

# Duboko učenje 1- Jesenski ispitni rok (06.09.2023.)

1. Koji izraz se koristi za ažuriranje skrivenog stanja povratne ćelije s dugoročnom memorijom (Long Short Term Memory)?

- a.  $h^{(t)} = i^{(t-1)} \circ h^{(t-1)} + (1 - i^{(t-1)}) \circ \hat{h}^{(t-1)}$
- b.  $h^{(t)} = \sigma(W_{hh} \cdot h^{(t-1)} + W_{xh} \cdot x^{(t)} + b)$
- c.  $c^{(t)} = f^{(t)} \circ c^{(t-1)} + (g^{(t)}) \circ \hat{c}^{(t)}$
- d.  $h^{(t)} = \tanh(i^{(t-1)} \circ h^{(t-1)} + (1 - i^{(t-1)}) \circ \hat{h}^{(t-1)})$

2. Koju tehniku ne ubrajamo u regularizaciju?

- a. učenje sa zaletom
- b. usrednjivanje predikcije većeg broja modela
- c. normalizacija po grupi podataka
- d. rano zaustavljanje optimizacije

3. Funkcija rotira točke ravnine za kut od 35 stupnjeva. Navedi dimenzije Jakobijeve matrice te funkcije.

- a. ta funkcija nema derivaciju
- b. 1x2
- c. 2x2
- d. 1x1

4. Zašto su konvolucijski slojevi ponekad prikladniji od potpuno povezanih slojeva?

- a. zbog regularizacijskog efekta usred dijeljenja parametara
- b. zbog povećanja kapaciteta
- c. zbog usklađenosti s aktivacijom ReLU
- d. zbog jednostavnije implementacije

5. Za koji od parametara običnog povratnog modela postoji mogućnost eksplodirajućeg odnosno nestajućeg gradijenta?

- a.  $W_{xh}$
- b.  $W_{hh}$
- c. b
- d.  $W_{hx}$

6. Kolika je površina ispod krivulje preciznosti i odziva za binarnu klasifikaciju dvaju podataka ako su oznake  $Y=[0,1]$ , a predikcije  $P(Y=1|x)=[0.9,0.8]$ .
- 0.5
  - 1.0
  - 0.8
  - 0.4
7. Koja inačica kontrasnog gubitka je povezana s unakrsnom entropijom?
- gubitak N parova
  - trojni gubitak
  - meki trojni gubitak
  - trojni gubitak sa skalarnim produktom
8. Razmatramo L2 regulariziranu funkciju gubitka dubokog modela tijekom provedbe jednog koraka stohastičkog gradijentnog spusta. negativni gradijent regularizacije pomiče model u smjeru:
- suprotnom od ishodišta prostora modela
  - okomitom na gradijent ne-regularizirane funkcije gubitka
  - ne utječe na pomak modela
  - ishodišta prostora modela
9. Razmatramo višerazrednu logističku regresiju s  $n$  značajki na ulazu. Ako prilikom učenja tog modela koristimo stohastičko izostavljanje značajki (dropout), evaluacijom tako naučenog modela možemo dobiti:
- geometrijsku sredinu predikcije  $O(2^n)$  modela
  - aritmetičku sredinu predikcije  $O(n^2)$  modela
  - aritmetičku sredinu predikcije  $O(2^n)$  modela
  - aritmetičku sredinu predikcije  $O(n)$  modela
10. Koliko iznosi  $\text{softmax}([\ln 2, 0])$ ?
- nije definirano jer je jedan d ulaza jednak 0
  - 0.55
  - $[2/3, 1/3]$
  - 100%

**11. Za aktivacijsku funkciju latentnog sloja gotovo nikad nećemo koristiti:**

- a.  $g(x|\alpha, \beta) = \beta + \alpha \cdot x$
- b.  $g(x|\alpha) = \max(0, x) + 0.001 \cdot \min(0, x)$
- c.  $g(x) = \max(-0.1, x)$
- d.  $g(x) = (1 + e^{-\alpha x})^{-1}$

**12. Razmatramo računanje gradijenta funkcije cilja uzorkovanjem manjeg broja podataka (mini-grupe) umjesto uporabe čitavog skupa uzoraka za učenje. Ta ideja je opravdana jer preciznost gradijenta:**

- a. raste ispodlinearno s povećanjem broja uzoraka
- b. raste kvadratno s povećanjem broja uzoraka
- c. nema nikakvog utjecaja na rad algoritma strojnog učenja
- d. ne ovisi o broju uzoraka mini-grupe