Duboko učenje

Unatražni prolaz kroz konvolucije

Siniša Šegvić

Sveučilište u Zagrebu

Fakultet elektrotehnike i računarstva

Sadržaj

- ponavljanje: afini slojevi, širenje gradijenata unatrag
- gradijenti parametara 1D konvolucije
  - jezgra dimenzije 1
  - jezgra dimenzije k
- gradijenti parametara 2D konvolucije
  - jezgra dimenzija k×k
  - zadatak s unatražnim prolazom kroz 2D konvoluciju
- 2D konvolucija nad tenzorima trećeg reda
- učenje konvolucije s korakom
- efikasna implementacija 2D konvolucije

UNATRAŽNI PROLAZ: DUBOKI MODEL

Duboki model: kompozicija parametarskih nelinearnih transformacija

Unaprijedni prolaz

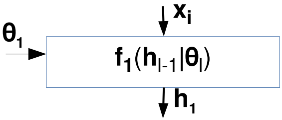
- ulaz: podatak $\mathbf{x}$, izlaz: $\mathrm{softmax}(\mathbf{h}_L)$

UNATRAŽNI PROLAZ: DUBOKI MODEL

Duboki model: kompozicija parametarskih nelinearnih transformacija

Unaprijedni prolaz

- ulaz: podatak $\mathbf{x}$, izlaz: $\mathrm{softmax}(\mathbf{h}_L)$

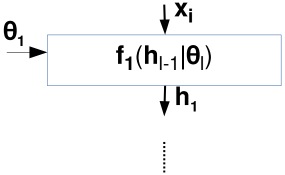


UNATRAŽNI PROLAZ: DUBOKI MODEL

Duboki model: kompozicija parametarskih nelinearnih transformacija

Unaprijedni prolaz

- ulaz: podatak $\mathbf{x}$, izlaz: $\mathrm{softmax}(\mathbf{h}_L)$

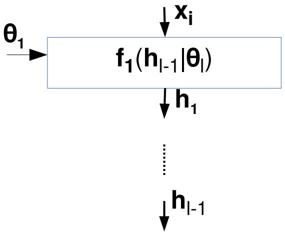


UNATRAŽNI PROLAZ: DUBOKI MODEL

Duboki model: kompozicija parametarskih nelinearnih transformacija

Unaprijedni prolaz

- ulaz: podatak $\mathbf{x}$, izlaz: $\mathrm{softmax}(\mathbf{h}_L)$

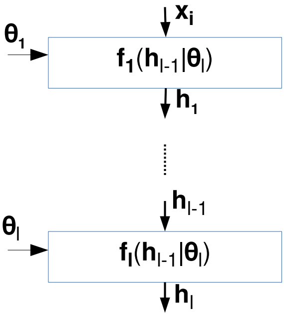


UNATRAŽNI PROLAZ: DUBOKI MODEL

Duboki model: kompozicija parametarskih nelinearnih transformacija

Unaprijedni prolaz

- ulaz: podatak $\mathbf{x}$, izlaz: $\mathrm{softmax}(\mathbf{h}_L)$

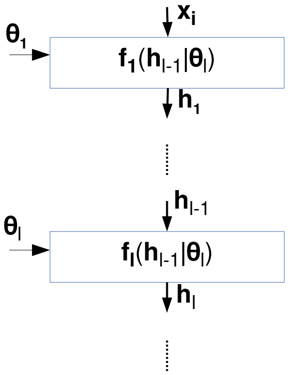


UNATRAŽNI PROLAZ: DUBOKI MODEL

Duboki model: kompozicija parametarskih nelinearnih transformacija

Unaprijedni prolaz

- ulaz: podatak $\mathbf{x}$, izlaz: $\mathrm{softmax}(\mathbf{h}_L)$

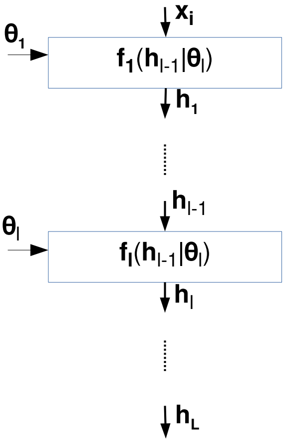


UNATRAŽNI PROLAZ: DUBOKI MODEL

Duboki model: kompozicija parametarskih nelinearnih transformacija

Unaprijedni prolaz

- ulaz: podatak $\mathbf{x}$, izlaz: $\mathrm{softmax}(\mathbf{h}_L)$



UNATRAŽNI PROLAZ: DUBOKI MODEL

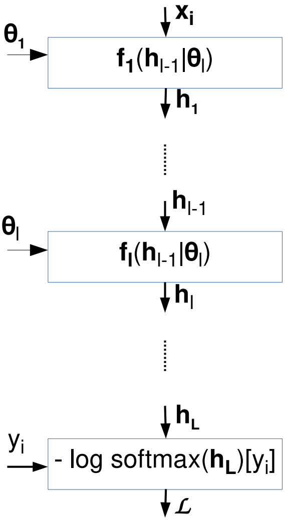
Duboki model: kompozicija parametarskih nelinearnih transformacija

Unaprijedni prolaz

- ulaz: podatak $\mathbf{x}$, izlaz: $\mathrm{softmax}(\mathbf{h}_L)$

Unatražni prolaz

- ulaz: predikcija, oznaka, aktivacije
- izlaz: gradijenti parametara $\partial\mathcal{L}/\partial\boldsymbol{\theta}_l$



UNATRAŽNI PROLAZ: DUBOKI MODEL

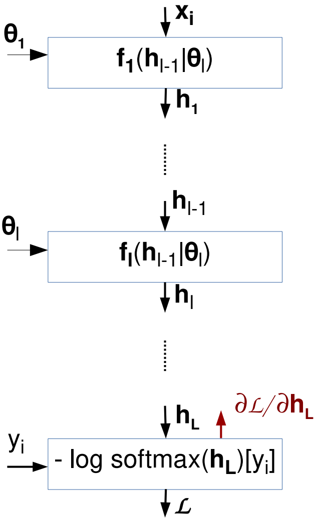
Duboki model: kompozicija parametarskih nelinearnih transformacija

Unaprijedni prolaz

- ulaz: podatak $\mathbf{x}$, izlaz: $\mathrm{softmax}(\mathbf{h}_L)$

Unatražni prolaz

- ulaz: predikcija, oznaka, aktivacije
- izlaz: gradijenti parametara $\partial\mathcal{L}/\partial\boldsymbol{\theta}_l$



UNATRAŽNI PROLAZ: DUBOKI MODEL

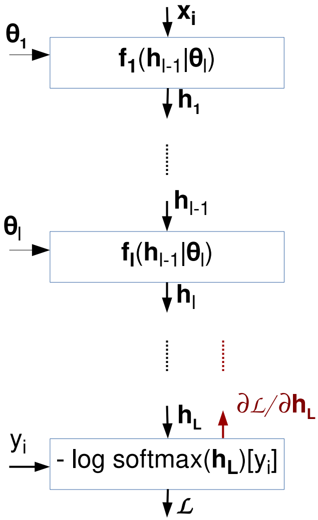
Duboki model: kompozicija parametarskih nelinearnih transformacija

Unaprijedni prolaz

- ulaz: podatak $\mathbf{x}$, izlaz: $\mathrm{softmax}(\mathbf{h}_L)$

Unatražni prolaz

- ulaz: predikcija, oznaka, aktivacije
- izlaz: gradijenti parametara $\partial\mathcal{L}/\partial\boldsymbol{\theta}_l$



UNATRAŽNI PROLAZ: DUBOKI MODEL

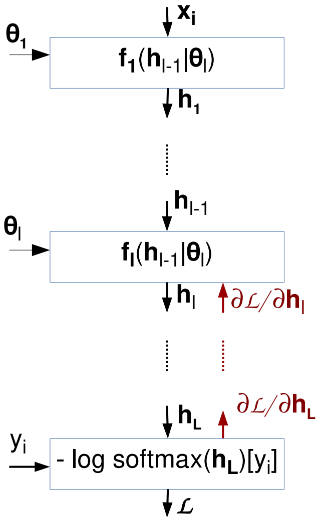
Duboki model: kompozicija parametarskih nelinearnih transformacija

Unaprijedni prolaz

- ulaz: podatak $\mathbf{x}$, izlaz: $\mathrm{softmax}(\mathbf{h}_L)$

Unatražni prolaz

- ulaz: predikcija, oznaka, aktivacije
- izlaz: gradijenti parametara $\partial\mathcal{L}/\partial\boldsymbol{\theta}_l$



UNATRAŽNI PROLAZ: DUBOKI MODEL

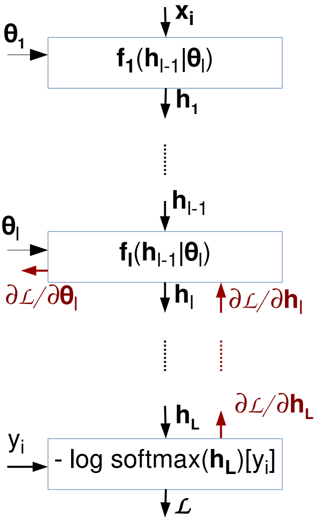
Duboki model: kompozicija parametarskih nelinearnih transformacija

Unaprijedni prolaz

- ulaz: podatak $\mathbf{x}$, izlaz: $\mathrm{softmax}(\mathbf{h}_L)$

Unatražni prolaz

- ulaz: predikcija, oznaka, aktivacije
- izlaz: gradijenti parametara $\partial\mathcal{L}/\partial\boldsymbol{\theta}_l$



UNATRAŽNI PROLAZ: DUBOKI MODEL

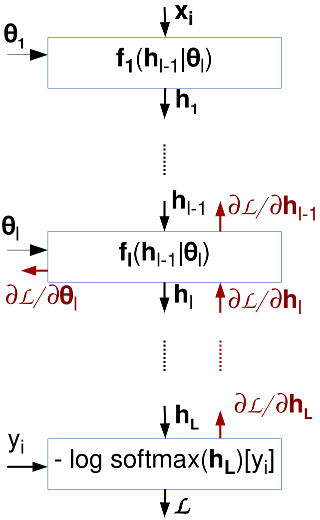
Duboki model: kompozicija parametarskih nelinearnih transformacija

Unaprijedni prolaz

- ulaz: podatak $\mathbf{x}$, izlaz: $\mathrm{softmax}(\mathbf{h}_L)$

Unatražni prolaz

- ulaz: predikcija, oznaka, aktivacije
- izlaz: gradijenti parametara $\partial\mathcal{L}/\partial\boldsymbol{\theta}_l$



UNATRAŽNI PROLAZ: DUBOKI MODEL

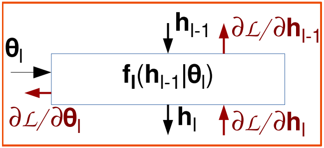
Duboki model: kompozicija parametarskih nelinearnih transformacija

Unaprijedni prolaz

- ulaz: podatak $\mathbf{x}$, izlaz: $\mathrm{softmax}(\mathbf{h}_L)$

Unatražni prolaz

- ulaz: predikcija, oznaka, aktivacije
- izlaz: gradijenti parametara $\partial\mathcal{L}/\partial\boldsymbol{\theta}_l$

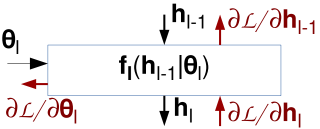


Definicija l-tog sloja:

- unaprijedni prolaz: $\mathbf{h}_l = \mathbf{f}_l(\mathbf{h}_{l-1} \mid \boldsymbol{\theta}_l)$
- određivanje gradijenata ulaza:
  $\partial\mathcal{L}/\partial\mathbf{h}_{l-1} = \partial\mathcal{L}/\partial\mathbf{h}_l \cdot \partial\mathbf{h}_l/\partial\mathbf{h}_{l-1}$
- određivanje gradijenata parametara:
  $\partial\mathcal{L}/\partial\boldsymbol{\theta}_l = \partial\mathcal{L}/\partial\mathbf{h}_l \cdot \partial\mathbf{h}_l/\partial\boldsymbol{\theta}_l$

UNATRAŽNI PROLAZ: DUBOKI MODEL, DETALJI

Pogledajmo dimenzije Jakobijana $\partial\mathcal{L}/\partial\mathbf{h}_l$:



- $\partial\mathcal{L}/\partial\mathbf{h}_l$: isto kao i $\mathbf{h}_l^\top$
  - potpuno povezani sloj: $1 \times D_l$
  - konvolucijski sloj (Torch): $[n, c, h, w]$

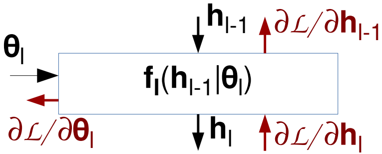
Dimenzije Jakobijana $\partial\mathcal{L}/\partial \boldsymbol{W}_l$ = $(\partial\mathcal{L}/\partial \boldsymbol{W}_{luv})_{u,v=1,1}^{H,W}$:

- $\partial\mathcal{L}/\partial \boldsymbol{W}_l$: isto kao i $\mathbf{W}_l$
  - potpuno povezani sloj: $D_l \times D_{l-1}$
  - jednostavan konvolucijski sloj: $k \times k$
  - konvolucijski sloj (Torch): $[c_{\text{out}}, c_{\text{in}}, k, k]$
- $\partial\mathcal{L}/\partial\mathrm{vec}(\mathbf{W}_l)$: isto kao i $\mathrm{vec}(\mathbf{W}_l)^\top$
  - potpuno povezani sloj: $1 \times D_l \cdot D_{l-1}$
  - jednostavan konvolucijski sloj: $1 \times k \cdot k$

UNATRAŽNI PROLAZ: DUBOKI MODEL, DETALJI (2)

Dimenzije Jakobijana $\partial \mathbf{h}_l / \partial \mathbf{h}_{l-1}$:

- u potpuno povezanom sloju:
  - gusta matrica $D_l \times D_{l-1}$
- u konvolucijskom sloju:
  - rijetki tenzor reda 4 (odnosno 6)
  - nepraktično odrediti ga izravno!

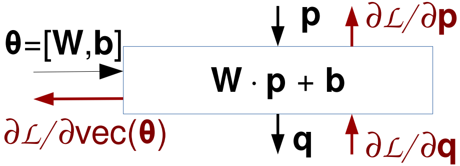


Dimenzije Jakobijana $\partial \mathbf{h}_l / \partial \mathbf{W}_l$:

- u potpuno povezanom sloju:
  - $\partial \mathbf{h}_l / \partial \mathbf{W}_l$: rijetki tenzor trećeg reda
  - $\partial \mathbf{h}_l / \partial \mathrm{vec}(\mathbf{W}_l)$: rijetka matrica $D_l \times D_l \cdot D_{l-1}$
  - u praksi koristimo postupak koji uzima u obzir rijetkost $\partial \mathbf{h}_l / \partial \mathbf{W}_l$
- u konvolucijskom sloju: gusti tenzor reda 4 (odnosno 7)
  - vidjet ćemo da komponente tog tenzora odgovaraju posmaknutom ulazu sloja $\partial \mathbf{h}_l / \partial \mathbf{W}_{luv} = \mathrm{shift}(\mathbf{h}_{l-1}, u - o_k, v - o_k)$, $o_k = \lfloor k/2 \rfloor + 1$

UNATRAŽNI PROLAZ: POTPUNO POVEZANI SLOJ

Unaprijedni prolaz: množenje matricom plus pomak



Gradijenti ulaza: $\partial\mathcal{L}/\partial\mathbf{p} = \partial\mathcal{L}/\partial\mathbf{q} \cdot \partial\mathbf{q}/\partial\mathbf{p} = \partial\mathcal{L}/\partial\mathbf{q} \cdot \mathbf{W}$

Gradijenti vektoriziranih parametara:

$$\partial\mathcal{L}/\partial\mathrm{vec}(\mathbf{W}) = \partial\mathcal{L}/\partial\mathbf{q} \cdot \partial\mathbf{q}/\partial\mathrm{vec}(\mathbf{W})$$

Problem: dimenzije $\partial\mathbf{q}/\partial\mathrm{vec}(\mathbf{W})$

- Jakobijan dimenzija dim($\mathbf{q}$)$\times$ dim($\mathbf{q}$)·dim($\mathbf{p}$)
- npr. za MNIST: $784 \cdot 10^2$ množenja po podatku

UNATRAŽNI PROLAZ: POTPUNO POVEZANI SLOJ (DETALJI)

Jakobijan $\partial\mathbf{q}/\partial\mathbf{W}$ je rijedak zbog specifične strukture ovisnosti:

- svaki izlaz q_j ovisi samo o jednom retku težina - $\mathbf{W}_{j,:}$
- s druge strane, $\partial\mathbf{q}/\partial\mathbf{p}$ je gust jer svaki q_j ovisi o svakom p_i

Ideja 1: iskoristiti strukturu za efikasniji izračun gradijenata parametara

$$\partial\mathcal{L}/\partial\mathbf{W}_{j,:} = \partial\mathcal{L}/\partial q_j \cdot \partial q_j/\partial\mathbf{W}_{j,:} = \partial\mathcal{L}/\partial q_j \cdot \mathbf{p}^\top$$

Ideja 2: izračunati gradijente svih redaka vanjskim produktom

$$\partial\mathcal{L}/\partial\mathbf{W} = [\partial\mathcal{L}/\partial\mathbf{q}]^\top \cdot \mathbf{p}^\top$$

U literaturi se nađe i sljedeći izraz (veza s pravilom ulančavanja?):

$$\partial\mathcal{L}/\partial\mathbf{W}^\top = \mathbf{p} \cdot \partial\mathcal{L}/\partial\mathbf{q}$$

Npr. za MNIST: $784 \cdot 10$ množenja po podatku

- $10\times$ brže nego prije
- više od 10 dimenzija u skrivenom sloju $\Rightarrow$ veće ubrzanje

UNATRAŽNI PROLAZ: POTPUNO POVEZANI SLOJ, GRUPA

Gubitak grupe je prosječan gubitak preko svih podataka:

$$\partial\mathcal{L}/\partial\mathbf{W} = \frac{1}{N}\sum_{i=1}^{N}\frac{\partial\mathcal{L}_i}{\partial\mathbf{W}}$$

Ideja 3: pozbrajati doprinose podataka grupe matričnim množenjem

$$\partial\mathcal{L}_i/\partial\mathbf{W} = [\partial\mathcal{L}_i/\partial\mathbf{q}]^\top \cdot \mathbf{p}_i^\top$$

$$\Rightarrow$$

$$\partial\mathcal{L}/\partial\mathbf{W} = \frac{1}{N}\cdot\mathbf{G}\cdot\mathbf{P}, \text{ gdje su:}$$

$$\mathbf{G}_{:,i} = [\partial\mathcal{L}_i/\partial\mathbf{q}]^\top$$

$$\mathbf{P}_{i,:} = \mathbf{p}_i^\top$$

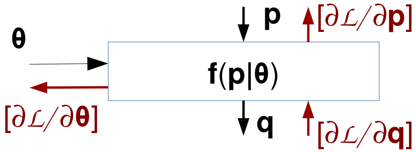
To implicira sljedeću proceduru:

- unaprijedni prolaz pamti ulazne aktivacije svih podataka $\mathbf{P} = \left\{\mathbf{p}_i^\top\right\}$
- backprop računa sve gradijente po izlazu sloja $\mathbf{G} = \{\partial\mathcal{L}_i/\partial\mathbf{q}\}^\top$
- gradijenti gubitka grupe po parametrima su: $\frac{1}{N}\cdot\mathbf{G}\cdot\mathbf{P}$

UNATRAŽNI PROLAZ: SUČELJE SLOJEVA

Vidimo da svaki sloj možemo opisati sljedećim sučeljem:

```
class Layer: # ...
  def fwd_pass(self, p):
    self.p=p
    q=f(p, self.theta)
    return q
  def bwd_pass(self, dq):
    dp = g(dq, self.theta)
    self.dtheta = h(dq, self.p)
    return dp
```



Svaki sloj mora pamtiti:

- parametre (potrebni za gradijente po ulazu sloja)
- ulazne aktivacije (potrebne za gradijente po parametrima)
- gradijente parametara (potrebni za učenje)

UNATRAŽNI PROLAZ: ALGORITAM

Sad smo spremni skicirati općeniti backprop:

- ovaj algoritam pokrećemo pozivom `loss.backward()` u torchu
- više o unatražnim automatskim gradijentima: [gavranovic18sem]

```
def backprop(L, x,y):
  # forward pass
  q=x # input
  for l in L:
    q=l.fwd_pass(q)
  # backward pass, q=logits
  dq=grad_loss(q, y)
  for l in reversed(L):
    dq=l.bwd_pass(dq)
```

U ovu formulaciju mogu se uklopiti svi slojevi koji:

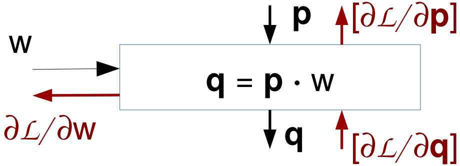
- zadovoljavaju sučelje `Layer`
- imaju dimenzije ulaza i izlaza kompatibilne sa susjedima

Sadržaj

- ponavljanje: unatražno učenje, potpuno povezani slojevi
- **gradijenti parametara 1D konvolucije**
  - jezgra dimenzije 1
  - jezgra dimenzije k
- gradijenti parametara 2D konvolucije
  - jezgra dimenzija k×k
  - zadatak s unatražnim prolazom kroz 2D konvoluciju
- 2D konvolucija nad tenzorima trećeg reda
- učenje konvolucije s korakom
- efikasna implementacija 2D konvolucije

Vektori: skalarna jezgra, jednadžbe

Unaprijedni prolaz: $\mathbf{q} = w \cdot \mathbf{p}$



Gradijenti po ulazu:

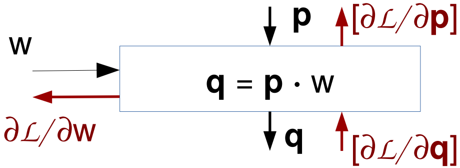
$$\begin{aligned}
\partial\mathcal{L}/\partial\mathbf{p} &= \partial\mathcal{L}/\partial\mathbf{q} \cdot \partial\mathbf{q}/\partial\mathbf{p} \\
&= \partial\mathcal{L}/\partial\mathbf{q} \cdot (w \cdot \mathbf{I}) \\
&= w \cdot \partial\mathcal{L}/\partial\mathbf{q}
\end{aligned}$$

Gradijenti po parametrima:

$$\begin{aligned}
\partial\mathcal{L}/\partial w &= \partial\mathcal{L}/\partial\mathbf{q} \cdot \partial\mathbf{q}/\partial w \\
&= \partial\mathcal{L}/\partial\mathbf{q} \cdot \mathbf{p}
\end{aligned}$$

Vektori: skalarna jezgra, jednadžbe (2)

Unaprijedni prolaz: $\mathbf{q} = w \cdot \mathbf{p}$



Gradijenti po ulazu: $\partial\mathcal{L}/\partial\mathbf{p} = w \cdot \partial\mathcal{L}/\partial\mathbf{q}$

Gradijenti po parametrima: $\partial\mathcal{L}/\partial w = \partial\mathcal{L}/\partial\mathbf{q} \cdot \mathbf{p}$

Kakve razlike možemo očekivati u "pravoj" konvoluciji?

- gradijenti po ulazu biti će složeniji: $\mathbf{q}_i$ ovisi o <u>susjedstvu</u> $\mathcal{N}(\mathbf{p}_i)$
- imat ćemo više gradijenata po parametrima, ali oni će biti jednako složeni kao i ovdje

Vektori: skalarna jezgra, kôd

Sloj kojeg smo upravo definirali možemo opisati sljedećim kodom:

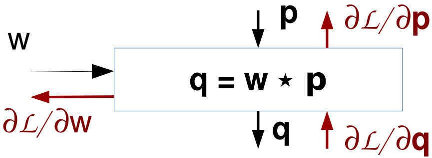
```
class LayerConv1x1: #...
  def fwd_pass(self, p):
    self.p=p
    return self.w*self.p
  def bwd_pass(self,dq):
    dp = self.w * dq
    self.dw = np.dot(dq,self.p)
    return dp
```

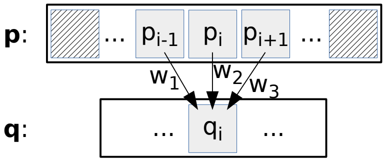
Kao i ranije, sloj pamti aktivacije, parametre i gradijente po parametrima

Sloj mora dodatno omogućiti:

- pristup parametru i njegovom gradijentu
- postavljanje parametra (inicijalizacija, učenje)

Vektori: jezgra dimenzije k, naprijed





Kod 1D konvolucije i-ta izlazna aktivacija odgovara **skalarnom produktu** isječka (crop) ulaza i jezgre:

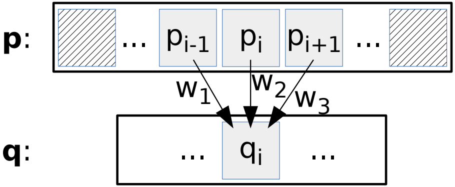
$$q_i = \mathrm{crop}_k(\mathbf{p}, i)^\top \cdot \mathbf{w}$$

Primjerice, za k=3, i=5 imali bismo: $q_5 = p_4 \cdot w_1 + p_5 \cdot w_2 + p_6 \cdot w_3$

Općenita jednadžba i-te aktivacije ($o_k = \lfloor k/2 \rfloor + 1$, o_3=2):

$$q_i = \mathrm{crop}_k(\mathbf{p}, i)^\top \cdot \mathbf{w} = \sum_{u=1}^{k} p_{i-o_k+u} \cdot w_u$$

VEKTORI: JEZGRA DIMENZIJE K, NATRAG W (1)



$$q_i = \sum_{u=1}^{k} p_{i-o_k+u} \cdot w_u$$

Za gradijent po parametru w_u postavljamo pitanje: kako q_i ovisi o w_u?

$$\partial q_i / \partial w_u = p_{i-o_k+u}$$

Traženi gradijent je skalarni produkt gradijenata po izlazu i posmaknutog (shift) ulaza ("ilegalne pristupe" rješava nadopunjavanje):

$$\partial \mathcal{L} / \partial w_u = \sum_i \partial \mathcal{L} / \partial q_i \cdot \partial q_i / \partial w_u = \sum_i \partial \mathcal{L} / \partial q_i \cdot p_{i-o_k+u}$$

$$= \partial \mathcal{L} / \partial \mathbf{q} \cdot \mathrm{shift}(\mathbf{p}, -o_k + u)$$

Vektori: jezgra dimenzije k, natrag w (2)

Raspišimo dobivene gradijente po parametrima za k=3:

$$\partial\mathcal{L}/\partial w_1 = \sum_i \partial\mathcal{L}/\partial q_i \cdot p_{i-1} = \partial\mathcal{L}/\partial\mathbf{q} \cdot \mathrm{shift}(\mathbf{p}, -1)$$

$$\partial\mathcal{L}/\partial w_2 = \sum_i \partial\mathcal{L}/\partial q_i \cdot p_i = \partial\mathcal{L}/\partial\mathbf{q} \cdot \mathrm{shift}(\mathbf{p}, 0)$$

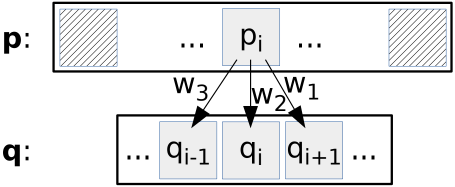
$$\partial\mathcal{L}/\partial w_3 = \sum_i \partial\mathcal{L}/\partial q_i \cdot p_{i+1} = \partial\mathcal{L}/\partial\mathbf{q} \cdot \mathrm{shift}(\mathbf{p}, 1)$$

Vidimo da se gradijenti svih parametara mogu dobiti unakrsnom korelacijom nadopunjenog (pad) ulaza s gradijentima po izlazu:

$$\partial\mathcal{L}/\partial\mathbf{w} = \partial\mathcal{L}/\partial\mathbf{q} \star \mathrm{pad}(\mathbf{p}, \lfloor k/2 \rfloor)$$

Vektori: jezgra dimenzije k, natrag p (1)

Za gradijente po ulazu sloja ključno je pitanje: koji q_r ovise o p_i i kako?



$$\frac{\partial q_{i-1}}{\partial p_i} = w_3$$

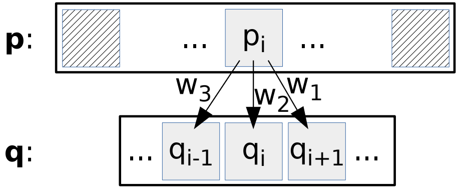
$$\frac{\partial q_{i+1}}{\partial p_i} = w_1$$

$$\frac{\partial q_{i+s}}{\partial p_i} = w_{o_k - s} = w_{2-s}$$

Fokusirajmo se na p_i kao na slici ($o_k = \lfloor k/2 \rfloor + 1$, o_3=2):

$$\begin{aligned}
\partial \mathcal{L}/\partial p_i &= \sum_r \partial \mathcal{L}/\partial q_r \cdot \partial q_r/\partial p_i \\
&= \sum_{s=-\lfloor k/2 \rfloor}^{\lfloor k/2 \rfloor} \partial \mathcal{L}/\partial q_{i+s} \cdot \partial q_{i+s}/\partial p_i \\
&= \sum_{s=-\lfloor k/2 \rfloor}^{\lfloor k/2 \rfloor} \partial \mathcal{L}/\partial q_{i+s} \cdot w_{o_k - s}
\end{aligned}$$

VEKTORI: JEZGRA DIMENZIJE K, NATRAG P (2)



$$\frac{\partial q_{i-1}}{\partial p_i} = w_3$$

$$\frac{\partial q_{i+1}}{\partial p_i} = w_1$$

$$\frac{\partial q_{i+s}}{\partial p_i} = w_{2-s}$$

$$\partial \mathcal{L}/\partial p_i = \sum_{s=-1}^{1} \partial \mathcal{L}/\partial q_{i+s} \cdot w_{o_k - s}$$

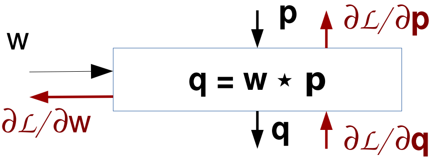
Vidimo da se gradijent po i-tom ulazu dobiva **skalarnim produktom** isječka (crop) gradijenata po izlazu i zrcaljene (flip) jezgre:

$$\partial \mathcal{L}/\partial p_i = \mathrm{sum}(\mathrm{crop}_k(\partial \mathcal{L}/\partial \mathbf{q}, i) \odot \mathrm{flip}(\mathbf{w}))$$

Gradijente po svim ulazima dobivamo **unakrsnom korelacijom** nadopunjenih (pad) gradijenata po izlazu i zrcaljene (flip) jezgre:

$$\partial \mathcal{L}/\partial \mathbf{p} = \mathrm{flip}(\mathbf{w}) \star \mathrm{pad}(\partial \mathcal{L}/\partial \mathbf{q}, \lfloor k/2 \rfloor)$$

Vektori: jezgra dimenzije k, sažetak



Unaprijedni prolaz:

$$q_i = \mathrm{crop}_k(\mathbf{p}, i)^\top \cdot \mathbf{w}$$

$$\mathbf{q} = \mathbf{w} \star \mathbf{p}$$

Unatražni prolaz po parametrima:

$$\partial\mathcal{L}/\partial\mathbf{w} = \partial\mathcal{L}/\partial\mathbf{q} \star \mathrm{pad}(\mathbf{p}, \lfloor k/2 \rfloor)$$

Unatražni prolaz po ulazima:

$$\partial\mathcal{L}/\partial\mathbf{p} = \mathrm{flip}(\mathbf{w}) \star \mathrm{pad}(\partial\mathcal{L}/\partial\mathbf{q}, \lfloor k/2 \rfloor)$$

Vektori: kod

```
def my_conv1d(data, kernel):
  data_pad = torch.nn.functional.pad(data,(1,1),'constant',0)
  return torch.nn.functional.conv1d(
    data_pad.reshape([1,1,-1]),
    kernel.reshape([1,1,-1])).squeeze()

D,K = 7,3
w = torch.tensor(np.random.randn(K), requires_grad=True)
p = torch.tensor(np.random.randn(D), requires_grad=True)
q = my_conv1d(p, w)
torch.sum(q).backward()

dLdq = torch.ones(D, dtype=torch.float64) # dL/dq_j=1!
dLdp = my_conv1d(dLdq, torch.flip(w, (0,)))
dLdw = my_conv1d(p, dLdq)

print(dLdw, w.grad)
print(dLdp, p.grad)
```

Vektori: zglobnica

Unaprijedni prolaz nezavisno primjenjuje zglobnicu u svakom ulazu:

$$q_i = \max(0, p_i)$$

Unatražni prolaz (svaki q_i ovisi samo o p_i):

$$\begin{aligned}
\partial\mathcal{L}/\partial\mathbf{p} &= \partial\mathcal{L}/\partial\mathbf{q} \cdot \partial\mathbf{q}/\partial\mathbf{p} \\
&= \partial\mathcal{L}/\partial\mathbf{q} \cdot \begin{bmatrix} [\![p_1 > 0]\!] & & & \\ & [\![p_2 > 0]\!] & & \\ & & \ddots & \\ & & & [\![p_n > 0]\!] \end{bmatrix} \\
&= \partial\mathcal{L}/\partial\mathbf{q} \odot [\![\mathbf{p} > 0]\!]
\end{aligned}$$

Posljednji izraz primjenljiv je i na tenzore višeg reda.

Vektori: globalno sažimanje maksimumom

Globalno sažimanje maksimumom reducira ulaz u skalar:

$$q = \mathrm{maxpool}(\mathbf{p})$$

Unatražni prolaz izražavamo funkcijom onehot

- vrijedi npr $\mathrm{onehot}^4(2) = \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix}$

$$\begin{aligned} \partial\mathcal{L}/\partial\mathbf{p} &= \partial\mathcal{L}/\partial q \cdot \partial q/\partial\mathbf{p} \\ &= \partial\mathcal{L}/\partial q \cdot \big[\, [\![\arg\max(\mathbf{p}) = 1]\!], [\![\arg\max(\mathbf{p}) = 2]\!],\ \ldots \\ &\qquad\qquad\qquad\qquad \ldots,\ [\![\arg\max(\mathbf{p}) = \dim(\mathbf{p})]\!] \,\big] \\ &= \partial\mathcal{L}/\partial q \cdot \mathrm{onehot}^n(\arg\max\mathbf{p}) \end{aligned}$$

Posljednji izraz primjenljiv je i na tenzore višeg reda, samo tada umjesto n navodimo $\mathrm{shape}(\mathbf{p})$

Vektori: sažimanje s jezgrom k i korakom k

Sažimanje $\mathrm{maxpool}_k$ zasebno reducira nepreklapajuće **regije** ulaza veličine k:

$$\mathbf{q} = \mathrm{maxpool}_k(\mathbf{p}) \ .$$

Ulaz se poduzorkuje $k\times$: $\dim(\mathbf{p}) = N \Rightarrow \dim(\mathbf{q}) = \lceil N/k \rceil$.

Primjer:

$$\mathrm{maxpool}_2\left(\begin{bmatrix}1 & 2 & 3 & 4 & 5 & 6\end{bmatrix}\right) = \begin{bmatrix}2 & 4 & 6\end{bmatrix}$$

Vektori: sažimanje s jezgrom k i korakom k (natrag)

Unatražni prolaz izražavamo funkcijom $\mathrm{embed}_k^n(\mathbf{x}, p)$:

- ulaz: $\mathbf{x}$, $\dim(\mathbf{x}) = k$,
- izlaz: $\mathbf{x}'$, $\dim(\mathbf{x}') = n$, takav da $\mathbf{x}'_{[p:p+k]} = \mathbf{x}$
- vrijedi: $\mathrm{onehot}_1^n(p) = \mathrm{embed}_1^n(1, p)$

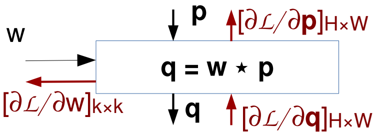
Unatražni prolaz širi $\partial\mathcal{L}/\partial q_i$ kao kod globalnog sažimanja (onehot_1^n) te rezultat ugrađuje u $\partial\mathcal{L}/\partial\mathbf{p}$ primjenom funkcije embed_k^n:

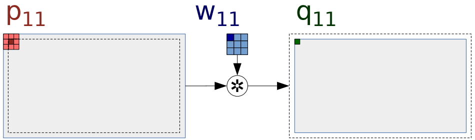
$$
\begin{aligned}
\partial\mathcal{L}/\partial\mathbf{p} &= \sum_{i=0}^{\lfloor N/k\rfloor-1} \partial\mathcal{L}/\partial q_i \cdot \mathrm{embed}_k^N(\partial q_i/\partial\mathbf{p}_{[k\cdot i:k\cdot i+k]}, k\cdot i) \\
&= \sum_{i=0}^{\lfloor N/k\rfloor-1} \partial\mathcal{L}/\partial q_i \cdot \mathrm{embed}_k^N(\mathrm{onehot}_1^k(\arg\max \mathbf{p}_{[k\cdot i:k\cdot i+k]}), k\cdot i)
\end{aligned}
$$

Sadržaj

- ponavljanje: unatražno učenje, potpuno povezani slojevi
- gradijenti parametara 1D konvolucije
  - jezgra dimenzije 1
  - jezgra dimenzije k
- **gradijenti parametara 2D konvolucije**
  - jezgra dimenzija k×k
  - zadatak s unatražnim prolazom kroz 2D konvoluciju
- 2D konvolucija nad tenzorima trećeg reda
- učenje konvolucije s korakom
- efikasna implementacija 2D konvolucije

Matrice: jezgra kxk, naprijed





Izlazne aktivacije 2D konvolucije odgovaraju **skalarnom produktu** pravokutnog isječka ulaza (crop) i jezgre:

$$q_{ij} = \mathrm{sum}(\mathrm{crop}_{k\times k}(\mathbf{p}, i, j) \odot \mathbf{w})$$

Raspišimo jednadžbu jedne aktivacije uz $o_k = \lfloor k/2 \rfloor + 1$:

$$q_{ij} = \mathrm{sum}(\mathrm{crop}_{k\times k}(\mathbf{p}, i, j) \odot \mathbf{w}) = \sum_{uv} \mathbf{p}_{i-o_k+u,j-o_k+v} \cdot \mathbf{w}_{uv} \ .$$

Npr. za i=j=1 i k=3 (o_k=2):

$$\mathbf{q}_{11} = \sum_{uv=1}^{3} \mathbf{p}_{u-1,v-1} \cdot \mathbf{w}_{uv}$$

Matrice: kxk, natrag

Dimenzije tenzora koji sudjeluju u unatražnom prolazu:

- ulaz (gradijent po izlazu): $\partial\mathcal{L}/\partial\mathbf{q}$ - H$\times$W
- izlaz (gradijent po ulazu): $\partial\mathcal{L}/\partial\mathbf{p}$ - H$\times$W
  - pretp: unaprijedni prolaz koristio nadopunjavanje
- izlaz (gradijent po parametrima): $[\partial\mathcal{L}/\partial\mathbf{w}]$ - (k$\times$k)

Specifičnost: izlazi $\mathbf{q}$ ne ovise o svim ulazima $\mathbf{p}$

Opći oblik gradijenata po jednom pikselu ulaza:

$$\partial\mathcal{L}/\partial p_{ij} = \partial\mathcal{L}/\partial\mathrm{vec}(\mathbf{q}) \cdot \partial\mathrm{vec}(\mathbf{q})/\partial p_{ij} = \textstyle\sum_r \partial\mathcal{L}/\partial q_r \cdot \partial q_r/\partial p_{ij}$$

Ključna pitanja za gradijente po ulazima:

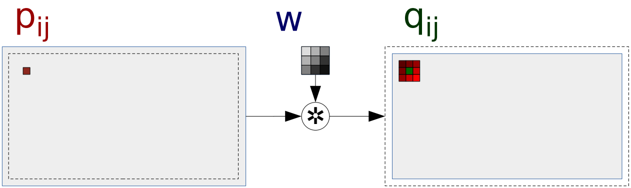
- koji q_r-ovi ovise o p_{ij}?
- za koje r je $\partial q_r/\partial p_{ij} \neq 0$?

MATRICE: KXK, NATRAG, ULAZI

Prepostavimo da je u cijelom $\mathbf{p}$-u upaljen <u>samo jedan</u> piksel: p_{ij}

Tada će u cijelom $\mathbf{q}$ biti upaljeni pikseli:

$q_{i+s,j+t} = p_{ij} \cdot w_{o_k-s,o_k-t}$, gdje su $s, t \in -\lfloor k/2 \rfloor .. \lfloor k/2 \rfloor$



Označeni pikseli tenzora $\mathbf{q}$: "zona utjecaja" piksela p_{ij}

- susjedstvo veličine k×k čije je središte q_{ij}
- za te piksele je $\partial q_r / \partial p_{ij} \neq 0$!
- vrijedi: $\partial q_{i+s,j+t} / \partial p_{ij} = \mathbf{w}_{o_k-s,o_k-t}$
- donji desni element susjedstva - gornji lijevi element jezgre:
  - $\partial q_{i+1,j+1} / \partial p_{ij} = \mathbf{w}_{1,1}$, uz k=3

Matrice: kxk, natrag, ulazi (2)

Što se zbiva kada upalimo sve aktivacije $\mathbf{p}$?

- aktivacije $\mathbf{q}$ odgovaraju zbroju doprinosa svih p_{ij}
- to znači da gradijenti $\partial q_{i+s,j+t}/\partial p_{ij}$ ostaju isti.

Sada možemo zapisati gradijente po ulazu:

$$\begin{aligned} \partial \mathcal{L}/\partial p_{ij} &= \sum_{s,t} \partial \mathcal{L}/\partial q_{i+s,j+t} \cdot \partial q_{i+s,j+t}/\partial p_{ij} \\ &= \sum_{s,t} \partial \mathcal{L}/\partial q_{i+s,j+t} \cdot w_{o_k-s,o_k-t} \end{aligned}$$

Oni odgovaraju **skalarnom produktu** pravokutnog isječka (crop) gradijenata po izlazu i zrcaljene $(\mathrm{flip2d})$ jezgre (kao i u 1D):

$$\partial \mathcal{L}/\partial p_{ij} = \mathrm{sum}(\mathrm{crop}_{k\times k}(\partial \mathcal{L}/\partial \mathbf{q}, i, j) \odot \mathrm{flip2d}(\mathbf{w}))$$

Matrice: kxk, natrag, ulazi (3)

Vidjeli smo da se gradijenti po ulazu dobivaju klizanjem zrcaljene jezgre po gradijentima po izlazu:

$$\partial\mathcal{L}/\partial p_{ij} = \mathrm{sum}(\mathrm{crop}_{k\times k}(\partial\mathcal{L}/\partial\mathbf{q}, i, j) \odot \mathrm{flip2d}(\mathbf{w}))$$

Sada je jasno da se gradijenti po ulazu dobivaju unakrsnom korelacijom gradijenta po izlazu sa zrcaljenom jezgrom $\mathrm{flip2d}(\mathbf{w})$.

Ako želimo dobiti gradijente u svim pikselima ulaza, gradijente po izlazu moramo nadopuniti s $\lfloor k/2 \rfloor$ nula sa <u>svake</u> strane:

$$\partial\mathcal{L}/\partial\mathbf{p} = \mathrm{flip2d}(\mathbf{w}) \star \mathrm{pad}(\partial\mathcal{L}/\partial\mathbf{q}, \lfloor k/2 \rfloor)$$

Ako u unaprijednom prolazu nismo koristili nadopunjavanje, onda moramo nadopuniti s k-1 nula (sa svake strane)

Matrice: kxk, transponirana konvolucija

Posljednju operaciju nazivamo transponiranom konvolucijom

- konvoluirani tenzor se nadopuni $\lfloor k/2 \rfloor$ nula
- konvolucijska jezgra se zrcali po obje osi
- stariji naziv: ~~dekonvolucija~~; bolje: unatražna konvolucija

Uvodimo novi operator: $\mathbf{w} \star^{\top} \mathbf{x} \triangleq \mathrm{flip2d}(\mathbf{w}) \star \mathrm{pad}(\mathbf{x}, \lfloor k/2 \rfloor)$

Kasnije ćemo upoznati dodatna svojstva transponirane konvolucije

- kod tenzora trećeg reda, jezgru treba zrcaliti samo po prostornim osima
- može se koristiti i u unaprijednom prolazu

Matrice: k×k, natrag, parametri

Ako ugasimo sve elemente jezgre osim w_{uv}, $u, v \in 1..k$, dobivamo:

$$q_{ij} = p_{i-o_k+u,j-o_k+v} \cdot w_{uv}, \forall i, j$$

Vidimo da parametri utječu na sve izlaze $\rightarrow$ potrebno uzeti u obzir sve lokacije izlaznog tenzora:

$$\begin{aligned}\partial\mathcal{L}/\partial w_{uv} &= \sum_{ij} \partial\mathcal{L}/\partial q_{ij} \cdot \partial q_{ij}/\partial w_{uv} \\ &= \sum_{ij} \partial\mathcal{L}/\partial q_{ij} \cdot p_{i-o_k+u,j-o_k+v}\end{aligned}$$

Ovaj rezultat možemo interpretirati kao redukciju Hadamardovog (elementwise) produkta gradijenata izlaza i posmaknutih ulaza:

$$\begin{aligned}\partial\mathcal{L}/\partial w_{uv} &= \sum_{ij} \partial\mathcal{L}/\partial q_{ij} \cdot p_{i-o_k+u,j-o_k+v} \\ &= \mathrm{sum}(\mathrm{shift}(\mathbf{p}, -o_k + u, -o_k + v) \odot \partial\mathcal{L}/\partial\mathbf{q})\end{aligned}$$

MATRICE: K×K, NATRAG, PARAMETRI (2)

Kada upalimo sve piksele jezgre $\mathbf{w}$:

- aktivacije $\mathbf{q}$ odgovaraju zbroju doprinosa svih w_{uv}
- to znači da gradijenti $\partial q_{ij}/\partial w_{uv}$ ostaju isti

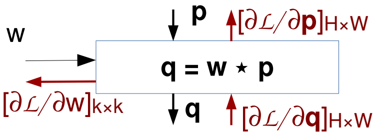
Gradijente po svim težinama sada možemo dobiti unakrsnom korelacijom gradijenta po izlazu s nadopunjenim ulazom:

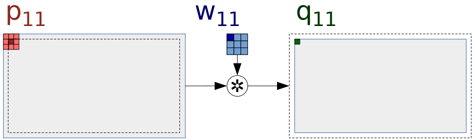
$$\partial\mathcal{L}/\partial\mathbf{w} = \partial\mathcal{L}/\partial\mathbf{q} \star \mathrm{pad}(\mathbf{p}, \lfloor k/2 \rfloor)$$

Prethodna jednadžba vrijedi ako izlaz i ulaz imaju istu rezoluciju zbog nadopunjavanja u unaprijednom prolazu.

Ako unaprijedni prolaz nije koristio nadopunjavanje, onda i gradijente po parametrima možemo izračunati bez nadopunjavanja.

MATRICE: K×K, NATRAG, SAŽETAK





Unaprijedni prolaz:

$$q_{ij} = \mathrm{sum}(\mathrm{crop}_{k\times k}(\mathbf{p}, i, j) \odot \mathbf{w})$$

$$\mathbf{q} = \mathbf{w} \star \mathbf{p}$$

Unatražni prolaz po parametrima:

$$\partial\mathcal{L}/\partial\mathbf{w} = \partial\mathcal{L}/\partial\mathbf{q} \star \mathrm{pad}(\mathbf{p}, \lfloor k/2 \rfloor)$$

Unatražni prolaz po ulazima:

$$\partial\mathcal{L}/\partial\mathbf{p} = \mathrm{flip2d}(\mathbf{w}) \star \mathrm{pad}(\partial\mathcal{L}/\partial\mathbf{q}, \lfloor k/2 \rfloor)$$

Matrice: zadatak

Razmatramo konvolucijski model za klasificiranje jednokanalne slike:

- konvolucija 3x3 bez pomaka i nadopunjavanja s aktivacijom ReLU
  - dvije izlazne mape značajki računaju jezgre $\mathbf{w}_1$ i $\mathbf{w}_2$
- globalno sažimanje maksimumom
- potpuno povezani sloj bez pomaka ($\mathbf{W}$), softmaks

Inicijalizacija:

- w_{121}=-1, w_{122}=1, svi ostali elementi $\mathbf{w}_1$ su 0
- w_{232}=-1, w_{222}=1, svi ostali elementi $\mathbf{w}_2$ su 0
- $W_{11} = W_{22}$=1, svi ostali elementi $\mathbf{W}$ su 0

Na ulazu je 4×4 matrica $\mathbf{x}$, x_{22}=1, x_{33}=1, svi ostali elementi su 0

Odredite gradijente negativne log-izglednosti uz y=2

Matrice: rješenje

```
import torch
import numpy as np

x = np.zeros([4,4])
x[1,1] = x[2,2] = 1
x = x.reshape([1,*x.shape])
x = torch.tensor(x, requires_grad=True)

w1 = np.zeros([3,3])
w1[1,0], w1[1,1] = -1,1
w1 = w1.reshape([1,1,*w1.shape])
w1 = torch.tensor(w1, requires_grad=True)

w2 = np.zeros([3,3])
w2[2,1], w2[1,1] = -1,1
w2 = w2.reshape([1,1,*w2.shape])
w2 = torch.tensor(w2, requires_grad=True)
W = torch.tensor(np.eye(2), requires_grad=True)
```

Matrice: rješenje (2)

```
f1 = torch.nn.functional.conv2d(x, w1)
f1r = torch.nn.functional.relu(f1)
f1m = torch.nn.functional.max_pool2d(f1r, [2,2])

f2 = torch.nn.functional.conv2d(x, w2)
f2r = torch.nn.functional.relu(f2)
f2m = torch.nn.functional.max_pool2d(f2r, [2,2])

h = torch.concat([f1m,f2m]).squeeze()
s = W @ h

for t in [f1,f2,h,s]: t.retain_grad()
L = torch.nn.functional.cross_entropy(s, torch.tensor(1))
L.backward()

for t in [s,h,W,f1,f2,w1,w2]:
  print(t.data, t.grad.data, sep='\n', end='\n\n')
```

Matrice: rješenje (3)

$$X = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, y = 2$$

$$\frac{\partial L}{\partial \mathbf{w_1}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.5 \end{bmatrix}$$

$$\frac{\partial L}{\partial \mathbf{w_2}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -0.5 & 0 \\ 0 & 0 & -0.5 \end{bmatrix}$$

Interpretacija: nakon koraka optimizacije grana 1 će davati slabiji a grana 2 jači odziv.

Sadržaj

- ponavljanje: unatražno učenje, potpuno povezani slojevi
- gradijenti parametara 1D konvolucije
  - jezgra dimenzije 1
  - jezgra dimenzije k
- gradijenti parametara 2D konvolucije
  - jezgra dimenzija k×k
  - zadatak s unatražnim prolazom kroz 2D konvoluciju
- **2D konvolucija nad tenzorima trećeg reda**
- učenje konvolucije s korakom
- efikasna implementacija 2D konvolucije

Višekanalni 2D slučaj: naprijed

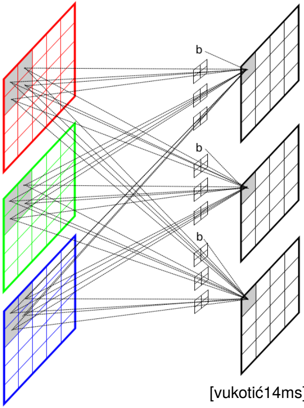
Zadržavamo standardnu sintaksu:

$$\mathbf{q} = \mathbf{w} \star \mathbf{p} + \text{broadcast}(\mathbf{b})$$

Izlaz $\mathbf{q}^{(g)}$ zbraja konvolucije odgovarajućih kriški ulaza i g-te konvolucijske jezgre:

$$\mathbf{q}^{(g)} = \sum_f \mathbf{w}^{(g,f)} \star \mathbf{p}^{(f)}$$

$$q_{ij}^{(g)} = \sum_{fuv} p_{i-o_k+u,j-o_k+v}^{(f)} \cdot w_{uv}^{(g,f)}$$



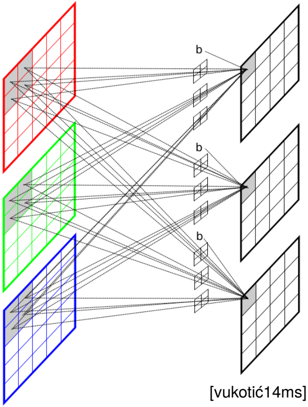
[vukotić14ms]

Gradijenti pomaka odgovaraju zbroju gradijenata izlaza:

$$\partial\mathcal{L}/\partial b_g = \sum_{ij} \partial\mathcal{L}/\partial q_{ij}^{(g)} \cdot \partial q_{ij}^{(g)}/\partial b_g = \sum_{ij} \partial\mathcal{L}/\partial q_{ij}^{(g)}$$

Višekanalni 2D slučaj: natrag, ulazi

Ulaz $\mathbf{p}^{(f)}$ utječe na svaku krišku $\mathbf{q}^{(g)}$ preko konvolucije s $\mathbf{w}^{(g,f)}$



[vukotić14ms]

Stoga, gradijenti po f-toj kriški ulaza zbrajaju gradijente iz svih $\mathbf{q}^{(g)}$:

$$\partial\mathcal{L}/\partial\mathbf{p}_{ij}^{(f)} = \sum_g \sum_{s,t} \partial\mathcal{L}/\partial\mathbf{q}_{i+s,j+t}^{(g)} \cdot w_{o_k-s,o_k-t}^{(g,f)}$$

Kao i ranije, te gradijente možemo izraziti konvolucijom sa zrcaljenom jezgrom:

$$\partial\mathcal{L}/\partial\mathbf{p}^{(f)} = \sum_g \mathrm{flip2d}(\mathbf{w}^{(g,f)}) \star \mathrm{pad}(\partial\mathcal{L}/\partial\mathbf{q}^{(g)}, \lfloor k/2 \rfloor)$$

Posljednja operacija odgovara transponiranoj konvoluciji za 3D tenzore:

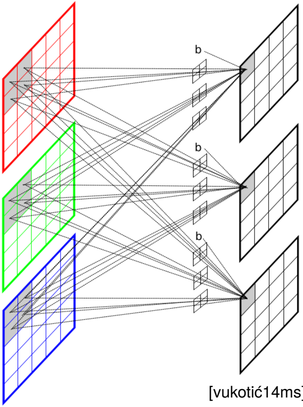
$$\mathbf{w} \star^\top \mathbf{x} \triangleq \sum_g \mathrm{flip2d}(\mathbf{w}^{(g,f)}) \star \mathrm{pad}(\mathbf{x}^{(g)}, \lfloor k/2 \rfloor)$$

Višekanalni 2D slučaj: natrag, parametri

Izlazni tenzor zbraja doprinose odgovarajućih kriški jezgre i ulaza

- f-ta kriška g-te jezgre utječe na g-tu krišku izlaza konvolucijom sa f-tom kriškom ulaza

Zato gradijente po parametrima računamo odvojeno za svaku krišku jezgre



[vukotić14ms]

Konačni izraz je potpuno isti kao i kod 2D tenzora:

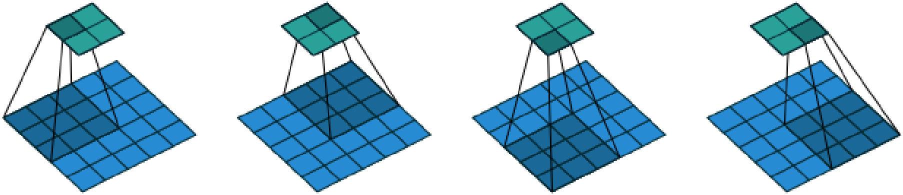
$$\partial\mathcal{L}/\partial\mathbf{w}^{(g,f)} = \partial\mathcal{L}/\partial\mathbf{q}^{(g)} \star \mathrm{pad}(\mathbf{p}^{(f)}, \lfloor k/2 \rfloor)$$

Sadržaj

- ponavljanje: unatražno učenje, potpuno povezani slojevi
- gradijenti parametara 1D konvolucije
  - jezgra dimenzije 1
  - jezgra dimenzije k
- gradijenti parametara 2D konvolucije
  - jezgra dimenzija k×k
  - zadatak s unatražnim prolazom kroz 2D konvoluciju
- 2D konvolucija nad tenzorima trećeg reda
- **učenje konvolucije s korakom**
- efikasna implementacija 2D konvolucije

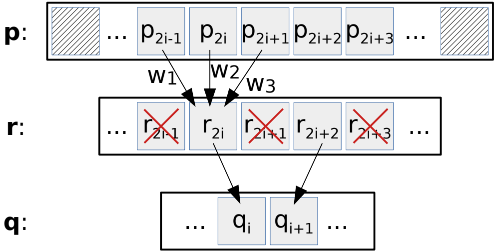
KORAK: NAPRIJED

Konvolucija s izlaznim korakom s ostavlja svaki s-ti element izlaza



[dumoulin16arxiv]

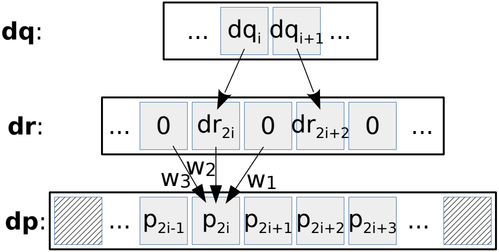
Kao da smo nakon standardne konvolucije proveli poduzorkovanje $\times s$:



Korak: natrag ulančavanjem

Vraćamo se kroz konvoluciju i poduzorkovanje $\times$s:

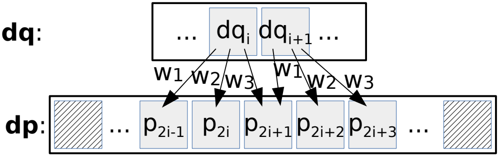
- nakon što gradijenti prođu kroz poduzorkovanje, gradijent svakog preskočenog elementa izlaza konvolucije postaje nula
- nastavljamo normalnim unatražnim prolazom kroz konvoluciju.
- efektivno, receptivno polje transformacije $\partial\mathcal{L}/\partial\mathbf{q} \to \partial\mathcal{L}/\partial\mathbf{p}$ za s$\times$ je manje od receptivnog polja konvolucije s istom jezgrom
  - zato ovu operaciju neki zovu frakcijska konvolucija (Theano)



Korak: natrag izravno

Množenje s nulama možemo izbjeći akumuliranjem doprinosa izlaza:

- višestruko pisanje na istu lokaciju je loše za paralelnu izvedbu, ali ovdje može biti metoda izbora



Ovakva operacija ne može se izraziti konvolucijom

- $\Rightarrow$ izvodimo je posebnim algoritmom

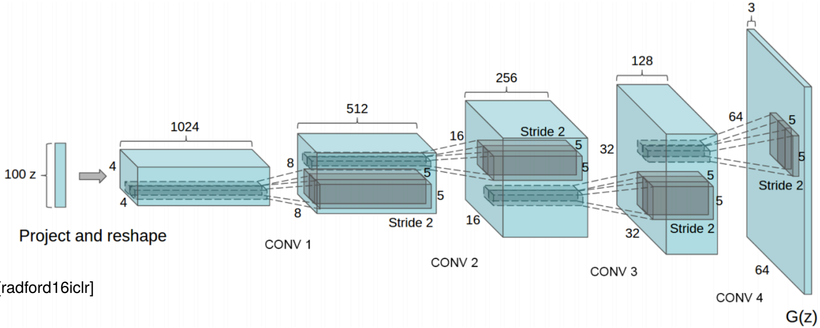
$$\partial\mathcal{L}/\partial\mathbf{p}^{(f)} = \sum_{ij}\sum_{g}\partial\mathcal{L}/\partial q_{ij}^{(g)} \cdot \partial q_{ij}^{(g)}/\partial\mathbf{p}^{(f)}$$

$$= \sum_{ij=0,0}^{shape(\mathbf{q})}\sum_{g}\mathrm{embed}_{k\times k}^{H\times W}(\partial\mathcal{L}/\partial q_{ij}^{(g)} \cdot \mathbf{w}^{(g,f)}, [s\cdot i, s\cdot j])$$

KORAK: NATRAG IZRAVNO (2)

Transponirana konvolucija s korakom u unaprijednom prolazu:

- standardni štos za povećati rezoluciju reprezentacije
- za raspoznavanje malih slika (npr. CIFAR)
- kod generativnih modela s niskodimenzionalnim latentnim prostorom (npr. DCGAN, VAE)



[radford16iclr]

Korak: alternativa

Rezoluciju latentne reprezentacije možemo povećati i bilinearnim naduzorkovanjem nakon čega tipično ide normalna konvolucija:

- učinak je vrlo sličan transponiranoj konvoluciji s korakom
- razlika: rupe u naduzorkovanom ulazu pune se intepoliranim značajkama (umjesto nulama)
- veća složenost ali ista brzina u praksi
- nama je takav postupak radio bolje [kreso17iccvw]

Sadržaj

- ponavljanje: unatražno učenje, potpuno povezani slojevi
- gradijenti parametara 1D konvolucije
  - jezgra dimenzije 1
  - jezgra dimenzije k
- gradijenti parametara 2D konvolucije
  - jezgra dimenzija k×k
  - zadatak s unatražnim prolazom kroz 2D konvoluciju
- učenje konvolucije s korakom
- 2D konvolucija nad tenzorima trećeg reda
- **efikasna implementacija 2D konvolucije**

Implementacija: uvod

Ideja: predstaviti konvoluciju kao matrični umnožak

- kapitalizirati GEMM za efikasnu evaluaciju

Dvije izvedbe ideje:

- izravnati ulazne i izlazne aktivacije (loša ideja)
- izravnati težine konvolucijske jezgre (dobra ideja)

Novije implementacije koriste Winogradov konvolucijski algoritam

- ukupni broj množenja više nego dvostruko manji
- Nervana [lavin15arxiv], cudnn v5.1+

Implementacija: ravnanje aktivacija

$\mathbf{q}_{3\times3} = \mathbf{w}_{3\times3} \star \mathbf{p}_{5\times5}$

$\mathrm{vec}(\mathbf{q}_{3\times3}) = \mathrm{doubly\_circ}(\mathbf{w}_{3\times3}) \cdot \mathrm{vec}(\mathbf{p}_{5\times5})$

$$
\begin{bmatrix} q_{11} \\ q_{12} \\ q_{13} \\ q_{21} \\ q_{22} \\ q_{23} \\ q_{31} \\ q_{32} \\ q_{33} \end{bmatrix}
=
\begin{bmatrix}
w_{11} & w_{12} & w_{13} & \cdot & \cdot & w_{21} & w_{22} & w_{23} & \cdot & \cdot & w_{31} & w_{32} & w_{33} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & w_{11} & w_{12} & w_{13} & \cdot & \cdot & w_{21} & w_{22} & w_{23} & \cdot & \cdot & w_{31} & w_{32} & w_{33} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & w_{11} & w_{12} & w_{13} & \cdot & \cdot & w_{21} & w_{22} & w_{23} & \cdot & \cdot & w_{31} & w_{32} & w_{33} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & w_{11} & w_{12} & w_{13} & \cdot & \cdot & w_{21} & w_{22} & w_{23} & \cdot & \cdot & w_{31} & w_{32} & w_{33} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & w_{11} & w_{12} & w_{13} & \cdot & \cdot & w_{21} & w_{22} & w_{23} & \cdot & \cdot & w_{31} & w_{32} & w_{33} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & w_{11} & w_{12} & w_{13} & \cdot & \cdot & w_{21} & w_{22} & w_{23} & \cdot & \cdot & w_{31} & w_{32} & w_{33} & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & w_{11} & w_{12} & w_{13} & \cdot & \cdot & w_{21} & w_{22} & w_{23} & \cdot & \cdot & w_{31} & w_{32} & w_{33} & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & w_{11} & w_{12} & w_{13} & \cdot & \cdot & w_{21} & w_{22} & w_{23} & \cdot & \cdot & w_{31} & w_{32} & w_{33} & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & w_{11} & w_{12} & w_{13} & \cdot & \cdot & w_{21} & w_{22} & w_{23} & \cdot & \cdot & w_{31} & w_{32} & w_{33}
\end{bmatrix}
\begin{bmatrix} p_{11} \\ p_{12} \\ p_{13} \\ p_{14} \\ p_{15} \\ p_{21} \\ p_{22} \\ p_{23} \\ p_{24} \\ p_{25} \\ p_{31} \\ p_{32} \\ p_{33} \\ p_{34} \\ p_{35} \\ p_{41} \\ p_{42} \\ p_{43} \\ p_{44} \\ p_{51} \\ p_{52} \\ p_{53} \\ p_{54} \\ p_{55} \end{bmatrix}
$$

Vremenska složenost: O(W^2H^2) vs O(WHk^2)

Implementacija: ravnanje jezgri

Ideja: presložiti ulaz da konvolucija odgovara matričnom množenju:

$$\operatorname{vec}(\mathbf{q}) = \operatorname{im2row}(\mathbf{p}) \cdot \operatorname{vec}(\mathbf{w})$$

Na slici dolje $\mathbf{X}$ odgovara $\mathbf{p}$, dok $\mathbf{X}^{\mathrm{ROWS}}$ odgovara $\operatorname{im2row}(\mathbf{p})$:

$\boldsymbol{X}$

x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
x_{21}	x_{22}	x_{23}	x_{24}	x_{25}
x_{31}	x_{32}	x_{33}	x_{34}	x_{35}
x_{41}	x_{42}	x_{43}	x_{44}	x_{45}
x_{51}	x_{52}	x_{53}	x_{54}	x_{55}

$\boldsymbol{X}^{\mathrm{ROWS}}$

x_{11}	x_{12}	x_{13}	x_{21}	x_{22}	x_{23}	x_{31}	x_{32}	x_{33}
x_{12}	x_{13}	x_{14}	x_{22}	x_{23}	x_{24}	x_{32}	x_{33}	x_{34}
x_{13}	x_{14}	x_{15}	x_{23}	x_{24}	x_{25}	x_{33}	x_{34}	x_{35}
x_{33}	x_{34}	x_{35}	x_{43}	x_{44}	x_{45}	x_{53}	x_{54}	x_{55}

Vremenska složenost: O(WHk^2) vs O(WHk^2)

Prostorna složenost: O(WHk^2) vs O(WH)

Implementacija: ravnanje jezgri

Ideja: presložiti ulaz da konvolucija odgovara matričnom množenju:

$$\mathrm{vec}(\mathbf{q}) = \mathrm{im2row}(\mathbf{p}) \cdot \mathrm{vec}(\mathbf{w})$$

Na slici dolje $\mathbf{X}$ odgovara $\mathbf{p}$, dok $\mathbf{X}^{\mathrm{ROWS}}$ odgovara $\mathrm{im2row}(\mathbf{p})$:

$\boldsymbol{X}$

x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
x_{21}	x_{22}	x_{23}	x_{24}	x_{25}
x_{31}	x_{32}	x_{33}	x_{34}	x_{35}
x_{41}	x_{42}	x_{43}	x_{44}	x_{45}
x_{51}	x_{52}	x_{53}	x_{54}	x_{55}

$\boldsymbol{X}^{\mathrm{ROWS}}$

x_{11}	x_{12}	x_{13}	x_{21}	x_{22}	x_{23}	x_{31}	x_{32}	x_{33}
x_{12}	x_{13}	x_{14}	x_{22}	x_{23}	x_{24}	x_{32}	x_{33}	x_{34}
x_{13}	x_{14}	x_{15}	x_{23}	x_{24}	x_{25}	x_{33}	x_{34}	x_{35}
x_{33}	x_{34}	x_{35}	x_{43}	x_{44}	x_{45}	x_{53}	x_{54}	x_{55}

Vremenska složenost: O(WHk^2) vs O(WHk^2)

Prostorna složenost: O(WHk^2) vs O(WH)

Implementacija: ravnanje jezgri

Ideja: presložiti ulaz da konvolucija odgovara matričnom množenju:

$$\mathrm{vec}(\mathbf{q}) = \mathrm{im2row}(\mathbf{p}) \cdot \mathrm{vec}(\mathbf{w})$$

Na slici dolje $\mathbf{X}$ odgovara $\mathbf{p}$, dok $\mathbf{X}^{\mathrm{ROWS}}$ odgovara $\mathrm{im2row}(\mathbf{p})$:

$\boldsymbol{X}$

x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
x_{21}	x_{22}	x_{23}	x_{24}	x_{25}
x_{31}	x_{32}	x_{33}	x_{34}	x_{35}
x_{41}	x_{42}	x_{43}	x_{44}	x_{45}
x_{51}	x_{52}	x_{53}	x_{54}	x_{55}

$\boldsymbol{X}^{\mathrm{ROWS}}$

x_{11}	x_{12}	x_{13}	x_{21}	x_{22}	x_{23}	x_{31}	x_{32}	x_{33}
x_{12}	x_{13}	x_{14}	x_{22}	x_{23}	x_{24}	x_{32}	x_{33}	x_{34}
x_{13}	x_{14}	x_{15}	x_{23}	x_{24}	x_{25}	x_{33}	x_{34}	x_{35}
x_{33}	x_{34}	x_{35}	x_{43}	x_{44}	x_{45}	x_{53}	x_{54}	x_{55}

Vremenska složenost: O(WHk^2) vs O(WHk^2)

Prostorna složenost: O(WHk^2) vs O(WH)

Implementacija: ravnanje jezgri

Ideja: presložiti ulaz da konvolucija odgovara matričnom množenju:

$$\mathrm{vec}(\mathbf{q}) = \mathrm{im2row}(\mathbf{p}) \cdot \mathrm{vec}(\mathbf{w})$$

Na slici dolje $\mathbf{X}$ odgovara $\mathbf{p}$, dok $\mathbf{X}^{\mathrm{ROWS}}$ odgovara $\mathrm{im2row}(\mathbf{p})$:

$\boldsymbol{X}$

x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
x_{21}	x_{22}	x_{23}	x_{24}	x_{25}
x_{31}	x_{32}	x_{33}	x_{34}	x_{35}
x_{41}	x_{42}	x_{43}	x_{44}	x_{45}
x_{51}	x_{52}	x_{53}	x_{54}	x_{55}

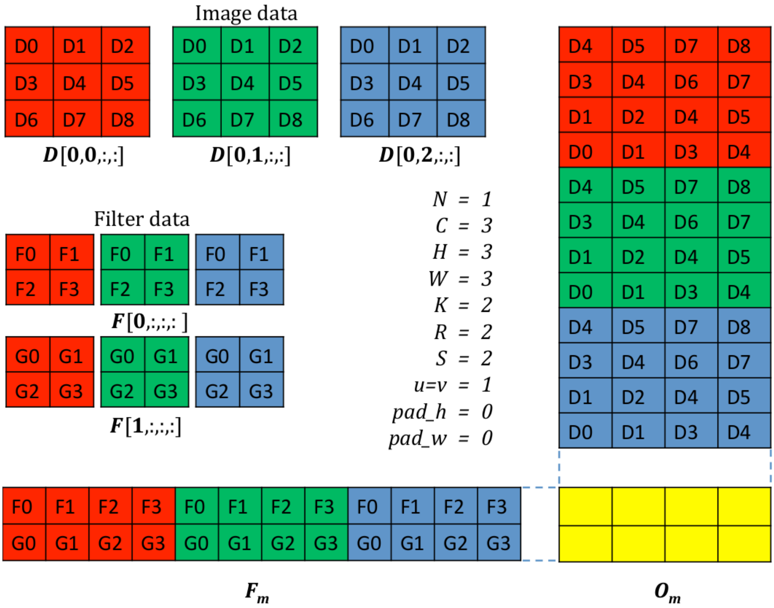
$\boldsymbol{X}^{\mathrm{ROWS}}$

x_{11}	x_{12}	x_{13}	x_{21}	x_{22}	x_{23}	x_{31}	x_{32}	x_{33}
x_{12}	x_{13}	x_{14}	x_{22}	x_{23}	x_{24}	x_{32}	x_{33}	x_{34}
x_{13}	x_{14}	x_{15}	x_{23}	x_{24}	x_{25}	x_{33}	x_{34}	x_{35}
x_{33}	x_{34}	x_{35}	x_{43}	x_{44}	x_{45}	x_{53}	x_{54}	x_{55}

Vremenska složenost: O(WHk^2) vs O(WHk^2)

Prostorna složenost: O(WHk^2) vs O(WH)

Implementacija: ravnanje jezgri (tenzori 3. reda)



[chetlur14arxiv]

Implementacija: natrag - parametri

Podsjetimo se, konvoluciju smo izrazili matričnim produktom:

$$\mathrm{vec}(\mathbf{q}) = \mathrm{im2row}(\mathbf{p}) \cdot \mathrm{vec}(\mathbf{w})$$

Odatle se jasno vidi da mora biti:

$$\frac{\partial \mathrm{vec}(\mathbf{q})}{\partial \mathrm{vec}(\mathbf{w})} = \mathrm{im2row}(\mathbf{p})$$

Konačno, gradijente funkcije gubitka s obzirom na parametre (konvolucijsku jezgru) računamo kao:

$$\begin{aligned} \partial \mathcal{L} / \partial \mathrm{vec}(\mathbf{w}) &= \partial \mathcal{L} / \partial \mathrm{vec}(\mathbf{q}) \cdot \partial \mathrm{vec}(\mathbf{q}) / \partial \mathrm{vec}(\mathbf{w}) \\ &= \partial \mathcal{L} / \partial \mathrm{vec}(\mathbf{q}) \cdot \mathrm{im2row}(\mathbf{p}) \end{aligned}$$

Implementacija: natrag - ulazi

Podsjetimo se kako izgledaju gradijenti po ulazu:

$$\partial\mathcal{L}/\partial\mathbf{p} = \mathrm{flip2d}(\mathbf{w}) \star \mathrm{pad}(\partial\mathcal{L}/\partial\mathbf{q}, \lfloor k/2 \rfloor)$$

Gradijente izlaza treba presložiti u konvolucijsku matricu $\mathbf{G}_{\mathbf{q}}^{\mathrm{ROWS}}$:

$$\mathbf{G}_{\mathbf{q}}^{\mathrm{ROWS}} = \mathrm{im2row}(\mathrm{pad}(\partial\mathcal{L}/\partial\mathbf{q}, \lfloor k/2 \rfloor))$$

Konvoluciju sada možemo provesti matričnim umnoškom:

$$\mathbf{G}_{\mathbf{p}}^{\mathrm{VEC}} = \mathbf{G}_{\mathbf{q}}^{\mathrm{ROWS}} \cdot \mathrm{vec}(\mathrm{flip2d}(\mathbf{w}))$$

Konačni rezultat dobivamo preoblikovanjem $\mathbf{G}_{\mathbf{p}}^{\mathrm{VEC}}$:

$$\partial\mathcal{L}/\partial\mathbf{p} = \mathrm{reshape}(\mathbf{G}_{\mathbf{p}}^{\mathrm{VEC}}, [H, W])$$

Ako smo u unaprijednom prolazu imali izlazni korak:

- $\mathbf{G}_{\mathbf{q}}^{\mathrm{ROWS}}$ je rijetka matrica
- bolju performansu postiže izravna transponirana konvolucija

ZAHVALA

Ova predavanja proizišla su iz istraživanja koje je financirala Hrvatska zaklada za znanost projektom I-2433-2014 MultiCLoD.

`http://multiclod.zemris.fer.hr`