



M Ű E G Y E T E M 1 7 8 2

Budapesti Műszaki és Gazdaságtudományi Egyetem

Villamosmérnöki és Informatikai Kar

Irányítástechnika és Informatika Tanszék

# M9 – Osztályozás mély neurális hálókkal

MÉRÉSI ÚTMUTATÓ

IRÁNYÍTÁSTECHNIKA ÉS KÉPFELDOLGOZÁS  
LABORATÓRIUM 1.

Szemenyei Márton

# Tartalomjegyzék

<b>1. Neurális hálók</b>	<b>2</b>
<b>2. A mérés környezete</b>	<b>3</b>
2.1. PyTorch . . . . .	3
2.2. Haladó Python . . . . .	5
2.3. Haladó NumPy . . . . .	6
<b>3. Mérési feladatok</b>	<b>9</b>
<b>4. Hasznos kódrészletek</b>	<b>10</b>
<b>5. Ellenőrző kérdések</b>	<b>13</b>

# 1 Neurális hálók

A számítógépes látás a számítástudomány egyik legrohamosabban fejlődő területe, amelynek egyre több gyakorlati felhasználása létezik. Ez a tendencia nem meglepő, hiszen az ember az érzékszervei közül a szemre hagyatkozik a leginkább a napi feladatának ellátásakor. Ebből következik, hogy ha a számítógépes látás algoritmusai képesek megközelíteni, vagy akár felülmúlni az emberi látás képességeit, akkor számos fontos feladatot leszünk képesek automatizálni. Könnyen belátható azonban az is, hogy a gyakorlatban ez egy rendkívül nehéz feladat, mivel általában olyan feladatokat tudunk könnyedén algoritmusok formájában leírni, ahol a feladat elvégzésének a menetét pontosan, tudatos szinten megértjük. A szemből érkező jelek feldolgozásának azonban a jelentős része tudatalatti szinten történik, így aligha tudjuk ezeket a folyamatokat könnyedén megérteni, és algoritmusok formájában lemásolni.

Ennek a problémának a megoldásához a mesterséges intelligencia, ezen belül is a gépi tanulás módszereihez kell nyúlnunk. A tanuló számítógépes látás legegyszerűbb formája az osztályozás, vagyis amikor egy képhez egyetlen címkét rendelünk, amely a képen található objektum kategóriáját kódolja. Bizonyos esetekben a címke mellé már egy az adott objektumot körbefogó téglalapot is rendelünk, ebben az esetben lokalizációról beszélhetünk. Adott esetben ennél magasabb szintű információt is szeretnénk kinyerni a képből: érdekelhet minket az, hogy melyek azok a pixelek az adott képen, amelyek az észlelt osztályhoz tartoznak. Amennyiben ezt az információt egy képsorozaton minden pixelre meghatározzuk, akkor követésről beszélhetünk.

A neurális hálózatokról és azok tanításáról bővebben a Számítógépes Látórendszerek c. tárgy jegyzetében [1] olvashatunk.

## 2 A mérés környezete

A mérés során a *PyCharm* elnevezésű IDE áll rendelkezésre, amely rendkívül sokoldalú szolgáltatásokkal könnyíti meg a szoftverfejlesztést, például konfigurálható automatikus formázási lehetőségek állnak rendelkezésünkre. További részletekért érdemes lehet a JetBrains ide vonatkozó weboldalát [2] felkeresni. Függvények, objektumok esetében a **Ctrl+P** billentyűkombináció pop-up segítségként szolgálva mutatja nekünk a paramétereket. A mérés során használt programnyelv a Python 3-as verziója lesz.

A Python programnyelvhez számos hasznos függvénykönyvtár tartozik, melyek a mérési feladatok megvalósítását nagymértékben megkönnyítik. A Python nyelv egyik rendkívül kényelmes funkciója a beépített package manager, amelynek segítségével az egyes könyvtárak automatikusan telepíthetők, telepítsük után pedig minden további beállítás nélkül használhatók. A Pythonhoz két ilyen package manager is tartozik, az egyik a Pip, amely a legtöbb telepíthető Python verzió mellé automatikusan települ, a másik pedig az Anaconda [3], ami a könyvtárkezelési funkciókon túl virtuális környezeteket is képes kezelni.

A Python egyik legfontosabb függvénykönyvtára a Numpy, amely tömbök kezelésére, illetve számtalan numerikus algoritmus használatára ad lehetőséget. A Numpy funkcionalitását kiegészíti a Matplotlib, melynek segítségével különböző ábrákat készíthetünk a tömbjeinkről. Egy harmadik rendkívül hasznos könyvtár család a scikit, ami számos tudományos számításához szükséges alkönyvtárt foglal össze. A scikit-image képek kezelésére, a scikit-learn gépi tanulás algoritmusok használatára, míg a scikit-fuzzy fuzzy logika használatára ad lehetőséget. Ezek a könyvtárak tulajdonképpen együttesen kiadják a Matlab funkcionalitásának jelentős részét.

### 2.1. PyTorch

A PyTorch [4], [5] a Facebook Research által fejlesztett Deep Learning keretrendszer, melyet a mérés során használni fogunk, így a következőkben röviden áttekintésre kerülnek a fontosabb koncepciók. A PyTorch egyik előnye a TensorFlow-val, Keras-szal szemben, hogy a hálózatot leíró gráf dinamikusan kerül létrehozásra,

nincs fordítási szakasza a tanításnak – ez a kevesebb lépés mellett akár azt is lehetővé teszi számunkra, hogy a háló rétegeit tanítás közben változtassuk (mert az egyik hiperparaméterünk lehet a lineáris rétegek száma).

A PyTorch alapja az akár GPU-n is futtatható tenzor típus, mely egy pratikus tagfüggvényekkel felvértezett, multidimenzionális tömböt megvalósító osztály – tehát felettebb alkalmas például képek tárolására is.

Tenzorokon alapul a rétegeket/hálózatokat megvalósító Module osztály, melyből minden saját hálózatunknak örökölnie kell, így kerülnek regisztrálásra az előre-, ill. visszatérjesztést megvalósító függvényeink. Szerencsére nem kell a hibavisszatérjesztést kézzel megírni, a PyTorch-beépített automatikus differenciálásra képes modulja (autograd) megteszi ezt helyettünk.

Fontos megjegyezni, hogy az automatikus differenciálás nem szimbolikus differenciálás (azaz nem az a célja, hogy paraméteresen felírja egy bonyolult összefüggés deriváltfüggvényét) és nem is numerikus differenciálás (tehát az sem cél, hogy iteratívan közelítsünk egy értéket, mert a kvantálás/kerekítés jelentős eltérésekben nyilvánulhat meg). Automatikus differenciálás során egy számítási gráf segítségével, a láncszabályt alkalmazva vagyunk képesek, viszonylag jól skálázhatóan, parciális deriváltakat kiszámítani, ami feltétlenül szükséges gradiens-alapú tanulóalgoritmusok esetében.

A PyTorch optim nevű almodulja tartalmazza a különböző, a tanítás során használt algoritmusokat, ezek közül nekünk most csak az SGD-re (stochastic gradient descent) lesz szükségünk. Az SGD objektum számára megadható a modell, a tanulási ráta (learning rate), a momentum, ill. van lehetőség normán alapuló regularizáció konfigurálására is.

Ahhoz, hogy tanítani tudjunk egy hálózatot, a következő struktúrájú programot kell megvalósítanunk és iteratívan meghívunk. A következőkben a PyTorch egyik kiegészítésének (Ignite) forráskódjából látható egy részlet, ami tömören magába foglalja a szükséges lépéseket – a mérésen ugyanezen lépések segítségével, de egy kicsit más struktúrában fogjuk a tanítást végezni.

```
def _update(batch):
    model.train()
    optimizer.zero_grad()
    x, y = _prepare_batch(batch, device=device)
    y_pred = model(x)
    loss = loss_fn(y_pred, y)
    loss.backward()
    optimizer.step()
    return loss.item()
```

Részleteiben a következő történik: tudatnunk kell a programmal, hogy most tanítjuk a hálót (hogy pl. a dropout csak ilyen esetben kerüljön aktiválásra). Ezt követően a gradiensek előző értékét kell nulláznunk (ugyanis megtehetnénk, hogy valamihez még felhasználjuk, ezért nem történik automatikusan). Ahhoz, hogy a következő előreterjesztést véghez vihessük, elő kell készíteni a következő minibatchet (ezt a kódban egy saját függvény, `_prepare_batch` jelöli). Az előreterjesztés végrehajtása felettébb egyszerű, magának a modellnek adjuk át argumentumként a minibatchet. A hiba kiszámítása a következő lépés, ezt az általunk konfigurált hibafüggvény segítségével tehetjük meg, majd a hiba backward tagfüggvényét meghívva végrehajtjuk a hibavisszaterjesztést. Nagyon fontos az utolsó lépés az optimalizáló algoritmus (SGD) lépésének végrehajtása, így kerülnek a súlyok frissítésre.

Számos gyakran használt adatbázis elérhető a torchvision modulból, ezek előkészítése, ill. adat augmentációra is használható transzformációk is rendelkezésünkre állnak. Ezen felül van lehetőség DataLoader objektumok segítségével (torch.utils.data modul) az adatok betöltését automatizálni. Továbbá lehetőségünk van például a tanulási ráta automatikus megváltoztatására adott szabályrendszer szerint.

## 2.2. Haladó Python

A korábbi mérés(ek) során megismerkedtünk a Python nyelvvel, a jelen mérés során azonban a nyelv bonyolultabb funkcióira is szükségünk lesz. Az alábbiakban tekintsük át legfontosabb adattípusokat és nyelvi elemeket.

A Python legfontosabb konténer osztályai a list és a tuple. Ezek közt a különbség, hogy a tuple immutable, vagyis nem módosítható. Létrehozásuk és az elemeik elérése az alábbi módon történhet:

```
myList = ["wow", "you", "can", "mix", 2, "or", "more", "different", "types"]
myTuple = ("wow", "you", "can", "mix", 2, "or", "more", "different", "types")
myList[4] = "two" # 2 becomes 'two'
myTuple[4] = "two" # Error
```

Érdemes tudni, hogy habár (1D) listák esetében a `+` operátor értelmezett, ez a listákat összefűzi, nem pedig összeadja.

```
myList = [1,2]+[3,4]
>>> [1,2,3,4] # Not [4,5]
```

Listákhoz új elemet az alábbi módon lehet hozzáadni:

```
myList = [1,2,3,4]
myList.append(5) # [1,2,3,4,5]
```

Listák, vagy Tuple-ök elemein az alábbi módokon lehet végigiterálni:

```
for elem in myList:
    print(elem)

for i, elem in enumerate(myList):
    print(i,elem) # i is the index
```

Adott esetben egyszerre több (egyenlő elemszámú) listán is végigiterálhatunk egyidejűleg:

```
if len(list1) == len(list2):
    for (elem1,elem2) in zip(list1, list2):
        print(elem1,elem2)
```

Érdemes megjegyezni, hogy az 'in' kulcsszó feltételeknél is használható annak eldöntésére, hogy egy adott érték szerepel-e a listában:

```
myList = [1,2,3,4]
if 1 in myList:
    print("1 is in the list ")
if 5 not in myList:
    print("5 is not in the list ")
```

A Python programnyelv támogat bizonyos funkcionális programozási elemeket is, melyek közül az egyik leghasznosabb az úgynevezett list comprehension. Ezt akkor használhatjuk, ha szeretnénk egy lista minden elemén elvégezni egy műveletet és ezek eredményeiből egy újabb listát készíteni. Az alábbi példa egy lista minden elemét négyzetre emeli:

```
myList = [1,2,3,4]
mySqrList = [elem**2 for elem in myList] # [1,4,9,16]
```

A list comprehension a korábban említett for ciklus változatokkal (zip, enumerate) is használható, az új lista elemének kiszámolásához pedig bármilyen érvényes Python kód írható. Ezen felül a kifejezés feltétellel is kiegészíthető:

```
myList = [9,-16,25,-4]
mySqrList = [sqrt(elem) for elem in myList if elem >= 0] # [3,5]
```

## 2.3. Haladó NumPy

Fontos megjegyezni, hogy a Python list és tuple osztályok különböző típusú objektumokat tartalmazhatnak, vagyis egy lista egyik eleme lehet egy másik lista. Ez lehetőséget nyújt arra, hogy a lista osztályt N-dimenziós tömbként használjuk, azonban

egy listákat tartalmazó listában az egyes részlisták különböző hosszúak és típusúak lehetnek, ami hagyományos tömböknél nem fordulhat elő. Éppen ezért méret- és típus konzisztens tömbök tárolására a NumPy array osztályt célszerű használni.

A NumPy tömbök között definiált a  $+$ ,  $-$ ,  $*$  és a  $/$  művelete, azonban ezeket elemenként végzi, így a tömbök méretének kompatibilisnek kell lennie. A NumPy broadcasting szabályai alapján két tömb az alábbi feltételek teljesítése esetén kompatibilis méretű:

1. A két tömb minden mérete megegyezik
2. Amelyik dimenzió mentén nem egyeznek meg a méretek, ott az egyik tömb mérete 1

A tömbök méretét az alábbi módon ellenőrizhetjük:

```
arr1.shape # [2,4,6]
arr2.shape # [1,4,1] - Compatible
arr3.shape # [3,4,6] - Incompatible
```

Bizonyos esetekben előfordulhat, hogy új elemet szeretnénk a tömbhöz adni, vagy két tömböt össze szeretnénk fűzni, amelyeket az alábbi módon tehetünk meg:

```
extArr = np.append(arr,elem,dim)
catArr = np.concatenate((arr1,arr2 ,..., arrn),dim)
```

A *dim* változó azt adja meg, hogy melyik dimenzió mentén történjen a kiegészítés/összefűzés. Ha None értéket adunk meg (nem adjuk meg a paramétert), akkor a NumPy megpróbálja magától kitalálni.

Gyakran előfordul, hogy egy általunk használt tömbben van egy extra dimenzió, aminek az értéke egy. Például, ha van egy 2D mátrixokból álló tömbünk, amelyiknek kivesszük egy elemét. Ez az extra dimenzió sok fejfájást tud okozni, mert ha konkatenálni, vagy mátrixokat szorozni szeretnénk, a NumPy nem megfelelő méret hibákat fog dobni. Ennek megoldására a squeeze parancs való.

```
mtxArr.shape # [N,3,3] - N 3x3 matrices
myMtx = mtxArr[i]
myMtx.shape # [1,3,3]
myMtx = mtxArr[i].squeeze()
myMtx.shape # [3,3]
```

Előfordulhat az is, hogy a tömbjeink megfelelő méretűek, de valamilyen oknál fogva a dimenziók sorrendje eltér. Például képek esetén elterjedt konvenció a  $H \times W \times Ch$  sorrend használata, a Deep Learning könyvtárak viszont a  $Ch \times H \times W$  sorrendet preferálják. Erre megoldás a NumPy transpose függvény.



```
myArr.shape # [N,3]
np.transpose(myArr).shape # [3,N]

# Több dimenzióban
myImg.shape # [1080,1920,3]
np.transpose(myImg,(2,0,1)).shape # [3,1080,1920]
```

Ezen felül érdemes még említeni, hogy a NumPy képes különböző lineáris algebrai műveletek elvégzésére, mint például a mátrixszorzás (amennyiben a dimenziók megfelelőek), valamint az invertálás:

```
myMtx = np.array ([[1,2,3],[4,5,6],[7,8,9]])
myInv = np.linalg.inv(myMtx)

myData = np.randn((50,3)) # 50 3D coordinates
myTrData = np.matmul(myMtx,myData) # myTrData = myMtx*myData
```

Fontos megemlíteni, hogy a *numpy.ndarray* típuson elvégezhető műveletek nagy része pontosan ugyanúgy működik *torch.tensor* típuson is, bár esetenként ugyanaz a függvény más néven érhető el. A *numpy.transpose* például *torch.permute*.

## 3 Mérési feladatok

A mérés folyamán az alábbi feladatokat kell elvégezni:

1. Készítsen egy egyszerű néhány rétegből álló neurális háló architektúrát az 3.1. Táblázatban megadott paraméterek alapján.
2. Készítsen eljárást egyszerű neurális háló tanítására a méréshez használt adatbázison.
3. Végezzen el különböző tanításokat a 3.2. Táblázatban megadott hiperparaméterek alapján.
4. Valósítson meg egy egyszerű közlekedési tábla detektáló és osztályozó algoritmust a mérésvezető által szolgáltatott képekre.

Típus	Csatornák száma	Méret	Stride
Konvolúciós	$N_{CH}$	$3 \times 3$	1
Konvolúciós	$N_{CH} * 2$	$3 \times 3$	2
Konvolúciós	$N_{CH} * 2$	$3 \times 3$	1
Konvolúciós	$N_{CH} * 4$	$3 \times 3$	2
Konvolúciós	$N_{CH} * 4$	$3 \times 3$	1
Konvolúciós	$N_{CH} * 8$	$3 \times 3$	2
Konvolúciós	$N_{CH} * 8$	$3 \times 3$	1
Konvolúciós	$N_{CH} * 16$	$3 \times 3$	2
Konvolúciós	$N_{CH} * 16$	$3 \times 3$	1
Konvolúciós	$N_{CH} * 32$	$3 \times 3$	2
Osztályozó	55	$1 \times 1$	1

**3.1. táblázat.** *Az elkészítendő neurális háló architektúrája.*

$N_{CH}$	LR	Eta_min	Momentum	Decay
8	1e-1	1e-2	0.1	1e-3
8	1e-1	1e-2	0.9	1e-4
8	1	1e-2	0.9	1e-1
8	1e-3	1e-4	0.9	1e-5
16	1e-1	1e-2	0.9	1e-5

**3.2. táblázat.** *A Vizsgálandó hiperparaméter kombinációk.*

## 4 Hasznos kódrészletek

Különböző rétegek létrehozása PyTorch környezetben

```
self.conv = nn.Conv2d(inplanes, planes, size, padding=size//2, stride=stride)
self.bn = nn.BatchNorm2d(planes)
```

Rétegek meghívása

```
out = self.bn(F.relu(self.conv(x)))
```

Térbeli dimenziók eldobása az osztályozó réteg után

```
out = out.squeeze()
```

Adat augmentációs függvények

```
transforms.RandomCrop(32, padding=4),
transforms.ColorJitter(brightness=0.2, contrast=0.2, saturation=0.2, hue=0.1),
transforms.ToTensor(),
transforms.Normalize((0.49139968, 0.48215827, 0.44653124),
                      (0.24703233, 0.24348505, 0.26158768))
```

Különböző augmentációs módszerek összefűzése

```
transform = transforms.Compose( <list of transforms> )
```

Adatbázis létrehozása

```
dataSet = torchvision.datasets.ImageFolder(root=root, transform=transform)
```

Adat betöltő létrehozása

```
dataLoader = torch.utils.data.DataLoader(trainSet, batch_size=512, shuffle=True,
                                         num_workers=4)
```

Költségfüggvény létrehozása

```
criterion = nn.CrossEntropyLoss()
```

Optimalizáló módszer létrehozás

```
optimizer = optim.SGD(net.parameters(), lr=0.1, momentum=0.9, nesterov=True,
                      weight_decay=1e-4)
```

Tanulási ráta ütemező létrehozása

```
scheduler = lr_scheduler.CosineAnnealingLR(optimizer,numEpoch,eta_min)
```

Neurális háló módjainak állítása

```
net.train()  
net.eval()
```

Progress bar készítése

```
# Create progress bar  
bar = progressbar.ProgressBar(0, len(trainLoader), redirect_stdout=False)  
    # Inside the loop:  
    bar.update(i)  
# Upon finishing:  
bar.finish()
```

Dataloaderen történő iterálás:

```
# Epoch loop  
for i, (inputs, labels) in enumerate(trainLoader, 0):
```

CUDA konverzió

```
if haveCuda:  
    inputs = inputs.cuda()
```

Optimizer kinullázása

```
optimizer.zero_grad()
```

Forward

```
outputs = net(inputs)
```

Költség

```
loss = criterion(outputs, labels)
```

Backward

```
loss.backward()
```

Egy lépés gradiens módszerrel

```
optimizer.step()
```

Tanulási ráta ütemező léptetése

```
scheduler.step()
```

Modell mentése

```
torch.save(net, root + '/model.pth')
```

OpenCV kép PyTorch tenzorra történő konvertálása

```
# BGR to RGB
img2 = cv2.cvtColor(img, cv2.COLOR_BGR2RGB)
# Convert to PIL image
pil_im = Image.fromarray(img2)
# Apply transform function, unsqueeze and convert to cuda
imageT = transform(pil_im).unsqueeze(0).cuda()
```

Küszöbözés határok közt

```
img_bin = cv2.inRange(img_hsv,np.array([0,100,80]),np.array([255,255,255]))
```

Morfológia

```
SE = np.ones((3, 3), np.uint8)
img_bin = cv2.morphologyEx(img_bin, cv2.MORPH_CLOSE, SE,iterations=4)
```

Kontúrok keresése

```
contours, __ = cv2.findContours(binImage,cv2.RETR_EXTERNAL,cv2.
    CHAIN_APPROX_SIMPLE)
```

Bounding box meghatározása (x,y,w,h)

```
rect = cv2.boundingRect(contour)
```

Kontúr területe

```
area = cv2.contourArea(contour)
```

Képrészlet kivágása (hard copy)

```
imgROI = img[y1:y2,x1:x2].copy()
```

Kép átméretezése

```
imgROI = cv2.resize(imgROI,(32,32))
```

Numpy - torch konverzió és dimenziók cserélése

```
imgROI = torch.Tensor(imgROI)
permuted = imgROI.permute(2,0,1)
```

Téglalap rajzolása és kiírás

```
cv2.rectangle(img,(x1,y1),(x2,y2),(255,0,255),2)
cv2.putText(img,className,(x1,y1-10),cv2.FONT_HERSHEY_COMPLEX_SMALL
    ,1.0,(0,255,0))
```

## 5 Ellenőrző kérdések

1. Ismertesse röviden a Hinge és a Kereszt-entrópia költségfüggvényeket! Mi az előnyük/hátrányuk?
2. Hogyan működik a gradiens módszer? Milyen fontos kiegészítései vannak?
3. Mire való a backpropagation? Hogyan működik?
4. Ismertesse egy konvolúciós háló és gyakori rétegei felépítését!
5. Milyen módszerekkel kerülhető el az overfitting? Ismertesse ezeket egy-egy mondatban!
6. Milyen módszerekkel javítható egy neurális háló konvergenciája? Ismertesse ezeket egy-egy mondatban!
7. Ismertesse röviden a PyTorch könyvtárat! Milyen alapvető adattípusai, moduljai léteznek? Milyen szolgáltatásokat nyújt neurális hálók számára?
8. Készítsen egy egyszerű szkriptet Python nyelven:
  - (a) Töltse be a torch könyvtár nn nevű modulját!
  - (b) Definiáljon egy új, MyNet nevű osztályt, melynek ősosztálya az nn.Module
  - (c) Az osztály konstruktorában hozzon létre egy mult nevű tagváltozót, értéke legyen 5.
  - (d) Definiáljon egy forward nevű tagfüggvényt, melynek bemeneti paramétere x, és a mult nevű tagváltozóval szorozza meg, az eredményt pedig adja vissza.
9. Hogyan néz ki egyetlen tanulási ciklus PyTorch-ban? (Pszedó kód elég)

# Bibliography

- [1] M. Szemenyei, *Számítógépes Látórendszerek*, M. Szemenyei, Ed. BME, 2019. [Online]. Available: <http://deeplearning.iit.bme.hu/jegyzetFull.pdf>.
- [2] *PyCharm Quick Start Guide*. [Online]. Available: <https://www.jetbrains.com/help/pycharm/quick-start-guide.html>.
- [3] *Anaconda*. [Online]. Available: <https://www.anaconda.com/>.
- [4] *PyTorch Documentation*. [Online]. Available: <https://pytorch.org/docs/stable/index.html>.
- [5] *PyTorch Tutorials*. [Online]. Available: <https://pytorch.org/tutorials/index.html>.