

MESTERSÉGES INTELLIGENCIA

2. forduló



A kategória támogatója: Ulyssys Kft.

Ismertető a feladatlaphoz

Kérjük, hogy a feladatlap indítása előtt mindenképp olvasd el az alábbi útmutatót:

Helyezéseket a 4. forduló után mutatunk, százalékos formában: adott kategóriában a TOP 20-40-60%-hoz tartozol.

A feltűnően rövid idő alatt megoldott feladatlapok kizárást vonnak maguk után, bármilyen más gyanús esetben fenntartjuk a jogot a forduló érvénytelenítésére!

Jó versenyzést kívánunk!



Ebben a fordulóban 4 feladat lesz, három könnyebb és egy komplexebb. Saját GPU-ra nem lesz szükséged, helyette nyugodtan használj Colab-ot (<https://colab.research.google.com/>) !

1. feladat 2 pont

Melyik a kakukktojás?

Válasz

- ☐ Mask-R CNN
- ☐ YOLO v4
- ☐ SAM
- ☐ BERT

2. feladat 4 pont

Párosítsd össze a háló implementációkat a legvalószínűbb feladatukkal!

A válaszokhoz tartozó betűjeleket a hálók sorrendjében add meg a szövegdobozban, például így: "uvxy", amennyiben az első háló az U feladatot, a második a V-t... stb oldja meg.

Hálók:

1)

```
model = tf.keras.Sequential([
    layers.Embedding(length, embedding_dim),
    layers.Dropout(0.2),
    layers.GlobalAveragePooling1D(),
    layers.Dropout(0.2),
    layers.Dense(len(label_dict))])
```

2)

```
model = tf.keras.Sequential([
    layers.LSTM(64, input_shape=input_shape, return_sequences=True),
    layers.Dropout(0.2),
    layers.LSTM(128, return_sequences=True),
    layers.Dropout(0.2),
    layers.LSTM(256),
    layers.Dropout(0.2),
    layers.Dense(1) ])
```

3)

```
model = tf.keras.Sequential([
    layers.Conv2D(64, (3, 3), activation='relu', padding='same', input_shape=input_shape),
    layers.MaxPooling2D((2, 2)),
    layers.Conv2D(128, (3, 3), activation='relu', padding='same'),
    layers.MaxPooling2D((2, 2)),
    layers.Conv2D(256, (3, 3), activation='relu', padding='same'),
    layers.MaxPooling2D((2, 2)),
    layers.UpSampling2D((2, 2)),
    layers.Conv2D(128, (3, 3), activation='relu', padding='same'),
    layers.UpSampling2D((2, 2)),
    layers.Conv2D(64, (3, 3), activation='relu', padding='same'),
    layers.UpSampling2D((2, 2)),
    layers.Conv2D(32, (3, 3), activation='relu', padding='same'),
    layers.Conv2D(1, (1, 1), activation='sigmoid')])
```

4)

```
model = tf.keras.Sequential([layers.Input((500,500,3)),
    layers.Conv2D(filters=32,kernel_size=(3,3),activation="relu"),
    layers.MaxPooling2D((2,2)),
    layers.Conv2D(filters=64,kernel_size=(3,3),activation="relu"),
    layers.MaxPooling2D((2,2)),
    layers.Flatten(),
    layers.Dense(128,activation='relu'),
    layers.Dense(10,activation="softmax")])
```

Feladatok:

- a) idősoros előrejelzés
- b) szöveg klasszifikáció
- c) kép szegmentáció
- d) kép klaszifikáció

Válasz

A legjobb gyógyszer:

Egy kedves ismerősödet igen ritka betegséggel diagnosztizálták. Több kísérleti fázisú gyógyszer is rendelkezésre áll a piacon, de még kevés az adat melyik a legmegfelelőbb.

Te azonban találtál egy kicsi adathalmazt az interneten, ugyanebben a betegségben szenvedő emberekről, mely megmondja, hogy nekik a rendelkezésre álló 5 gyógyszerből melyik vált be a legjobban.

Erre az adatra rátanítottál egy Gini Decision Treet, ami a teszt adaton nagyon jól teljesít. Nézd meg hogy a barátodnak eszerint milyen gyógyszer lenne a legjobb?

A barátod egy 30 éves, normális vérnyomású, normális koleszterinszintű nő, akinek a nátrium/kálium szintje 8.678.

Adott: https://colab.research.google.com/drive/1I3JAOQJ5KAgy-SIPRKhmcvk4kdzgNk_8

Válasz

- ☐ drugA
- ☐ drugB
- ☐ drugC
- ☐ drugX
- ☐ drugY

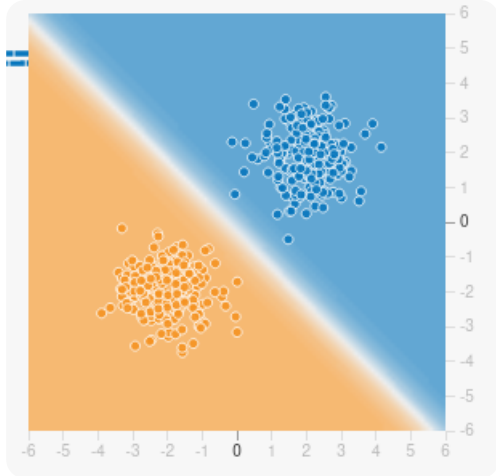
4. feladat 5 pont

Neural Network playground

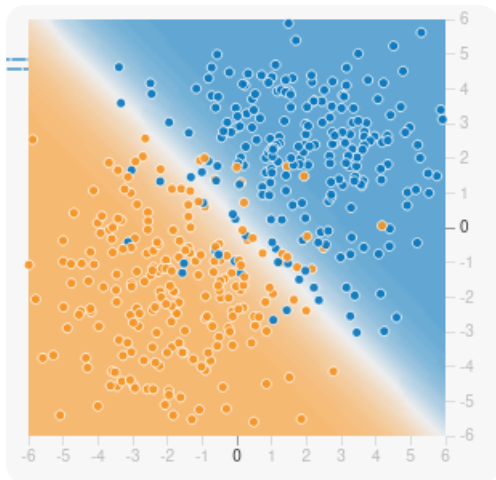
Van egy adathalmazunk, amiben minden input 2 featurrel rendelkezik, melyek értéke -6 és +6 közötti valós szám. 2-féle prediktálandó címke van, a "narancs" és a "kék". Ha ábrázolni szeretnénk az adathalmazt, akkor a feature-öket használhatjuk koordinátaként, és az így kapott pontokat elhelyezhetjük egy 2D koordináta-rendszerben. Minden pont egy 12 egység oldalú négyzeten belülre fog esni. Tegyük fel, hogy esetünkben a "narancs" pontok a négyzet bal alsó, a "kék" pontok a négyzet jobb felső részében szimmetrikusan lévő centrum körül Gauss-eloszlással vannak elszórva. (Erre láthatsz alább két példát, amik csak a szórás mértékében térnek el.)

Könnyen rájöhetünk - így hogy a pontok eloszlását megsúgták - hogy ezen az adathalmazon milyen lenne az elméletben ideális modell. Statisztikai okokból, ha kellően sok adatpont áll rendelkezésre (és figyelembe vesszük a mintavételezés eloszlásának szimmetriáját), az következik, hogy az ideális klasszifikáló modell a (-6;6) és (6;-6) közötti átló alá eső pontokat "narancs"-nak, a fölé eső pontokat "kék"-nek klasszifikálja. Az alábbi ábrák háttérének színezése is ezt mutatja. Tekinthető ez a színezés úgy, mint a cél, amit egy tanító- és teszt-halmazon egyaránt jól működő modellről elvárhatunk.

kis szórás (csak szemléltetésképp):



nagy szórás (ilyennel fogunk dolgozni):



Az alábbi linken az alsó ábrához hasonló adathalmaz generálódik, aminek csak egy kis részét fogjuk a neurális hálónk betanítására használni (a többit a modell teszteléséhez használjuk).

Kattintsunk a linkre, és az oldal betöltését követően indítsuk el a tanítást a ">" gombbal! (Ha a tanítást reprodukálni kellene, nyissuk meg újra ugyanezt a linket, ez beállítja a random seedet is!)

Figyeljük meg, hogyan alakul a tanítás és a teszt vesztesége, illetve a predikció döntési felülete! A tanítást kb. 2000 epoch elteltével leállíthatjuk.

<https://playground.tensorflow.org/#activation=tanh&batchSize=8&dataset=gauss®Dataset=reg-plane&learningRate=0.03®ularizationRate=0&noise=50&networkShape=6,6,6&seed=0.45185&showTestData=false&discretize=false&percTrainData=20&x=true&y=true&xTimesY=false&xSquared=false&ySquared=false&cosX=false&sinX=false&cosY=false&sinY=false&collectStats=false&problem=classification&initZero=false&hideText=false>

Melyek az igaz állítások az alábbiak közül?

Válaszok

- ☐ A tanítás kései szakaszában a teszt loss csökken.
- ☐ Túltanulás jelei mutatkoznak.
- ☐ Alultanulás jeleit látjuk.
- ☐ Ha több tanítóadat lenne, enyhülne a fentiek közül igaz effektus.
- ☐ Ha az adatpontok kisebb szórással lennének generálva, a betanult modell kevésbé lenne sikeres a tesztalmonon.

- ☐ A döntési felület tagolttá válik: elkülönülnek olyan területek, amik csupán néhány távolszóródott adatpontot tartalmaznak.
- ☐ Az outputot 1 perceptron állítja elő.
- ☐ Ha az összes rejtett réteget ("HIDDEN LAYERS") töröljük, a modell akkor is képes lehet megtanulni a fenti értelemben ideális döntési felületet.

Megoldások beküldése