

Magyar MI Diákolimpia

Klasszifikáló Klón

Feladatleírás

2025. május 24.

1. Klasszifikáló Klón

Adott egy előre betanított neurális háló, amelyet klasszifikációs feladatra tanítottunk be egy meghatározott **training** adathalmazon. A feladat megoldásához a rendelkezésekre áll:

- A betanított neurális háló;
- a tanító adathalmaz (`X_train`) és címkék (`y_train`);
- a teszt adathalmaz (`X_train_small`) és címkék (`y_train_small`) egy kis részhalmaza.

A cél egy olyan gépi tanulási modell elkészítése, amely **nem neurális háló alapú**. Ennek a modellnek az a feladata, hogy **a teljes teszt adathalmazon a lehető legjobban közelítse meg a betanított neurális háló kimenetét**, azaz próbálja meg utánózni annak viselkedését.

2. A feladat pontozása

A végleges modell a teljes teszt adathalmazon lesz kiértékelve. A pontozás az accuracy metrika szerint, a következő skála alapján történik (az intervallum jobb oldala nyílt):

- 0 - 0.5: 0 pont
- 0.5 - 0.62: 20 pont
- 0.62 - 0.66: 30 pont
- 0.66 - 0.68: 40 pont
- 0.68 - 0.70: 50 pont
- 0.70 - 0.71: 70 pont
- 0.71 - 0.728: 80 pont
- 0.728 - 0.80: 90 pont
- 0.80 - 1: 100 pont

Maximális pontszám: 100 pont

A pontozás során az első cellában beállított seed lesz használva. Ezen felül javasolt a modellek számára is random állapotot beállítani, mivel ez befolyásolhatja hogy milyen eredményt érnek el. A kiértékelés során minden kód csupán egyszer lesz futtatva.

3. Technikai tudnivalók

A feladat megoldásához elegendő a mellékelt `.ipynb` notebook fájl, a titkos modell súlyai, valamint az adathalmaz. A kiértékeléshez egy `train-test.pt` nevű fájl is rendelkezésre áll, azonban ezt kizárólag a megoldások végén, a modell kiértékelésére használhatjátok fel. A tanítás vagy a fejlesztés során **tilos** használni. Segítségképpen több hasznos könyvtár és útmutató is megtalálható a notebook elején.

A feladat megoldása után, függetlenül attól, hogy Google Colabban vagy lokálisan dolgoztál, a `.ipynb` fájlt kell feltöltened a CMS rendszerbe a megfelelő feladathoz. Más fájl feltöltésére nincs szükség.

4. Hasznos források

-  Scikit-Learn - Dokumentáció
-  Catboost - Dokumentáció
-  LightGBM - Dokumentáció
-  XGBoost - Dokumentáció