

2. alkalom – A gépi tanulás: elmélet

Informatikai és távközlési alapok II. · 12. Erbium

Programozás vagy gépi tanulás?

	Hagyományos programozás	Gépi tanulás
Ki írja a szabályt?	az ember	a gép tanulja meg példákból
Mi megy be?	szabály + adat	adat + a helyes válaszok
Mi jön ki?	eredmény	szabály (modell) , amely új adatra is választ ad
Példa	„ha a levélben szerepel az <i>ingyen</i> szó, spam”	a gép sok ezer megjelölt levélből maga jön rá, mi jellemző a spamre

A gépi tanulás fogalmai

Fogalom	Jelentése	Példa
Tanító adat	a példák, amelyekből a gép tanul	10 000 kép állatokról
Címke	a helyes válasz egy példához	„macska”, „kutya”
Modell	a megtanult „szabály”	a program, amely egy új képre megmondja, mi van rajta
Tesztelés	a modell kipróbálása olyan adaton, amelyet tanulás közben nem látott	1000 új kép, ismert válasszal

A tanulás három fajtája

Fajta	Hogyan tanul?	Példa
Felügyelt	a tanító adatban benne van a helyes válasz (címke)	spamszűrő, arcfelismerés
Felügyelet nélküli	nincs címke: a gép maga keres csoportokat, mintázatokat	a webáruház vásárlóinak csoportosítása
Megerősítéses	próbálkozik, és jutalmat vagy büntetést kap	sakk- és go-programok, robotok

Hol találkozol vele?

beszéd- és arcfelismerés · fordítóprogramok · **ajánlórendszerek** (YouTube, TikTok, Netflix: a nézési szokásaidból tanulnak) · **célzott reklámok** · spamszűrő · navigáció (a forgalom előrejelzése)

Veszélyek, amelyekre figyelni kell

- **Torzítás (bias):** ha a tanító adat egyoldalú, a modell is egyoldalúan dönt. Például ha egy arcfelismerőt főleg egyféle arcokon tanítottak, a többin rosszabbul működik.
- **Információs buborék:** az ajánló csak azt mutatja, amit már szeretsz, így egyre kevesebb mást láatsz.
- **Adatvédelem:** a modellek rólunk gyűjtött adatokból tanulnak (mit nézünk, hol vagyunk, mit vásárolunk).
- **A modell is tévedhet:** amit sosem látott (egy új fajta példa), azt rosszul sorolhatja be.

Fogalmak röviden

gépi tanulás · tanító adat · címke · modell · tesztelés · felügyelt, felügyelet nélküli és megerősítéses tanulás
· ajánlórendszer · torzítás · információs buborék

A2 – Emberi osztályozó: így tanul a gép

Név: _____ Pár: _____

A feladat

Ti vagytok a „gépi tanulás”. Az 1–10. gyümölcs **címkézett** tanító adat. Találjatok olyan szabályt, amellyel be lehet sorolni őket, aztán használjátok a szabályt az A–F gyümölcsökre!

Tanító adat (címkével)

#	Szín	Méret	Tömeg	Címke
1	piros	kicsi	5 g	cseresznye
2	sárga	hosszú	120 g	banán
3	piros	közepes	150 g	alma
4	zöld	közepes	160 g	alma
5	narancssárga	közepes	180 g	narancs
6	sárga	hosszú	130 g	banán
7	piros	kicsi	6 g	cseresznye
8	zöld	nagy	3000 g	dinnye
9	narancssárga	közepes	200 g	narancs
10	zöld	nagy	4000 g	dinnye

A szabályotok (a „modell”)

Új adat (címké nélkül) – soroljátok be!

#	Szín	Méret	Tömeg	A ti címketek
A	sárga	hosszú	110 g	
B	piros	kicsi	4 g	
C	zöld	közepes	140 g	
D	zöld	nagy	3500 g	
E	narancssárga	közepes	190 g	
F	sárga	közepes	150 g	

Kérdések

1. Mit mondott a modelletek az F-re? Biztosan jó a válasz? Miért nem?

2. Mit kellene tenni, hogy a modell az F-et is jól felismerje?

3. Felügyelt vagy felügyelet nélküli tanulás volt ez? Miért?

Kilépő

Fejezd be a mondatot: „A gépi tanulás az, amikor...”
