

ZI 21/22 – Teorija

1. Označite razred koji ima metodu `backward_parameters`
 - (a) `SoftmaxCrossEntropyWithLogits`
 - (b) `ReLU`
 - ☒ (c) `L2Regularizer`
 - (d) `MaxPooling`
2. Koju od ponuđenih tehnika ne ubrajamo u regularizaciju?
 - (a) generiranje umjetnih podataka
 - (b) rano zaustavljanje optimizacije
 - (c) normalizacija po grupi podataka
 - ☒ (d) učenje sa zaletom
3. Koji od sljedećih izraza odgovara unakrsnoj entropiji?
 - (a) `torch.log_softmax(h2)`
 - ☒ (b) `(Yoh * logprobs).sum()`
 - (c) `loss.backward()`
 - (d) `torch.relu(torch.matmul(hfc1, wfc2) + bfc2)`
4. Transformaciju kojom modeliramo slučajne varijable s binomnom razdiobom možemo izraziti funkcijom:
 - (a) $A \cdot s + b$
 - (b) $\sqrt{|s|}$
 - (c) $e^s/(1-e^s)$
 - ☒ (d) $e^s/(1+e^s)$
5. Koji je odnos razreda `nn.Sequential` i `nn.Module`?
 - ☒ (a) `nn.Sequential` nasljeđuje `nn.Module`
 - (b) svi objekti tipa `nn.Module` sadrže barem jedan objekt tipa `nn.Sequential`
 - (c) svaki objekt tipa `nn.Module` stvara objekte tipa `nn.Sequential`
 - (d) `nn.Module` nasljeđuje `nn.Sequential`
6. Uzorkovanje manjeg broja uzoraka (mini-grupe) umjesto uporabe čitavog skupa uzoraka za učenje pri izračunu gradijenata opravdano je jer:
 - (a) preciznost određivanja gradijenta nema nikakav utjecaj na rad algoritma strojnog učenja
 - (b) preciznost određivanja gradijenta raste kvadratno s brojem uzoraka
 - ☒ (c) Preciznost određivanja gradijenta s povećanjem broja uzoraka raste podlinearno
 - (d) preciznost određivanja gradijenta ne ovisi o broju uzoraka mini-grupe

7. Promotrimo funkciju koja opisuje stvaranje slike projiciranjem točaka scene na slikovnu ravninu. Navedite dimenzije Jakobijanske matrice:

- (a) 1×1
- ☒ (b) 2×3
- (c) 3×3
- (d) ta f-ja nema derivacije

8. Navedite izraz za ažuriranje memorije LSTM ćelije:

- (a) $C(t) = W_{hh} \cdot c(t-1)$
- (b) $c(t) = f(t) \odot c(t-1)$
- ☒ (c) $c(t) = f(t) \odot c(t-1) + i(t) \odot \hat{c}(t)$
- (d) $c(t) = f(t) + c(t-1) + i(t) + \hat{c}(t)$

9. Izraz za trojni gubitak jest:

- ☒ (a) $\max\{d(a,p)-d(a,n)+m,0\}$
- (b) $\min\{d(n,p)+d(p,n)+m,0\}$
- (c) $\min\{d(a,p)+d(a,n)-m,0\}$
- (d) $\max\{d(n,p)+d(p,n)+m,0\}$

10. Razmatramo višerazrednu logističku regresiju s n značajki na ulazu. Ako prilikom učenja tog modela koristimo stohastičko izostavljanje značajki (dropout), jednim unaprijednim prolazom kroz tako naučeni model možemo dobiti:

- (a) aritmetičku sredinu predikcije $O(2^n)$ modela
- (b) aritmetičku sredinu predikcije $O(n^2)$ modela
- ☒ (c) geometrijsku sredinu predikcije $O(n^2)$ modela
- (d) aritmetičku sredinu predikcije $O(n)$ modela

11. Što možemo reći o odnosu $a = f(x)$ i $b = f(x - \delta \frac{df(x)}{dx})$?

- (a) $b \geq a$, ako je δ dovoljno mali
- (b) $a \leq b$
- (c) $b \leq a$, ako je δ dovoljno velik
- ☒ (d) $b \leq a$, ako je δ dovoljno mali

12. Razmatramo L2-regulariziranu funkciju gubitka dubokog modela tijekom provedbe jednog koraka SGD-a. Negativni gradijent regularizacije pomiče model u smjeru:

- ☒ (a) ishodišta prostora modela
- (b) suprotnom od ishodišta modela
- (c) ne utječe na pomak modela
- (d) okomitom na gradijent ne-regularizirane funkcije gubitka