bez kojeg paralelizacija pri obradi podataka tijekom učenja mreže ne bi bila moguća
 - (c) koji osigurava početne vrijednosti parametara mreže koje su bolje od slučajnih
 - (d) koji osigurava da su težine smještene u raspon između -6 i +6
7. Do problema eksplodirajućeg gradijenta može doći ako se podatci uzastopno množe matricom težina za koju vrijedi:
 - (a) da ima neke svojstvene vrijednosti koje su po iznosu oko 0
 - (b) da ima neke svojstvene vrijednosti koje su po iznosu dosta veće od 1
 - (c) da ima neke svojstvene vrijednosti koje su po iznosu dosta manje od 1
 - (d) da je singularna

8. Zašto pri optimizaciji radije promatramo log-izglednost nego izglednost?
- (a) log-izglednost podataka se brže evaluira od njegove izglednosti
 - (b) zbog boljih stat gradijenata nezavisnih podataka
 - (c) zbog aditivnosti gradijenata nezavisnih podataka
 - (d) zbog aditivnosti gradijenata međusobno ovisnih podataka
9. Uzorkovanje manjeg broja uzoraka (mini-grupe) umjesto uporabe čitavog skupa uzoraka za učenje pri izračunu gradijenata je opravdano jer:
- (a) preciznost određivanja gradijenta s povećanjem broja uzoraka raste ispodlinearno
 - (b) preciznost određivanja gradijenta ne ovisi o broju uzoraka mini-grupe
 - (c) preciznost određivanja gradijenta nema nikakvog utjecaja na rad algoritma strojnog učenja
 - (d) preciznost određivanja gradijenta raste kvadratno s brojem uzoraka
10. U kojoj od sljedećih situacija nam ne može pomoći učenje s momentom?
- (a) kod plitkih lokalnih minimuma
 - (b) kod „kanjona“
 - (c) kod „sedla“
 - (d) kod „ploha“
11. Kod normalizacije nad grupom normalizirani se podatci na kraju provode kroz afinu transformaciju:
- (a) kako bi se povećala ekspresivnost mreže u odnosu na onu bez normalizacije podataka
 - (b) kako bi se osiguralo da su podaci centrirani oko nule
 - (c) kako bi se osiguralo da mreža ne gubi ekspresivnost
 - (d) kako bi se smanjilo standardno odstupanje podataka
12. Tehnike regularizacije:
- (a) povećavaju kapacitet modela
 - (b) smanjuju pristranost modela
 - (c) povećavaju pristranost modela
 - (d) povećavaju kompleksnost modela
13. Koja je vremenska složenost računanja izlaza u svim vremenskim koracima u dvosmjernoj višeslojnoj povratnoj neuronskoj mreži (bidirectional RNN), gdje je dubina mreže D , a broj vremenskih korak odmatanja mreže T ?
- (a) $T + D - 1$
 - (b) $T \cdot D$
 - (c) T^2
 - (d) T

14. Koji je odnos između vektora $a = \text{softmax}(x)$ i vektora $b = \text{softmax}(x - \max(x))$
- (a) u implementaciji preferiramo a jer se računa brže
 - (b) matematički, vrijedi $\|a\| < \|b\|$
 - (c) matematički, vrijedi $a = b$
15. Rano zaustavljanje ima regularizacijski efekt jer:
- (a) osigurava brzo učenje
 - (b) ne koristimo skup za validaciju
 - (c) osigurava ograničenu normu vektora parametara modela
 - (d) osigurava rano pretraživanje prostora stanja
16. Razmatramo k -ti sloj dubokog modela koji provodi transformaciju $h_k = W_k \cdot h_{k-1} + b_k$. Zašto tijekom učenja treba pamtiti međurezultate h_k :
- (a) Radi regularizacije
 - (b) Detekcije numeričkih pogrešaka
 - (c) Računanje gradijenta po b_k
 - (d) Računanje gradijenta po W_k
17. Može li korak gradijentnog spusta povećati gubitak mini-grupe:
- (a) Da, ali samo ako je faktor pomaka prevelik
 - (b) Da, ali samo ako koristimo regularizaciju
 - (c) Da, ali samo ako je mini-grupa premala
 - (d) Da, ali samo ako je mini-grupa prevelika
18. Razmatramo slučajnu varijablu $Y = a \cdot x + b + W$, gdje x označava ulaz, a $W \sim N(0, 100)$. Algoritam strojnog učenja koji uči vezu između Y i X polinomom 10. stupnja ima:
- (a) Veliku varijancu, malu pristranost
 - (b) Nezamjetljivu varijancu, malu pristranost
 - (c) Ništa od navedenog nije točno
 - (d) Nezamjetljivu varijancu, veliku pristranost
19. Korištenje LSTM-a potpuno rješava:
- (a) Pamćenje dugih veza
 - (b) Broj parametara
 - (c) Problem eksplodirajućeg gradijenta
 - (d) Problem nestajućeg gradijenta
20. Zašto u konvolucijskoj mreži koristimo sažimanje:
- (a) Osiguravamo derivabilnost narednih konvolucijskih slojeva
 - (b) Isključivo da poboljšamo invarijantnost na pomake objekata
 - (c) Poboljšamo invarijantnost na pomake objekata i smanjimo dimenzionalnost izlaza
 - (d) Isključivo da smanjimo dimenzionalnost izlaza