Gépi tanulás

Képfelismerés gépi tanulásos módszerek alkalmazásával python nyelven

írta: Köő József Témavezető: Czúni lászló

2018

Tartalom

[Bevezetés 2](#_Toc511449831)

[Python 2](#_Toc511449832)

[Képadatbázis 2](#_Toc511449833)

[K-legközelebbi szomszéd 2](#_Toc511449834)

[Lineáris osztályozás 2](#_Toc511449835)

[Optimalizációs eljárások 2](#_Toc511449836)

[Neurális hálózatok 2](#_Toc511449837)

[Hiba visszaterjesztés 2](#_Toc511449838)

# Bevezetés

Célom a projekt során megismerni, és bemutatni a gépi tanulás néhány módszerét és alkalmazását. Néhány gyakorlati példán keresztül ismertetném az egyes megoldásokat adott problémára.

Forrásként főként a Stanford egyetem egy online elérhető, a CS231n nevű kurzusának anyagát használtam. A kurzus alapvetően a konvolúciós neurális hálókat hivatott bemutatni, de emellett bemutat még más tanulási módszereket is.

A projekt során a fő probléma, amin keresztül bemutatom az egyes tanulási algoritmusokat, egy osztályozás jellegű probléma. Egyszerű képek besorolása néhány kategóriába a tartalmuk alapján. Például egy macskát ábrázoló képet a macskák kategóriába kell besorolnunk.

# Python programnyelv

Mivel a forrásként megjelölt kurzusban is a Python programnyelvet használják, valamint napjainkban elég előrehaladott, és népszerű nyelv ez, ezért a problémák megoldásait én is Python nyelven prezentálom.

A Python egy általános célú szkript nyelv, amely fordítási időben futtatja a programot. A feladatainkra az teszi igazán alkalmassá, hogy igen erős támogatottsága miatt léteznek olyan könyvtárak hozzá, amelyek igen hatékonyan kezelik a tudományos számítási műveleteket. A numpy könyvtár tartalmazza azokat az eszközöket, amelyekkel igen egyszerű és hatékony kezelni a mátrix műveleteket. Könnyen dolgozhatunk vele többdimenziós adattömbökön, adathalmazokon is.

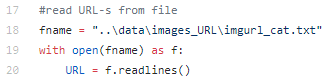
# Képadatbázis

A gépi tanulásos módszerek alapját a tanítási halmaz képezi. Szükséges a megléte egy olyan adathalmaznak, amely már osztályozott elemeket tartalmaz, és rendelkezik olyan címkével, amivel majd az új adatokat az algoritmusunk fog ellátni.

## ImageNET

A képek forrásaként először az Image-net.org weboldalt használtam. Itt elérhető több mint 14 millió kategorizált kép, és kategóriánként van lehetőség a letöltésükre. Van lehetőség továbbá a képek URL-jének letöltésére is. Ezek szintén kategóriák szerint érhetőek el egy-egy szövegfájlban felsorolásként. Ez utóbbi forrást választottam, amihez szükségeltetett egy feldolgozó program, ami letölti automatizáltan a képeket az URL-ek alapján.

A projektemben a *00\_image\_download\_from\_URL.py* fájl tartalmazza a kódot, ami a képek letöltéséért felelős. Az első részletben beolvastam a sorokat a fájlból egy tömbbe.



Így az URL tömb tartalmazza az összes elérési utat, ami macskát ábrázoló képekre mutat.



Egy ciklussal végighaladva a tömbön egyesével letöltöttem a képeket egy meghatározott mappába. A képek nevei és a mappa neve reprezentálta a rendszeremben a címkéket.

A képek letöltésekor azonban probléma lépett fel, ugyanis a fájl tartalmazott sok olyan elérési utat ahol már vagy nem létezett a kép, vagy nem volt elérhető. Így a teljes adathalmaznak csak egy részét tudtam ezzel a módszerrel megszerezni. Továbbá a sikeresen letöltött képek is tartalmaztak fals értékeket, ugyanis néhány oldalon a megszűnt képet egy hibát reprezentáló képpel helyettesítették, így ezt a programom már nem tudta kiszűrni, és címkézett adatként le is töltötte.

## CIFAR-10

Egy másik megoldást, a CIFAR-10 adatbázis jelentett. Ez egy speciális adathalmaz, ami 60000 darab 32x32 pixel méretű színes képet tartalmaz. A képek tíz osztályba vannak sorolva:

* repülőgép
* autó
* madár
* macska
* szarvas
* kutya
* béka
* ló
* hajó
* teherautó

Ezeket a képeket speciális formátumban lehet elérni. A letöltést követően 6 kiterjesztés nélküli fájlhoz jutunk hozzá. A 6 fájl összekeverve tartalmazza a képeket, mindegyikhez hozzárendelve egy osztályt jelentő számot. Öt fájlt jelöltek meg tanító halmazként, és egy fájlt tesztelő halmazként. Minden fájl azonos hosszúságú, és a következőképpen tárolja a kép adatokat:

Egy fájl 10000 képet tartalmaz. Egy képet egy „sor” reprezentál. Az első bájt az osztályt jelölő címkét tartalmazza (0-tól 9-ig), az utána következő 1024 bájt jelenti a kép vörös pixelkomponenseit, majd 1024 bájt a zöld, és újabb 1024 bájt a kék komponenseket tartalmazza.

# K-legközelebbi szomszéd

Tbd

# Lineáris osztályozás

Tbd

# Optimalizációs eljárások

Tbd

# Neurális hálózatok

Tbd

# Hiba visszaterjesztés

Tbd