Zadatak 1.
Odrediti strukturu, jednadžbe i ukupan broj parametara potpuno povezanog modela za 2D podatke ako znamo da su dimenzije slojeva: 5, 10, 5, 2.

Rješenje. Svaki Θ_i se sastoji od $\mathbf{W}_i$ i $\mathbf{b}_i$. Ako pretpostavimo da koristimo samo jedan primjer, struktura izgleda ovako:

Ulaz:

$$\mathbf{x} - D_{podatak} \times D_{br\_primjera} = 2 \times 1$$

1. sloj:

$$\begin{aligned}\mathbf{W}_1 - D_{podatak} \times D_1 = 5 \times 2\\ \mathbf{b}_1 - D_1 \times 1 = 5 \times 1\\ \mathbf{h}_1 - D_1 \times 1 = 5 \times 1\end{aligned}$$

2. sloj:

$$\begin{aligned}\mathbf{W}_2 - D_1 \times D_2 = 10 \times 5\\ \mathbf{b}_2 - D_2 \times 1 = 10 \times 1\\ \mathbf{h}_2 - D_2 \times 1 = 10 \times 1\end{aligned}$$

3. sloj:

$$\begin{aligned}\mathbf{W}_3 - D_2 \times D_3 = 5 \times 10\\ \mathbf{b}_3 - D_3 \times 1 = 5 \times 1\\ \mathbf{h}_3 - D_3 \times 1 = 5 \times 1\end{aligned}$$

4. sloj:

$$\begin{aligned}\mathbf{W}_4 - D_3 \times D_4 = 2 \times 5\\ \mathbf{b}_4 - D_4 \times 1 = 2 \times 1\\ \mathbf{o} - D_4 \times 1 = 2 \times 1\end{aligned}$$

Iz ove strukture se zatim računaju parametri prema dimenzija $\mathbf{W}_i$ i $\mathbf{b}_i$:

$$(2 \cdot 5 + 5 \cdot 1) + (5 \cdot 10 + 10 \cdot 1) + (10 \cdot 5 + 5 \cdot 1) + (5 \cdot 2 + 2 \cdot 1) = 142$$

Konačno, još je potrebno napisati jednadžbe koje su samo afine transformacije nakon kojih slijedi nelinearna aktivacija. One redom glase:

$$\mathbf{h}_1 = ReLU(\mathbf{W}_1\mathbf{x} + \mathbf{b}_1)$$
$$\mathbf{h}_2 = ReLU(\mathbf{W}_2\mathbf{h}_1 + \mathbf{b}_2)$$
$$\mathbf{h}_3 = ReLU(\mathbf{W}_3\mathbf{h}_2 + \mathbf{b}_3)$$
$$\mathbf{o} = softmax(\mathbf{W}_4\mathbf{h}_3 + \mathbf{b}_4)$$

Zadatak 2.
Za domaći rad dokazati:

$$\frac{dL_{CE}(\mathbf{y}, softmax(\mathbf{z}))}{dz_i} = softmax(z_i) - [\![y = i]\!]$$

U ovom izrazu y predstavlja indeks točnog razreda.

Rješenje. Izvedimo prvo gubitak pomoću negativne log izglednosti. Pretpostavljamo da su oznake $\mathbf{y}$ one hot enkodirane i da se ravnaju prema multinulijevoj razdiobi. Zbog toga izglednost možemo raspisati kao $\prod_{k=1}^{K} p_k^{y_k}$.

$$\begin{aligned}
L_{CE}(\mathbf{y}, softmax(\mathbf{z})) &= -\sum_{k=1}^{K} \log(p_k^{y_k}) \\
&= -\sum_{k=1}^{K} y_k \log \frac{\exp(z_k)}{\sum_j \exp(z_j)} \\
&= \sum_{k=1}^{K} y_k \Big(-z_k + \log \sum_j \exp(z_j) \Big) \\
&= -\sum_{k=1}^{K} y_k z_k + \sum_{k=1}^{K} y_k \log \sum_j \exp(z_j) \\
&= -z_y + \log \sum_j \exp(z_j)
\end{aligned}$$

Sada možemo derivirati izraz po z_i.

$$\begin{aligned}
\frac{dL_{CE}(\mathbf{y}, softmax(\mathbf{z}))}{dz_i} &= -[\![y = i]\!] + \frac{1}{\sum_j \exp(z_j)} \frac{d \sum_j \exp(z_j)}{dz_i} \\
&= -[\![y = i]\!] + \frac{1}{\sum_j \exp(z_j)} \frac{d \exp(z_i)}{dz_i} \\
&= -[\![y = i]\!] + \frac{\exp(z_i)}{\sum_j \exp(z_j)} \\
&= -[\![y = i]\!] + softmax(z_i)
\end{aligned}$$

Primijetimo da će kod posljednje jednakosti pri računanju funkcije gubitka varijabla y_k biti jednaka 1 samo kada vrijedi $k = y$ pa zato nestaju obje sume.

Zadatak 3.
Za domaći rad dokazati:

$$\frac{dL_{CE}(y, \sigma(z))}{dz} = \sigma(z) - y$$

Rješenje. Također krenimo s gubitkom kao i u prethodnom zadatku. Znamo da se oznaka y ravna po Bernoullijevoj razdiobi pa možemo izglednost zapisati kao $\sigma(z)^y\,(1-\sigma(z))^{1-y}$, nakon čega primijenjujemo log funkciju na umnožak.

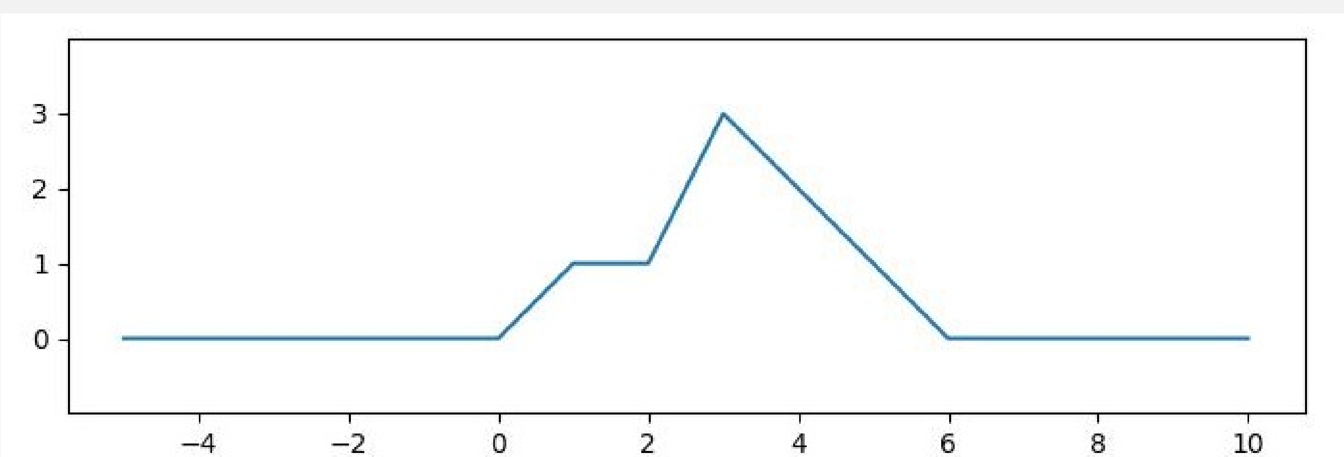
$$\begin{aligned} L_{CE}(y, \sigma(z)) &= -log\big(\sigma(z)^y\,(1-\sigma(z)^{1-y}\big) \\ &= -y\sigma(z) - (1-y)\,log(1-\sigma(z)) \end{aligned}$$

Deriviramo gubitak po z:

$$\begin{aligned} \frac{dL_{CE}(y, \sigma(z))}{dz} &= \left(-y \cdot \frac{1}{\sigma(z)} + (1-y)\frac{1}{1-\sigma(z)}\right)\sigma(z)(1-\sigma(z)) \\ &= -y\,(1-\sigma(z)) + (1-y)\,\sigma(z) \\ &= -y + y\sigma(z) + \sigma(z) - y\sigma(z) \\ &= \sigma(z) - y \end{aligned}$$

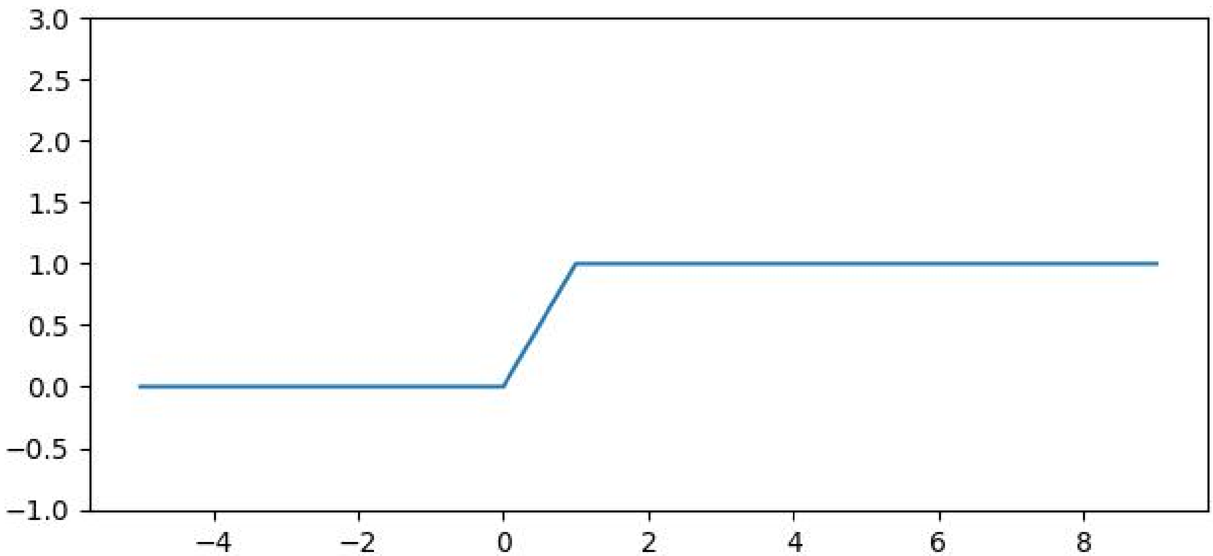
Ovdje valja napomenuti da je ovo identično prethodnom zadatku, ali za 2 klase. Drugim riječima, ovaj zadatak je poseban slučaj prethodnog.

Zadatak 4.
Konstruirajte dvorazinski model koji koristi linearne transformacije i zglobnicu da bi aproksimirao sljedeću funkciju:



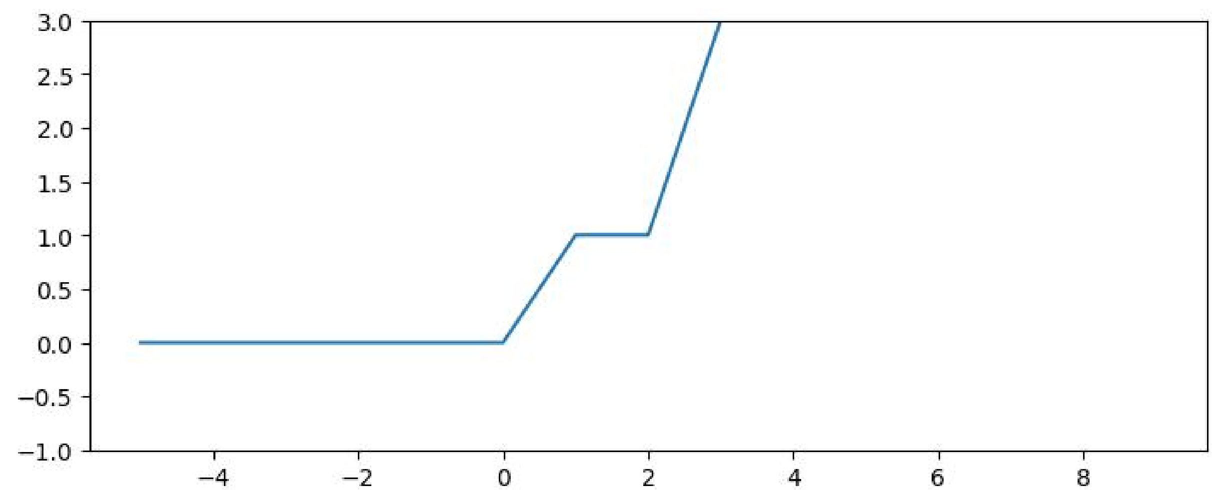
Rješenje. Krećemo sa segmentom $[0, 2]$. Ovaj dio linearno raste i to od $x = 0$ pa će prva funkcija koju koristimo biti $y_1 = \max(0, x)$. Nakon toga nam je potrebna funkcija koja će poništiti rastuću funkciju y_1 što znači da će ispred nje ići negativan predznak. Ona ima pomak u točki $x = 1$ pa ju je moguće definirati kao $y_2 = \max(0, x - 1)$.

Ove funkcije ćemo kombinirati kao $f = y_1 - y_2$ i dobiti sljedeći rezultat:



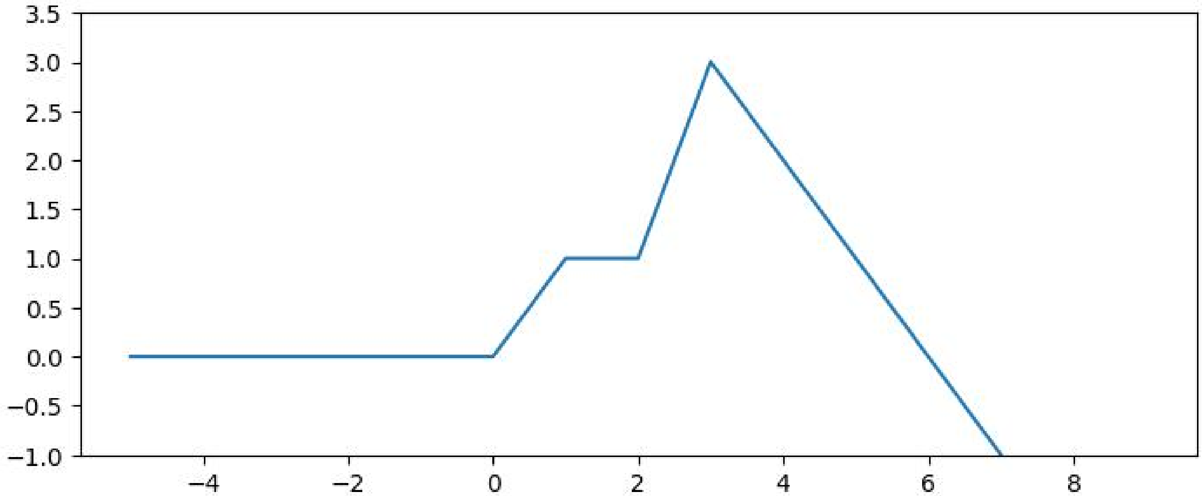
Slijedi segment $[2, 3]$. Ovdje funkcija duplo brže raste i tu komponentu je potrebno pomaknuti točku $x = 2$. Tu možemo definirati $y_3 = \max(0, 2(x - 2))$.

Sada kombinacijom $f = y_1 - y_2 + y_3$ dobivamo sljedeći rezultat:



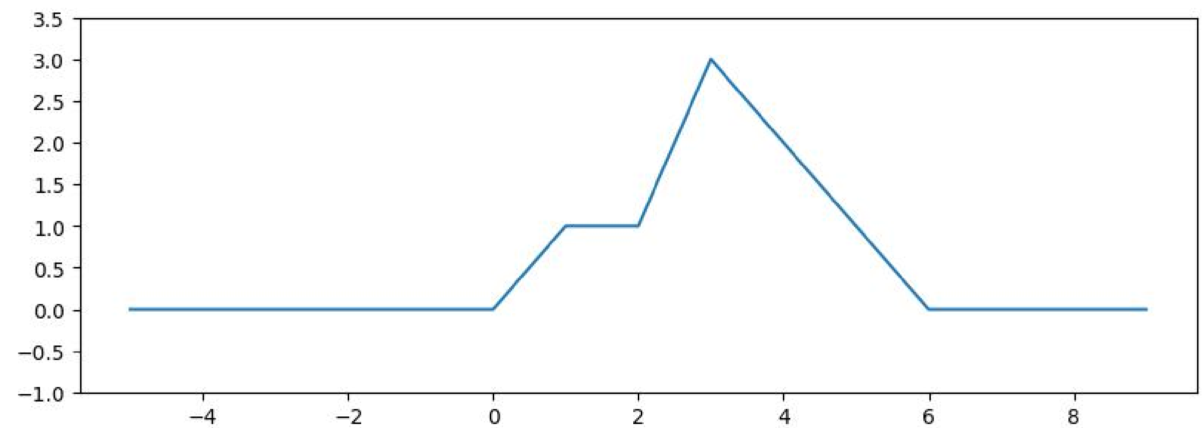
U segmentu $[3, 6]$ imamo padajući dio funkcije. Moramo poništiti utjecaj prethodne komponente funkcije i pobrinuti se za negativni nagib sljedeće komponente stavljanjem negativnog predznaka. Uzeti ćemo koeficijent nagiba 3, gdje 2 dolazi zbog poništavanja prethodne komponente i 1 zbog linearnog opadanja. To možemo ostvariti funkcijom $y_4 = \max(0, 3(x - 3))$.

Kombinacijom $f = y_1 - y_2 + y_3 - y_4$ dobivamo sljedeći rezultat:



I konačno na zadnjem segmentu je potrebno samo poništiti utjecaj prethodne komponente od točke $x = 6$ za što ćemo koristiti funkciju $y_5 = \max(0, x - 6)$. Ova komponenta će imati pozitivan predznak jer želimo poništiti utjecaj komponente s negativnim nagibom.

Konačno rješenje dobivamo kombinacijom $f = y_1 - y_2 + y_3 - y_4 + y_5$.



Zadatak 5.
Pokažite da je pristup računanja gradijenata gubitka po parametrima i po vektoriziranim parametrima ekvivalentan:

$$\frac{\partial L}{\partial vec(\mathbf{W}_k)} = \frac{\partial L}{\partial \mathbf{h}_k} \cdot \frac{\partial \mathbf{h}_k}{\partial vec(\mathbf{W}_k)}$$

$$\frac{\partial L}{\partial \mathbf{W}_k} = \left(\frac{\partial L}{\partial w_{k_{ij}}} \right) = \left[\frac{\partial L}{\partial \mathbf{h}_k} \right]^T \cdot \mathbf{h}_{k-1}^T$$

i objasnite njihovu složenost.

Rješenje. Možemo krenuti od prvog izraza. Zbog jednostavnosti s notacijom, pretpostavimo da računamo vrijednost vektora:

$$\mathbf{q} = \mathbf{W} \cdot \mathbf{p} + \mathbf{b}$$

Dimenzije ovih matrica i vektora su redom:

$$(n \times 1) = (n \times m) \cdot (m \times 1) + (n \times 1)$$

U matricama će to ovako izgledati:

$$\begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \end{bmatrix} = \begin{bmatrix} w_{11} \, w_{12} \, \dots \, w_{1m} \\ w_{21} \, w_{22} \, \dots \, w_{2m} \\ \vdots \\ w_{n1} \, w_{n2} \, \dots \, w_{nm} \end{bmatrix} \cdot \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_m \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

Kod ovakvih zadataka najlakši pristup je raspisati kako se računaju zasebni dijelovi vektora $\mathbf{q}$.

$$\begin{aligned} q_1 &= w_{11} \cdot p_1 + w_{12} \cdot p_2 + \ldots + w_{1m} \cdot p_m + b_1 \\ q_2 &= w_{21} \cdot p_1 + w_{22} \cdot p_2 + \ldots + w_{2m} \cdot p_m + b_2 \\ &\vdots \\ q_n &= w_{n1} \cdot p_1 + w_{n2} \cdot p_2 + \ldots + w_{nm} \cdot p_m + b_n \end{aligned}$$

Jakobijan $\frac{\partial L}{\partial vec(\mathbf{W}_k)}$ će izgledati ovako:

$$\begin{aligned} \frac{\partial L}{\partial vec(\mathbf{W}_k)} &= \frac{\partial L}{\partial \mathbf{q}} \cdot \frac{\partial \mathbf{q}}{\partial vec(\mathbf{W})} \\ &= \frac{\partial L}{\partial \mathbf{q}} \cdot \begin{bmatrix} \frac{\partial q_1}{\partial w_{11}} \, \frac{\partial q_1}{\partial w_{12}} \cdots \frac{\partial q_1}{\partial w_{1m}} \cdots \frac{\partial q_1}{\partial w_{n1}} \, \frac{\partial q_1}{\partial w_{n2}} \cdots \frac{\partial q_1}{\partial w_{nm}} \\ \frac{\partial q_2}{\partial w_{11}} \, \frac{\partial q_2}{\partial w_{12}} \cdots \frac{\partial q_2}{\partial w_{1m}} \cdots \frac{\partial q_2}{\partial w_{n1}} \, \frac{\partial q_2}{\partial w_{n2}} \cdots \frac{\partial q_2}{\partial w_{nm}} \\ \vdots \\ \frac{\partial q_n}{\partial w_{11}} \, \frac{\partial q_n}{\partial w_{12}} \cdots \frac{\partial q_n}{\partial w_{1m}} \cdots \frac{\partial q_n}{\partial w_{n1}} \, \frac{\partial q_n}{\partial w_{n2}} \cdots \frac{\partial q_n}{\partial w_{nm}} \end{bmatrix} \end{aligned}$$

Većina derivacija će biti jednaka nuli. Derivacija neće biti jednaka nuli kada u izrazu $\frac{\partial q_k}{\partial w_{ij}}$ vrijedi da je k jednak i:

$$\frac{\partial q_1}{\partial w_{11}} = p_1, \; \frac{\partial q_1}{\partial w_{12}} = p_2, \; \ldots \;, \frac{\partial q_1}{\partial w_{1m}} = p_m$$

$$\frac{\partial q_2}{\partial w_{21}} = p_1, \; \frac{\partial q_2}{\partial w_{22}} = p_2, \; \ldots \;, \frac{\partial q_2}{\partial w_{2m}} = p_m$$

$$\frac{\partial q_n}{\partial w_{n1}} = p_1, \; \frac{\partial q_n}{\partial w_{n2}} = p_2, \; \ldots \;, \frac{\partial q_n}{\partial w_{nm}} = p_m$$

Iz ovoga se može zaključiti kako će izgledati prethodno navedena matrica:

$$\frac{\partial \mathbf{q}}{\partial vec(\mathbf{W})} = \begin{bmatrix} p_1 & p_2 & \dots & p_m & 0 & 0 & \dots 0 & \dots & 0 & 0 \dots 0 \\ 0 & 0 & \dots & 0 & p_1 & p_2 & \dots p_m & \dots & 0 & 0 \dots 0 \\ & & & & & \vdots & & & & \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots 0 & \dots & p_1 & p_2 \dots p_m \end{bmatrix}$$

Može se primijetiti da je ovakva matrica vrlo rijetka i to će se odraziti u složenosti. Vektor $\frac{\partial L}{\partial \mathbf{q}}$ će biti dimenzija $1 \times n$, a matrica $\frac{\partial \mathbf{q}}{\partial vec(\mathbf{W})}$ dimenzija $n \times n \cdot m$. To znači da će složenost biti $O(n^2 \cdot m)$. Sad možemo promotriti čemu će biti jednak gradijent $\frac{\partial L}{\partial vec(\mathbf{W})}$.

$$\begin{bmatrix} \frac{\partial L}{\partial q_1} \frac{\partial L}{\partial q_2} \dots \frac{\partial L}{\partial q_n} \end{bmatrix} \cdot \begin{bmatrix} p_1 & p_2 & \dots & p_m & 0 & 0 & \dots 0 & \dots & 0 & 0 \dots 0 \\ 0 & 0 & \dots & 0 & p_1 & p_2 & \dots p_m & \dots & 0 & 0 \dots 0 \\ & & & & & \vdots & & & & \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots 0 & \dots & p_1 & p_2 \dots p_m \end{bmatrix}$$

$$= \begin{bmatrix} p_1 \cdot \frac{\partial L}{\partial q_1} & p_2 \cdot \frac{\partial L}{\partial q_1} & \dots & p_m \cdot \frac{\partial L}{\partial q_1} & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & p_1 \cdot \frac{\partial L}{\partial q_2} & p_2 \cdot \frac{\partial L}{\partial q_2} & \dots & p_m \cdot \frac{\partial L}{\partial q_2} & \dots & 0 & 0 & \dots & 0 \\ & & & & & \vdots & & & & & & & \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & p_1 \cdot \frac{\partial L}{\partial q_n} & p_2 \cdot \frac{\partial L}{\partial q_n} & \dots & p_m \cdot \frac{\partial L}{\partial q_n} \end{bmatrix}$$

Ako želimo izračunati $\frac{\partial L}{\partial \mathbf{W}}$ onda to možemo napraviti preslaganjem u početni oblik i primijetiti da se to može kompaktno zapisati u obliku vanjskog produkta:

$$\begin{bmatrix} p_1 \cdot \frac{\partial L}{\partial q_1} & p_2 \cdot \frac{\partial L}{\partial q_1} & \dots & p_m \cdot \frac{\partial L}{\partial q_1} \\ p_1 \cdot \frac{\partial L}{\partial q_1} & p_2 \cdot \frac{\partial L}{\partial q_1} & \dots & p_m \cdot \frac{\partial L}{\partial q_1} \\ \vdots & & & \\ p_1 \cdot \frac{\partial L}{\partial q_n} & p_2 \cdot \frac{\partial L}{\partial q_n} & \dots & p_m \cdot \frac{\partial L}{\partial q_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial L}{\partial q_1} \\ \frac{\partial L}{\partial q_2} \\ \vdots \\ \frac{\partial L}{\partial q_n} \end{bmatrix} \cdot \begin{bmatrix} p_1 \, p_2 \dots p_m \end{bmatrix} = \left[\frac{\partial L}{\partial \mathbf{q}}\right]^T \cdot \mathbf{p}^T$$

Ovaj pristup je bolji jer sada množimo vektore dimenzija $n \times 1$ i $1 \times m$ pa će složenost biti $O(n \cdot m)$.

Zadatak 6.

Razmatramo klasifikacijski konvolucijski model za sive slike 28×28. Arhitektura modela je:

- dva konvolucijska sloja bez nadopunjavanja (1. sloj - 16 kanala, 2. sloj - 32 kanala)
- jezgra 5×5 s pomakom
- aktivacija ReLU
- sažimanje maksimumom 2×2
- potpuno povezani sloj dimenzije 512 s pomakom + ReLU
- potpuno povezani sloj $D = 10$ s pomakom + softmax

Zadatci:

- Odredite dimenzije aktivacija, broj parametara i veličinu receptivnog polja u svim slojevima
- Natipkajte implementaciju u PyTorchu s metodama `__init__`, `fwd` i `loss`

Rješenje. Dimenzija aktivacija, broj parametara i veličina receptivnog polja su prikazane u tablici.

Sloj	Parametri	Tenzor	Receptivno polje
Slika	-	$1 \times 28 \times 28$	1×1
Conv1	$16 \cdot 1 \cdot 5 \cdot 5 + 16$	$16 \times 24 \times 24$	5×5
Relu1	-	$16 \times 24 \times 24$	5×5
Pool1	-	$16 \times 12 \times 12$	6×6
Conv2	$32 \cdot 16 \cdot 5 \cdot 5 + 32$	$32 \times 8 \times 8$	14×14
Relu2	-	$32 \times 8 \times 8$	14×14
Pool2	-	$32 \times 4 \times 4$	16×16
Flatten	-	512	16×16
FC1	$512 \cdot 512 + 512$	512	28×28
Relu3	-	512	28×28
FC2	$512 \cdot 10 + 10$	10	28×28
softmax	-	10	28×28

Na ulazu je slika dimenzija $1 \times 28 \times 28$ pa su to i dimenzije tenzora, a receptivno polje sadržava samo jedan element.

Za svaki konvolucijski sloj broj parametara će odgovarati $in\_channels \cdot out\_channels \cdot F \cdot F$. To je zato što nam je potrebna jezgra za svaku kombinaciju ulaznog i izlaznog kanala. Osim toga trebati će nam i pomak za svaki od izlaznih kanala tj. $out\_channels$. Tako za prvi konvolucijski sloj broj parametara je jednak $16 \cdot 1 \cdot 5 \cdot 5 + 16$.

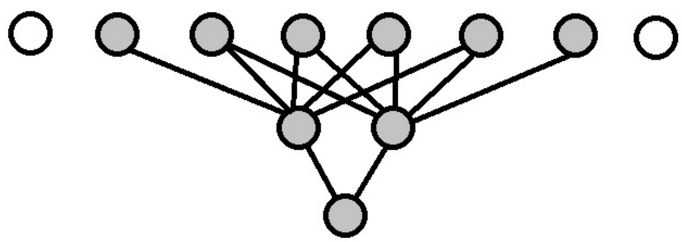
Kod konvolucije bez nadopunjavanja dimenzije tenzora će se smanjiti prema formuli $W - F + 1$ gdje je W širina ili visina tenzora (ovo dvoje će obično biti jednako). Zbog toga 2. i 3. dimenzija tenzora se mijenjaju u $28 - 5 + 1 = 24$.

Receptivno polje je najlakše odrediti gledamo li "unatrag". U prvom konvolucijskom sloju svaki piksel je izračunat na temelju 5 piksela u originalnoj slici.

ReLU aktivacijska funkcija nikako ne utječe na parametre, dimenzije, tenzora i receptivno polje.

Sljedeće je na redu sažimanje maksimumom. Ono nema nikakve parametre. Na veličinu tenzora utječe tako da u svakoj 2×2 regiji odabere maksimalnu vrijednost u svakom kanalu. Na taj način će se dimenzije tenzora prepoloviti.

Receptivno polje je ovdje kompliciranije odrediti. Lakše je ako se problem svede na 1D konvoluciju. Krenuvši unatrag jedan element računamo na temelju maksimuma 2 elementa sloja sažimanja. Svaki od tih slojeva sažimanja računamo na temelju 5 elemenata, ali oni imaju preklapanja, pa je na kraju samo 6 različitih elemenata.



Na isti način se sve operacija provode i za drugi konvolucijski sloj i sažimanje.

Kod Flatten sloja dimenzije tenzora su jednake umnošku elemenata u svakoj dimenziji iz prethodnog sloja, ali ovaj sloj nema nikakvih parametara. Kod potpuno povezanog sloja potrebne su težine za svaku kombinaciju ulaza i izlaza te pomak. Broj parametara ovdje će zato glasiti $512 \cdot 512 + 512$. Pošto je svaki element povezan sa svakim elementom iz prethodnog sloja receptivno polje će upravo odgovarati dimenzijama ulazne slike. Na isti način se određuju parametri za sljedeći potpuno povezani sloj.

U kodu bi to izgledalo ovako:

```python
class Model(nn.Module):
    def __init__(self):
        super().__init__()
        self.conv1 = nn.Conv2d(1, 16, kernel_size=5)
        self.conv2 = nn.Conv2d(16, 32, kernel_size=5)
        self.pool = nn.MaxPool2d(kernel_size=2, strides=2)
        self.l1 = nn.Linear(512, 512)
        self.l2 = nn.Linear(512, 10)

    def forward(self, x):
        x = F.relu(self.conv1(x))
        x = self.pool(x)
```

```
        x = F.relu(self.conv2(x))
        x = self.pool(x)
        x = x.flatten()
        x = F.relu(self.l1(x))
        x = F.softmax(self.l2(x))
        return x

def loss(y_pred, y):
    return F.cross_entropy(y_pred, y)
```

Zadatak 7.
Razmatramo konvolucijski model za klasificiranje jednokanalne slike:

- konvolucija 3×3 bez pomaka ($\mathbf{w}_1$, $\mathbf{w}_2$), nadopunjavanje, ReLU
- globalno sažimanje maksimumom
- pootpuno povezani sloj bez pomaka ($\mathbf{W}$)
- softmax

Inicijalizacija:

- $w_{121} = -1$, $w_{122} = 1$, ostali elementi $\mathbf{w}_1$ su 0
- $w_{232} = -1$, $w_{222} = 2$, ostali elementi $\mathbf{w}_2$ su 0
- $W_{11} = W_{22} = 1$, ostali elementi $\mathbf{W}$ su 0

Na ulazu je 4×4 matrica $\mathbf{x}$, $x_{23} = 1$, $x_{33} = 2$, svi ostali elementi su 0.

Odredite gradijente negativne log-izglednosti uz $y = 2$.

Rješenje. Odredimo prvo inicijalne parametre:

$$\mathbf{w}_1 = \begin{bmatrix} 0 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{w}_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 0 \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{W} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Operaciju konvolucije primijenjujemo na način da srednji element postavimo na gornji lijevi element slike i napravimo produkt po elementima i zbrojimo dobivene brojeve. Pomičemo se na sljedeći element i ponavljamo postupak dokle god nismo prošli sve elemente. Tamno iscrtane vrijednosti su maksimalne vrijednost i one će nam trebati za sljedeći korak.

Za prvi kanal to izgleda ovako:

$$\mathbf{s}_1 = \mathbf{X} * \mathbf{w}_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 2 & -2 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{h}_1 = ReLU(\mathbf{s}_1) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \mathbf{2} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Za drugi kanal:

$$\mathbf{s}_2 = \mathbf{X} * \mathbf{w}_2 = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{h}_2 = ReLU(\mathbf{s}_2) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{2} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Zatim slijedi primjena globalnog sažimanja maksimumom pa uzimamo maksimalnu vrijednost iz svake mape značajki.

$$GMP(\mathbf{h}_1) = 2, \quad GMP(\mathbf{h}_2) = 2,$$

Ovo konkateniramo u jedan vektor i množimo matricom $\mathbf{W}$:

$$\mathbf{s} = \mathbf{v} \cdot \mathbf{W} = \begin{bmatrix} 2 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 2 \end{bmatrix}$$

Konačno primijenimo softmax i dobivamo izlaz modela:

$$softmax(\mathbf{s}) = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix}$$

Unatražni prolaz započinjemo računanjem gradijenta vektora $\mathbf{s}$ po gubitku L:

$$\frac{\partial L}{\partial \mathbf{s}} = \mathbf{p} - \mathbf{y}_{OH} = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix} - \begin{bmatrix} 0 & 1 \end{bmatrix} = \begin{bmatrix} 0.5 & -0.5 \end{bmatrix}$$

Zatim slijedi gradijent gubitka po parametru $\mathbf{W}$ i gradijent gubitka po $\mathbf{v}$:

$$\frac{\partial L}{\partial \mathbf{W}} = \frac{\partial L}{\partial \mathbf{s}} \cdot \frac{\partial \mathbf{s}}{\partial \mathbf{W}} = \frac{\partial L}{\partial \mathbf{s}}^T \cdot \mathbf{v}^T = \begin{bmatrix} 0.5 \\ -0.5 \end{bmatrix} \cdot \begin{bmatrix} 2 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}$$

$$\frac{\partial L}{\partial \mathbf{v}} = \frac{\partial L}{\partial \mathbf{s}} \cdot \frac{\partial \mathbf{s}}{\partial \mathbf{v}} = \frac{\partial L}{\partial \mathbf{s}}^T \cdot \mathbf{W} = \begin{bmatrix} 0.5 & -0.5 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0.5 & -0.5 \end{bmatrix}$$

Prvu komponentu vektora $\frac{\partial L}{\partial \mathbf{v}}$ koristimo za unatražni prolaz prvog kanala, a drugu za unatražni prolaz drugog kanala. Kod računanja $\frac{\partial v_i}{\partial \mathbf{h}_i}$ koristimo matricu s nulama svugdje osim za maksimalan element gdje je vrijednost 1.

$$\frac{\partial L}{\partial \mathbf{h}_1} = \frac{\partial L}{\partial v_1} \cdot \frac{\partial v_1}{\partial \mathbf{h}_1} = 0.5 \cdot \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\frac{\partial L}{\partial \mathbf{h}_2} = \frac{\partial L}{\partial v_2} \cdot \frac{\partial v_2}{\partial \mathbf{h}_2} = -0.5 \cdot \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -0.5 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Slijedi gradijent gubitka po $\mathbf{s}_1$ i $\mathbf{s}_2$:

$$\frac{\partial L}{\partial \mathbf{s}_1} = \frac{\partial L}{\partial \mathbf{h}_1} \cdot \frac{\partial \mathbf{h}_1}{\partial \mathbf{s}_1} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \odot \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\frac{\partial L}{\partial \mathbf{s}_2} = \frac{\partial L}{\partial \mathbf{h}_2} \cdot \frac{\partial \mathbf{h}_2}{\partial \mathbf{s}_2} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -0.5 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \odot \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -0.5 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Još su preostali gradijenti po parametrima $\mathbf{w}_1$ i $\mathbf{w}_2$ i po ulazu $\mathbf{X}$.

$$\frac{\partial L}{\partial \mathbf{w}_1} = \frac{\partial L}{\partial \mathbf{s}_1} \cdot \frac{\partial \mathbf{s}_1}{\partial \mathbf{w}_1} = pad(\mathbf{X}) * \frac{\partial L}{\partial \mathbf{s}_1} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} * \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0.5 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\frac{\partial L}{\partial \mathbf{w}_2} = \frac{\partial L}{\partial \mathbf{s}_2} \cdot \frac{\partial \mathbf{s}_2}{\partial \mathbf{w}_2} = pad(\mathbf{X}) * \frac{\partial L}{\partial \mathbf{s}_2} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} * \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -0.5 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -0.5 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\frac{\partial L}{\partial \mathbf{X}} = pad(\frac{\partial L}{\partial \mathbf{s}_1}) * flip2d(\mathbf{w}_1) + pad(\frac{\partial L}{\partial \mathbf{s}_2}) * flip2d(\mathbf{w}_2) =$$

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} * \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -0.5 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} * \begin{bmatrix} 0 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} =$$

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -0.5 & 0 \\ 0 & 0 & 0.5 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0 \end{bmatrix}$$

Zadatak 8.
Pretpostavke:

- 3-dimenzionalni ulaz: $\mathbf{x} \in \mathbb{R}^3$
- dvije klase na izlazu: $\mathbf{y} \in \mathbb{R}^2$
- model: višerazredna logistička regresija

$$\mathbf{y} = softmax(\mathbf{W}\mathbf{x} + \mathbf{b})$$

- ulaz modela smo regularizirali dropoutom (bez mijenjanja aktivacija uz $p(\mu_{x_1}) = 0.2$, $p(\mu_{x_2}) = 0.5$, $p(\mu_{x_3}) = 0.7$

Potrebno je izračunati ulaz u softmax (eng. *logits*) za podatak $\mathbf{x} = [1\ 1\ 1]$, ako su parametri modela nakon učenja:

$$\mathbf{W} = \begin{bmatrix} -0.1 & 0.4 & -0.6 \\ 0.2 & -0.3 & 0.5 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 0.2 \\ -0.2 \end{bmatrix}$$

Rješenje. Izlazne težine aktivacije množimo s vjerojatnošću njene prisutnosti tijekom učenja:

$$\mathbf{w}'_{h_i} = p(\mu_{h_i}) \cdot \mathbf{w}_{h_i}$$

$$\mathbf{W}' = \begin{bmatrix} 0.2 & 0.5 & 0.7 \\ 0.2 & 0.5 & 0.7 \end{bmatrix} \odot \begin{bmatrix} -0.1 & 0.4 & -0.6 \\ 0.2 & -0.3 & 0.5 \end{bmatrix} = \begin{bmatrix} -0.02 & 0.2 & -0.42 \\ 0.04 & -0.15 & 0.35 \end{bmatrix}$$

Ulaz u softmaks onda glasi:

$$\mathbf{W}'\mathbf{x} + \mathbf{b} = \begin{bmatrix} -0.02 & 0.2 & -0.42 \\ 0.04 & -0.15 & 0.35 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0.2 \\ -0.2 \end{bmatrix} = \begin{bmatrix} -0.04 \\ 0.04 \end{bmatrix}$$

Zadatak 9.
Zadani su podaci x_1 do x_5.

Oznake podataka su redom $\mathbf{y} = [1, 0, 0, 1, 1]$

Udaljenosti od podatka x_1 iznose:

$$d(x_1, x_1) = 0.0$$
$$d(x_1, x_2) = 5.0$$
$$d(x_1, x_3) = 2.0$$
$$d(x_1, x_4) = 3.0$$
$$d(x_1, x_5) = 1.0$$

Odredite površinu ispod PR krivulje za predikcije modela u podatku x_1.

Rješenje. Prvo je potrebno rangirati podatke po udaljenosti:

$$\mathbf{x}_{sorted} = [x_5, x_3, x_4, x_2]$$

Za određivanje površine ispod krivulje prvo moramo izračunati krivulju. Potrebna nam je preciznost $P = \frac{TP}{TP+FP}$ i odziv $R = \frac{TP}{TP+FN}$.

Sada možemo napraviti tablicu:

Pozitivne predikcije	TP	FP	FN	P	R
$[x_5, x_3, x_4, x_2]$	2	2	0	0.5	1.0
$[x_5, x_3, x_4]$	2	1	0	0.66	1.0
$[x_5, x_3]$	1	1	1	0.5	0.5
$[x_5]$	1	0	1	1.0	0.5

Dogovorno se još dodaje i točka gdje je $R = 0$ i P za najmanji R tj. točka $R = 0$, $P = 1.0$. Površina ispod PR krivulje je onda:

$$AUPR = 0.5 \cdot 1.0 + 0.66 \cdot 1.0 = 0.83$$