DubuceZi 2023

1. Razmatramo afinu transformaciju zadanu sljedećom jednadžbom: q = W · p + b. Zadatci:

   (a) Napišite parcijalne derivacije izlaza po ulazu te izlaza po svim parametrima.

   (b) Za zadanu transformaciju implementirajte sučelje Layer po uzoru na drugu laboratorijsku vježbu. Implementirani sloj treba moći raditi s grupama podataka različitih dimenzija.

$$a) \quad q = Wp + b$$
$$\text{izlaz po ulazu:} \frac{\partial q}{\partial p} = W$$
$$\text{po parametrima: } W \text{ i } b$$
$$\frac{\partial q}{\partial W} = p$$
$$\frac{\partial q}{\partial b} = 1$$

b)

```
class Layer(nn.Module):
    def __init__(self, W, b):
        self.W = W
        self.b = b

    def forward(self, p):
        self.p = p
        return p @ self.W.T + self.b

    def backward_params(self, grads):
        dq_dW = grads.T @ self.p
        dq_db = np.sum(grads, axis=0)
        return dq_dW, dq_db

    def backward_inputs(self, grads):
        return grads @ self.W
```

2. Razmatramo klasifikacijski model s arhitekturom:

- konvolucijski sloj bez nadopunjavanja: 16 jezgara 7 × 7, korak 2, aktivacija ReLU;
- sažimanje maksimumom 2 × 2, korak 2;
- konvolucijski sloj bez nadopunjavanja: 32 jezgre 3 × 3, korak 1, aktivacija ReLU;
- sažimanje maksimumom 2 × 2, korak 2;
- pretvaranje u vektor;
- potpuno povezani sloj dimenzije 10 te aktivacija softmaks;

Svi slojevi imaju pomak (eng. bias). Na ulaz mreže dovodimo RGB slike veličine 62 × 62 piksela.

(a) Napišite dimenzije ulaznih i izlaznih tenzora za svaki sloj.

(b) Koliko ukupno parametara ima zadani model?

(c) Kolika su receptivna polja izlaza prvog i drugog sloja sažimanja? Pomoć: receptivno polje računamo s obzirom na piksele ulazne slike.

(d) Je li ovaj model prikladan za proizvoljnu prostornu dimenziju ulaza? Ako jest, dokažite to tako da provedete unaprijedni prolaz za proizvoljne dimenzije ulaza (nije potrebno računati stvarne mape značajki već dokaz ostvariti preko prikaza prostornih i semantičkih dimenzija mapa značajki). Ako nije, predložite minimalne promjene u arhitekturi tako da bude.

Formule:

$$n_{\text{out}} = \left\lfloor \frac{n_{\text{in}} + 2p - k}{s} \right\rfloor + 1$$

$$J_{\text{out}} = J_{\text{in}} \times s$$

$$r_{\text{out}} = r_{\text{in}} + (k - 1) \times j_{\text{in}}$$

sloj	param	tenzor	rf	Jin
Slika	$-$	$3 \times 62 \times 62$	1×1	1
Conv 7x7	$16 \times 3 \times 7 \times 7 + 16$	$16 \times 28 \times 28$	7×7	1
ReLU	$-$	$16 \times 28 \times 28$	7×7	2
Pool 2	$-$	$16 \times 14 \times 14$	9×9	2
Conv 3x3	$16 \times 32 \times 3 \times 3 + 32$	$32 \times 12 \times 12$	17×17	4
ReLU	$-$	$32 \times 12 \times 12$	17×17	4
Pool 2	$-$	$32 \times 6 \times 6$	21×21	4
Flatten	$-$	1152	21×21	8
FC1	$1152 \times 10 + 10$	10	62×62	8
Softmax	$-$	10	62×62	8

b) Parametri: 18538

c) 9×9 i 21×21

d) Nije moguce, treba koristiti adaptive pooling ili global polling prije potpuno povezanog sloja

3. Razmatramo običnu povratnu neuronsku mrežu koja klasificira nizove dvobitnih binarnih brojeva prema tome nalazi li se u nizu više parnih ili neparnih brojeva. Razredu 0 pripadaju nizovi koji imaju više (ili jednako) parnih nego neparnih brojeva, a razredu 1 nizovi koji imaju više neparnih brojeva. Primjerice niz 00, 01, 10, 11,00 sadrži tri parna i dva neparna broja, te zbog toga pripada razredu 0.

   (a) Odredite sve parametre i prikladne aktivacijske funkcije obične povratne neuronske ćelije i odgovarajućeg izlaznog sloja tako da model ispravno klasificira opisane nizove. Ulaz u mrežu $x^{(t)}$ neka bude binarni vektor-stupac s dva elementa.

   (b) Napišite jednadžbu ažuriranja stanja Vaše povratne ćelije, kao i jednadžbu izlaznog sloja i kriterij za određivanje razreda.

   (c) Koliko ukupno parametara ima Vaš model?

   (d) Demonstrirajte ispravnost Vašeg modela na sljedećem testnom nizu: 00, 10, 11.

$$0 \Rightarrow \text{više ili jednako parnih}$$
$$1 \Rightarrow \text{više neparnih}$$

$$\text{formule:}$$

$$h^t = W_{xh}x^{(t)} + W_{hh}h^{(t-1)} + b_h$$

$$o^t = \sigma(W_{hy}h + b_o)$$

$$\text{parametri:}$$

$$W_{xh} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}, \quad W_{hh} = \begin{bmatrix} 1 \end{bmatrix}, \quad b_h = -1$$
$$W_{hy} = 1, \quad b_0 = 0$$

$$\text{kriterij:}$$
$$O^{(t)} > 0.5 \Rightarrow 1$$
$$O^{(t)} \leq 0.5 \Rightarrow 0$$

c)

$$\text{broj parametara:}$$
$$W_{x_h} = 2, \quad W_{h_h} = 1, \quad b = 1$$
$$W_{hy} = 1, \quad b_o = 1$$
$$2 + 1 + 1 + 1 + 1 = 6 \text{ parametara}$$

d)

$$h^0 = \begin{bmatrix} 0 \end{bmatrix}$$
$$h^1 = \begin{bmatrix} 0 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \end{bmatrix} \cdot 0 - 1 = -1$$
$$h^1 = -1$$
$$h^2 = \begin{bmatrix} 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 1 \cdot (-1) - 1 = -2$$
$$h^2 = -2$$
$$h^3 = \begin{bmatrix} 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 1 \cdot (-2) - 1 = -1$$
$$h^3 = -1$$
$$O^t = \sigma(1 \cdot (-1) + 0)$$
$$O^t = 0.26894 \leq 0.5$$
$$O^t = 0 \Rightarrow \text{više parnih}$$

4. Razmatramo sijamsko učenje ugrađivanja u metrički latentni prostor primjenom kvadratnog kontrastnog gubitka. Ugrađivanje provodimo jednim konvolucijskim slojem s korakom jedan bez nadopunjavanja. Trenutni parametri konvolucijskog sloja su:

   $w = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}, \quad b = -0.1$

   Na ulazu su dva pozitivna podatka:

   $x_{p1} = \begin{bmatrix} 1.0 & 1.0 & 1.3 & 1.0 & 1.0 & 1.3 \end{bmatrix}, \quad x_{p2} = \begin{bmatrix} 3.3 & 3.3 & 2.7 & 3.3 & 3.3 & 2.7 \end{bmatrix}$

   Zadatci:

   (a) Napišite jednadžbe modela i gubitka za pozitivne i negativne parove.

   (b) Provedite unaprijedni prolaz i izračunajte gubitak.

   (c) Odredite gradijente s obzirom na parametre modela.

$$L_{pos} = \|f_\theta(x_q) - f_\theta(x_p)\|^2$$

$$\frac{\partial L_{pos}}{\partial f_p} = 2 \cdot (f_p - f_q)$$

$$L_{neg} = \max(0, m - \|f_\theta(x_q) - f_\theta(x_n)\|^2)$$
$$\frac{\partial L_{neg}}{\partial f_n} = 2\max(0, m - \|f_q - f_n\|) \cdot \frac{f_q - f_n}{\|f_q - f_n\|}$$

$$\text{model: conv1D (kernel\_size} = 3, \text{stride} = 1)$$
$$W = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \quad b = -0.1$$
$$x_{p1} = \begin{bmatrix} 1 & 1 & 1.3 & 1 & 1 & 1.3 \end{bmatrix}$$
$$x_{p2} = \begin{bmatrix} 2.3 & 3.3 & 2.7 & 3.3 & 3.3 & 2.7 \end{bmatrix}$$

b)

$$f_\theta(x_{p1}) = x_{p1} * W + b = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}$$
$$f_\theta(x_{p2}) = x_{p2} * W + b = \begin{bmatrix} 3 & 3 & 3 & 3 \end{bmatrix}$$
$$L_{pos} = \|f_\theta(x_{p1}) - f_\theta(x_{p2})\|^2$$
$$= 16$$

c)

$$\frac{\partial L_{pos}}{\partial f_{p1}} = 2\left(f_{p1} - f_{p2}\right) = \begin{bmatrix}-4 & -4 & -4 & -4\end{bmatrix}$$

$$\frac{\partial L_{pos}}{\partial f_{p2}} = 2\left(f_{p2} - f_{p1}\right) = \begin{bmatrix}4 & 4 & 4 & 4\end{bmatrix}$$

$$\frac{\partial L_{pos}}{\partial W} = \frac{\partial L_{pos}}{\partial f_{p1}} \frac{\partial fp_1}{\partial W} + \frac{\partial L_{pos}}{\partial fp_2} \frac{\partial fp_2}{\partial W}$$

$$\begin{aligned}\frac{\partial L_{pos}}{\partial fp_1} \cdot \frac{\partial fp_1}{\partial W} &= x_{p1} \star \begin{bmatrix}-4 & -4 & -4 & -4\end{bmatrix} \\ &= \begin{bmatrix}-17.2 & -17.2 & -18.4\end{bmatrix}\end{aligned}$$

$$\begin{aligned}\frac{\partial L_{pos}}{\partial fp_2} \cdot \frac{\partial fp_2}{\partial W} &= xp_2 \star \begin{bmatrix}4 & 4 & 4 & 4\end{bmatrix} \\ &= \begin{bmatrix}50.4 & 50.4 & 48\end{bmatrix}\end{aligned}$$

$$\frac{\partial L_{pos}}{\partial W} = \begin{bmatrix}33.2 & 33.2 & 29.6\end{bmatrix}$$

$$\frac{\partial L_{pos}}{\partial b} = \frac{\partial L_{pos}}{\partial f_{p1}} \cdot \begin{bmatrix}1\\1\\1\\1\end{bmatrix} + \frac{\partial L_{pos}}{\partial f_{p2}} \cdot \begin{bmatrix}1\\1\\1\\1\end{bmatrix} = [0]$$

5. Razmatramo utjecaj regularizacije normom vektora parametara modela na minimum funkcije gubitka. Na slici 1 prikazane su izokonture neregularizirane funkcije gubitka J, te točka njenog minimuma w^*. Zadaci:

   (a) Na odvojenim grafovima ucrtajte izokonture odgovarajuće norme parametara modela u slučajevima kada koristimo L_1 i L_2 regularizaciju.

   (b) Na istim grafovima ucrtajte točke w_1, w_2, w_3, w_4 koje odgovaraju minimumu regularizirane funkcije gubitka za sljedeće parametre regularizacijskog faktoraλ : $\lambda_1 = 0 < \lambda_2 < \lambda_3 < \lambda_4 = \infty$. Podsjetnik: funkcija gubitka s regularizatorom Ω $\mathrm{oblika\ je}$ $J + \lambda \cdot \Omega(\Theta)$

   (c) Po kojoj osi očekujemo veći pomak točke minimuma u slučaju regularizacije? Zašto?

   NE RADIMO REGULARIZACIJU