

ADATALAPÚ MEGOLDÁSOK

Kiselőadás témák

AWS

- Az Amazon Web Services (AWS) piacvezető az IaaS (Infrastructure-as-a-Service) és a PaaS (Platform-as-a-Service) felhőalapú rendszerek szolgáltatásában. Használható többek között adatok távoli eléréséhez, publikálásához, weboldalak/mobil applikációk készítéséhez és erőforrás igényes feladatok futtatásához.
- Kiselőadásod során ismertesd, a felhőalapú szolgáltatások szintjeit, melyik szint mire való. Fejtsd ki bővebben milyen adat alapú megoldásokhoz köthető szolgáltatások vannak. Összegezd, hogy milyen előnyei és hátrányai vannak a választott szolgáltatások használatának.

Kafka (elkelt)

- Apache Kafka egy adatalapú szolgáltatás, ahol a felhasználók adatokat cserélhetnek. Az adat publikálója nem egy konkrét helyre küldi az adatot, hanem a fogadók feliratkozhatnak az adat-folyamra. Az adatok korlátozott ideig a felhőben is tárolva vannak.
- Kiselőadásod során ismertesd a kafka működését. Fejtsd ki bővebben milyen adat alapú megoldásokhoz használható fel. Összegezd, hogy te miért vagy miért nem használnád ezt a szolgáltatást.

Scala (elkelt)

- A Scala egy java alapú programozási nyelv, amely egyszerre funkcionális és tartalmaz objektum orientált elemeket is.
- Kiselőadásod során ismertesd Scala programozási nyelv elemeit. Milyen adatalapú megoldásokhoz használható fel? Összegezd, hogy te miért vagy miért nem használnád ezt a programozási nyelvet.

Spark

- Az Apache Spark technológia segítségével elosztott számításokat lehet végezni. Az egyes egységek külön végzik el a számításaikat, majd ezek összesítésre kerülnek.
- Kiselőadásod során ismertesd a Spark működését. Milyen elemei vannak? Milyen felhasználási területei vannak? Összegezd, hogy te miért vagy miért nem használnád ezt a technológiát.

Web crawling (elkelt)

- A web crawling a weben fellelhető információk összegyűjtése és a fontos adatok, információk tárolása.
- Kiselőadásod során ismertesd a web crawling működését? Milyen megoldások találhatók? Milyen feladatokat kell web crawling során megoldani? Összegezd, hogy te miért vagy miért nem használnád az adott technológiát.

Bag of (Visual) Words (elkelt)

- A Bag of Words vagy másnéven szózsák modell egy reprezentáció készítő eljárás modellje, melynek alapötlete az, hogy egy dokumentumot a benne található szavak összességével vagy eloszlásával jellemzünk, miközben azok térbeli elhelyezkedésétől eltekintünk. Ennek egy vizuális tartalmakra kiterjesztett változata a BoVW, ahol képeket jellemzünk a rajtuk található vizuális kódszavak eloszlásával.
- Ismertesd a modell lépéseit és részfeladatait. Milyen gépi tanulási eljárások során szokás használni? Mik az előnyei és hátrányai?

Gaussian Mixture Modell

- A GMM egy nagyon sokoldalú modell, amely előre meghatározott számú Gauss függvény kombinációjaként épül fel.
- Kiselőadásod alatt ismertesd a felépítését és működését, valamint, hogy milyen típusú feladatokra/problémákra nyújt megoldást. Milyen gépi tanulási eljárás során szokták használni?

Active learning

- Felügyelt gépi tanulás esetén mindig valamilyen tanító adathalmazt használunk, azonban elképzelhető, hogy a valós címkék nem állnak rendelkezésünkre, és azokat időigényes, nehéz, esetleg drága előállítani. Ebben az esetben segíthet az aktív tanulás, melynek lényege, hogy a tanulórendszer maga választja ki, hogy mely mintákból szeretne tanulni, egy iteratív folyamat során.
- Részletesen ismertesd az aktív tanulás működését. Milyen különböző konfigurációkban képzelhető el a tanító halmaz kiválasztás? Milyen részfeladatai vannak az aktív tanulásnak és hogyan, milyen módszerek használhatók a kiválasztott tanító adathalmaz meghatározására?

Expectation Maximization

- Az EM egy iteratív, két lépéses, optimalizációs módszer. Az egyik lépés az expectation (E), a másik pedig a maximization (M).
- Részletesen ismertesd az eljárást működését. Milyen gépi tanulási módszerek részeként használjuk? Milyen előnyei és hátrányai vannak? Mutass be példákat.

XGBoosting

- Az XGBoosting egy olyan döntési fákra épülő technika, amely tartalmaz regularizációt, már az egyes döntési fák építése során.
- Ismerted az algoritmus működését. Mi az a regularizáció és miért hasznos? Milyen problémákra használható eredményesen?

Gradient descent

- A Gradient descent több mesterséges intelligencia algoritmus alapja. Tulajdonképpen egy eljárás arra, hogyan jussunk el egy konkrét állapotból egy kedvezőbb állapotra.
- Ismertesd az algoritmus működését. Milyen machine learning algoritmusok használják és melyik lépésben?

Recurrent Neural Network

- A recurrent neural network egy olyan neurális háló, amely szekvenciális adatokra illeszthető eredményesen, mert rendelkezik memóriával.
- Ismertesd a recurrent neural network működését. Milyen problémákra használható eredményesen? Mutass be példákat.

Convolutional Neural Network (elkelt)

- A Convolutional Neural Network egy főleg képfelismeréshez használt neurális háló. Működése során a konvolúciós szűrő segítségével a bementet tömöríti a neurális háló számára.
- Ismertesd a convolutional neural network működését. Milyen problémákra használható eredményesen? Mutass be példákat.

Generative Adversarial Networks (elkelt)

- A GAN-ok két versengő hálózatot valósítanak meg, ahol az egyik fél megpróbál minél jobb és pontosabb generált adatokat szolgáltatni, míg a másik megpróbálja eldönteni a kapott adatról, hogy az generált vagy „valódi”. A versengés során az egyik egyre jobb generáló, a másik egyre jobb megkülönböztető lesz.
- Részletesen ismertesd a generative adversarial network működését. Milyen részfeladatai vannak és milyen területeken használhatók? Mutass be példákat.

Zero-Shot Classification (elkelt)

- A ZSL (Zero-Shot Learning) egy olyan speciális eset, amikor adott kategóriákba egyetlen tanító minta sem tartozik.
- Ismertesd a Zero-Shot Learning és Zero-Shot Classification lényegét, típusait, és a szakirodalomban elterjedt technikákat az ilyen jellegű feladatok megoldására!

Viola-Jones detektor (elkelt)

- A Viola-Jones detektor egy Haar jellemzőkön alapuló kaszkádolt osztályozó AdaBoost tanulóalgoritmussal. A módszer legelterjedtebb alkalmazási példája az arcdetektálás.
- Kiselőadásodban részletesen ismertesd a Viola-Jones detektor működését, kitérve az integrálkép, az AdaBoost tanulóalgoritmus és a kaszkádolt struktúra építésének módszereire!

tf-idf

- A tf-idf (term frequency – inverse document frequency) egy kifejezés fontosságának mérésre használható adott dokumentumkészlet mellett.
- Ismertesd a tf-idf kiszámításának módját, előnyeit, hátrányait, illetve azt, hogy milyen eljárásokban szokás ezt használni! Mutass példát a működésére!

SIFT

- A SIFT (Scale Invariant Feature Transform) képi tartalmak kulcspontjainak meghatározására és a kulcspontokban jellemző vektorok kinyerésére kifejlesztett módszer.
- Ismertesd a SIFT működését részletesen! A lépéseket ábrákkal, példákkal illusztrálva mutasd be!