

Duboko učenje

završni ispit

1. (10 bodova) Razmatramo klasifikacijski konvolucijski model koji je prilagođen obradi sivih slika dimenzija 5×5 . Arhitektura modela je:

- konvolucijski sloj bez nadopunjavanja: dvije jezgre 3×3 bez pomaka, korak 1, aktivacija ReLU;
- globalno sažimanje maksimumom;
- softmaks.

Konvolucijske jezgre inicijalizirane su kako slijedi: $\mathbf{w}_0 = -\mathbf{w}_1 = 0.5 \cdot \begin{bmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{bmatrix}$. Točne

oznake podataka su $\mathbf{y}_0 = [1, 0]^\top$ i $\mathbf{y}_1 = [0, 1]^\top$.

Neka su na ulazu slike \mathbf{I}_0 i \mathbf{I}_1 čiji su pikseli $\mathbf{I}_0[2,1]$, $\mathbf{I}_0[2,2]$, $\mathbf{I}_0[2,3]$ te $\mathbf{I}_1[1,2]$, $\mathbf{I}_1[2,2]$, $\mathbf{I}_1[3,2]$ postavljeni na 1, dok su svi ostali pikseli postavljeni na 0 (indeksi kreću od nule).

Zadatci:

- (a) Odredite dimenzije svih aktivacijskih tenzora te broj parametara modela.
- (b) Odredite gradijente svih parametara u podacima \mathbf{I}_0 i \mathbf{I}_1 .

(10 bodova)

2. Napišite jednadžbe optimizacijskih algoritama SGD, SGD s momentom, RMS prop i ADAM. Navedite memorijske zahtjeve za svaki od tih algoritama u ovisnosti o broju parametara n . Raspišite prve dvije iteracije minimizacije polinoma $f(x) = x^2 - 4x + 2$ algoritmom ADAM počevši od $x=0$. Poznati su sljedeći hiper-parametri postupka: $\rho_1 = \rho_2 = 0.9$, $\delta = 0$, te korak $\varepsilon = 0.1$.

(10 bodova)

3. Razmatramo učenje višerazredne logističke regresije $\hat{\mathbf{y}} = \text{softmax}(\mathbf{W}\mathbf{x} + \mathbf{b})$ optimiranjem unakrsne entropije stohastičkim gradijentnim spustom uz zadanu stopu učenja $\delta = 1$. Zadan je sljedeći skup za učenje: $\mathbf{x}_0 = (2, 2)$, $\mathbf{y}_0 = (1, 0)$, $\mathbf{x}_1 = (-1, -1)$, $\mathbf{y}_1 = (0, 1)$. Početno stanje matrice težina i vektora pomaka zadano je s $\mathbf{W}^0 = \begin{bmatrix} 0.5 & -0.5 \\ 0.5 & -0.5 \end{bmatrix}$, $\mathbf{b}^0 = [0 \ 0]^\top$.

- (a) Odredite jednadžbe gradijenata svih parametara.
- (b) Provedite jednu epohu učenja počevši od \mathbf{W}^0 i \mathbf{b}^0 , uz regularizaciju ispuštanjem (engl. dropout) ulaznih značajki s vjerojatnošću $p(\mu_{x_i}) = 0.5, i = 0, 1$. Pretpostavite sljedeći raspored ispuštanja:
 - prva mini-grupa je $(\mathbf{x}_0, \mathbf{y}_0)$, ugašena je dimenzija 0,
 - druga mini-grupa je $(\mathbf{x}_1, \mathbf{y}_1)$, ugašena je dimenzija 1.

Izračunajte gradijente i nove vrijednosti parametara.

- (c) Nakon koraka učenja u prethodnom podzadatku, izračunajte izlaz naučenog modela za ulaz $\mathbf{x}_2 = (1, 0)$.

(10 bodova)

4. Razmatramo običnu povratnu neuronsku mrežu s prijenosnom funkcijom tangens hiperbolni na povratnoj vezi te funkcijom softmax na izlazu.

Rješavamo problem jezičnog modela na razini slova iz laboratorijske vježbe, pri čemu nam je veličina vokabulara 10. Uz pretpostavku standardne preciznosti decimalnih brojeva od 4 bajta, odredite maksimalnu veličinu skrivenog sloja kako bi parametri modela stali u 624 bajta. Pri izračunu zanemarite vektore pristranosti b i c .

(4 bodova)

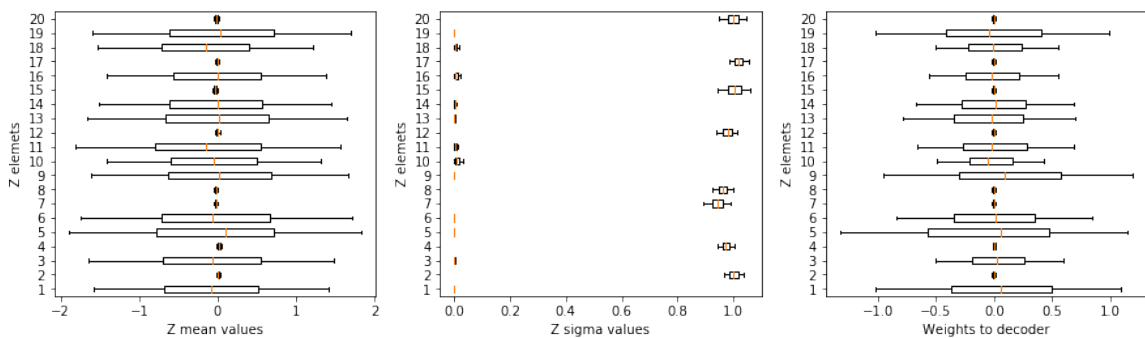
5. Razmatramo običnu povratnu neuronsku mrežu s identitetom u povratnoj vezi i bez izlaznog sloja, u kojoj su svi parametri skalari.

Ukoliko je poznato da je početno stanje $h^{(0)} = 64$, a ulazi u vremenskim koracima $x^{(1)} = z$, $x^{(2)} = z^2$ i $x^{(3)} = z^3$, odredite preostale parametre modela kako bi u skrivenom stanju mreže nakon trećeg koraka bila vrijednost polinoma $z + 2z^2 + 4z^3 + 1$.

(6 bodova)

6. Razmatramo varijacijski autoenkoder (VAE).

- Opišite regularizacijsku komponentu funkcije cilja VAE (što je, na što utječe...).
- Opišite reparametrizacijski trik.
- Kako se zove efekt koji je jasno uočljiv na slici te čega je on posljedica? Na slici su vizualizirane distribucije od $\mu_z(\mathbf{x})$ i $\sigma_z(\mathbf{x})$ za 20 elemenata skrivenog sloja, za sve uzorke \mathbf{x} iz skupa za treniranje te od svih težina koje povezuju pojedini z sa dekoderom.



(10 bodova)