



ELTE | IK
INFORMATIKAI KAR

BigData architektúrák és elemző módszerek

9. gyakorlat

BigData architektúrák és elemző módszerek

- Eddig: architektúrák (HDFS, Hadoop, Spark)
- Ezután: elemző módszerek
- Cél: adatokból információt kinyerni
 - Statisztikai elemzés
 - Gépi tanulás

Gépi tanulás

- Felügyelt tanulás (supervised learning)
- Nem felügyelt tanulás (unsupervised learning)
- Megerősítéssel tanulás (reinforcement learning)

Felügyelt tanulás

- Osztályozás (klasszifikáció, classification)
 - Jellemzője: diszkrét értékeket kell „kitalálni”
 - Példák: képek osztályozása, spam e-mail megjelölése, banki csalások
 - Modellek: döntési fa, k-NN, SVM, naive bayes
- Regresszió (regression)
 - Jellemzője: folytonos értéket kell megbecsülni
 - Példák: piaci árak előrejelzése, várható élettartam, időjárás előrejelzés
 - Módszerek: lineáris regresszió

Nem felügyelt tanulás

- Klaszterezés (clustering)
 - Célja: csoportok felfedezése
 - Példa: felhasználók csoportosítása
 - Módszerek: hierarchikus, k-means
- Asszociációs szabályok (association rules)
 - Célja: az adatokban lévő összefüggések feltárása
 - Példák: felhasználói kosár elemzés
 - Módszerek: apriori
- Ajánló rendszerek (recommender systems)
- Dimenzió redukció (dimension reduction)

Az elemzés folyamata

- Adat tisztítás (mai gyakorlat)
- Modell készítés
- Kiértékelés