

# Üzleti Elemzések Módszertana

## 5. Előadás: Együttes tanulás

Kuknyó Dániel  
Budapesti Gazdasági Egyetem

2023/24  
2.félév

1 Bevezetés

2 Adaptív turbózás

3 Gradiens turbózás

1 Bevezetés

2 Adaptív turbózás

3 Gradiens turbózás

# Az együttes tanulás mögötti intuíció

- Egy gazda szeretné lemérni, milyen a hőmérséklet a szőlős birtokán.
- A birtok egy hegyoldalban fekszik, ezért a szőlőtőkéket eltérő időjárási hatások érik.



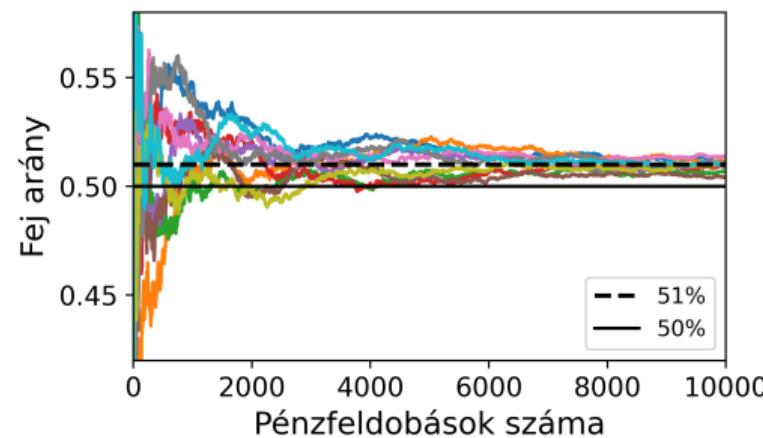
## Példa: szavazó osztályozók

A következő példában 10 osztályozó modell feladata, hogy megbecsüljék, melyik oldalára fog esni egy torzított pénzérme.

A pénzérme 51% valószínűsséggel esik fejre, 49% eséllyel pedig írásra.

1000 dobás után 75%, hogy a modellek valószínűsége fejt fog szavazni. Ugyanez a valószínűség 10000 dobás után 97%.

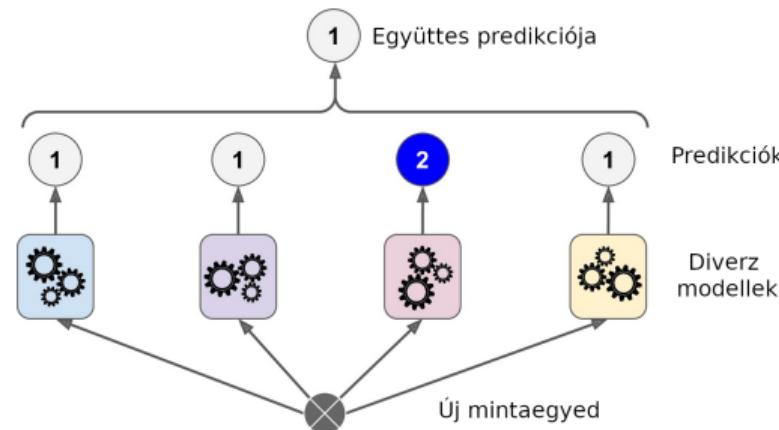
A szavazó osztályozók nagyobb pontosságot érnek el együttesen, mint a modellcsoport bármelyik tagja.



# Szavazó osztályozók

A szavazó osztályozó kifejezés modellek egy csoportjára utal, amelyben a modellek **egymástól függetlenül képesek predikciót adni** egy adott mintaegyedre vonatkozóan.

A szavazó osztályozó a végső predikciót úgy állítja elő, hogy **a benne lévő modellek predikciójait aggregálja** valamilyen módszertan szerint pl. kiválasztja belőle a leggyakoribb elemet.

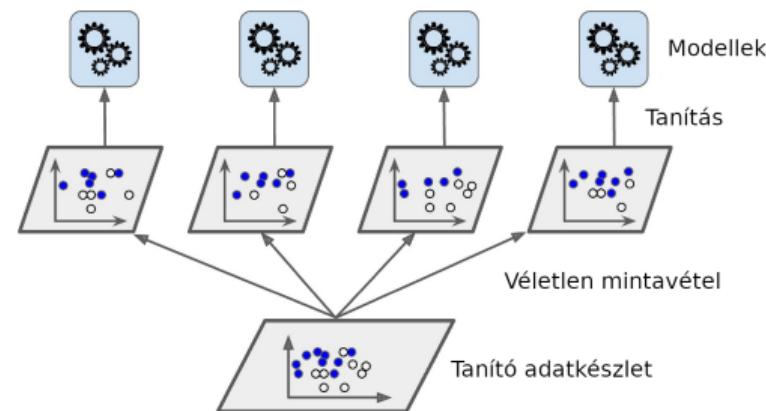


# Bagging és Pasting

Az együttes tanuló modellek taníthatók az adathalmaz különböző részhalmazain. Ez robusztusabb modellt fog eredményezni, ami jobb általánosító képességeket jelent éles felhasználásban.

## Bagging

Együttes tanulási módszer, melyben a modellek **visszatevés nélküli mintavétellel** kapják meg a saját tanító mintájukat.

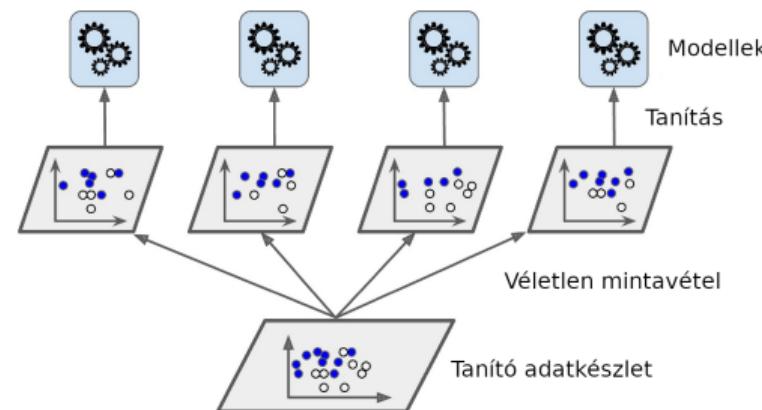


# Bagging és Pasting

Az együttes tanuló modellek taníthatók az adathalmaz különböző részhalmazain. Ez robusztusabb modellt fog eredményezni, ami jobb általánosító képességeket jelent éles felhasználásban.

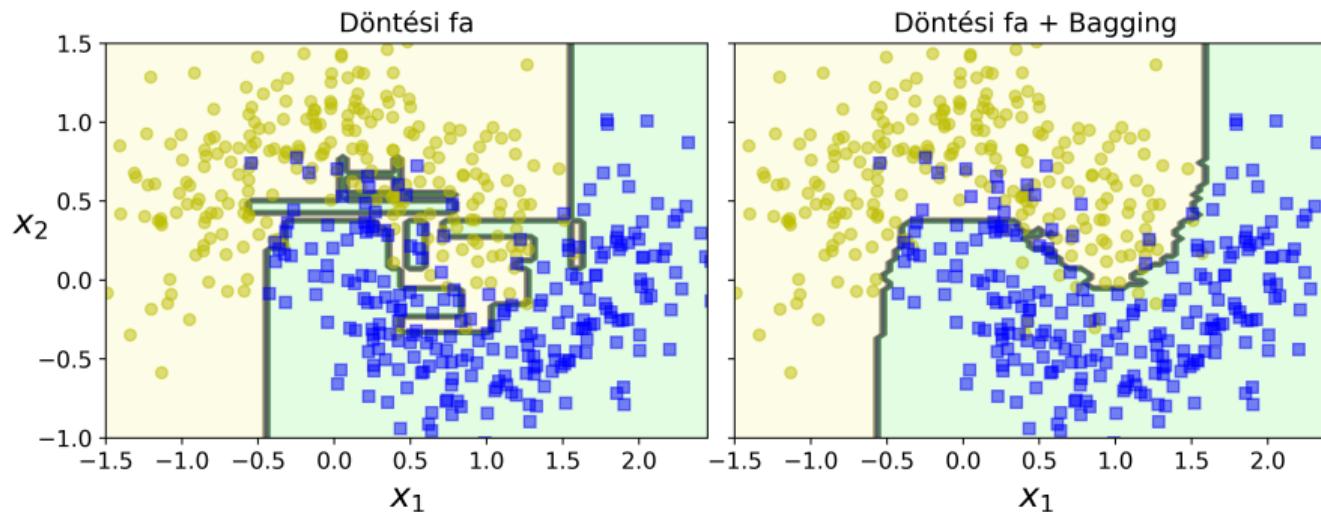
## Pasting

Együttes tanulási módszer, melyben a modellek **visszatevéses mintavétellel** kapják meg a saját tanító mintájukat.



## Bagging hatása a döntési határokra

A bal oldali ábrán egyetlen döntési fa határai láthatóak, a jobb oldalon pedig több bagging technikával tanított döntési fának a határait. A bagging modellcsoport egyértelműen jobb általánosító képességekkel rendelkezik mint a döntési fa.



1 Bevezetés

2 Adaptív turbózás

3 Gradiens turbózás

# Adaptív turbózás

## Adaptív turbózás

Az együttes tanulás súlyozott változata. A modellek szekvenciálisan állnak elő olyan módon, hogy az új modell mindenkorban tanul az elődje hibájából.

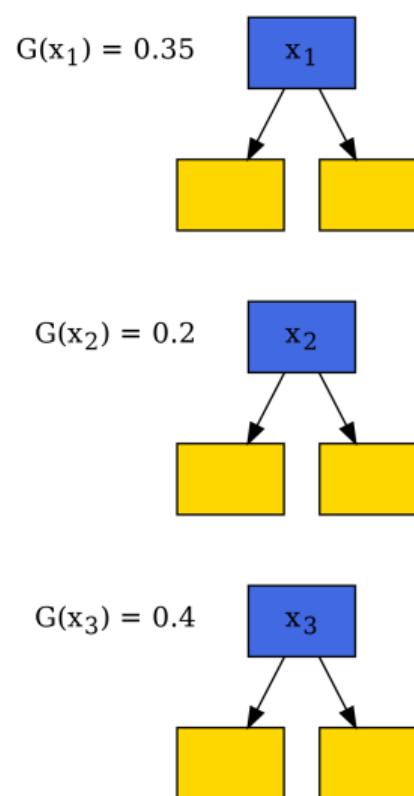
### Algoritmus 1: Adaptív turbózás

- ① Kezdeti súlyok rekordokhoz rendelése
- ② Modell illesztés minden változóra
- ③ Legjobb modell kiválasztása
- ④ Modell súlyozása
- ⑤ Egyedsúlyok frissítése
- ⑥ Adathalmaz újramintázása
- ⑦ Iteráció kilépésig

# Adaptív turbózás: lépésről lépésre

Az Adaptív turbózás algoritmusa:

- 1 Az algoritmus minden mintaegyedhez  $w = \frac{1}{n}$  kezdeti súlyt rendel, ahol  $n$  az összes a minta halmaz mérete.
- 2 minden  $x$  változóra egy döntési tönk kerül illesztésre. Ez egy lineáris döntési határ minden változóra.
- 3 Az a döntési tönk kerül kiválasztásra, amely a legjobban képes szeparálni az egyedekeket. Ez ebben az esetben az  $x_2$  változóhoz tartozó modell.



# Adaptív turbózás: lépésről lépésre

- 4 A modell súlyozása a teljes modellcsoportban. Egy tönk hibája azon mintaegyedek súlyainak összege, amelyeket helytelenül osztályozott:

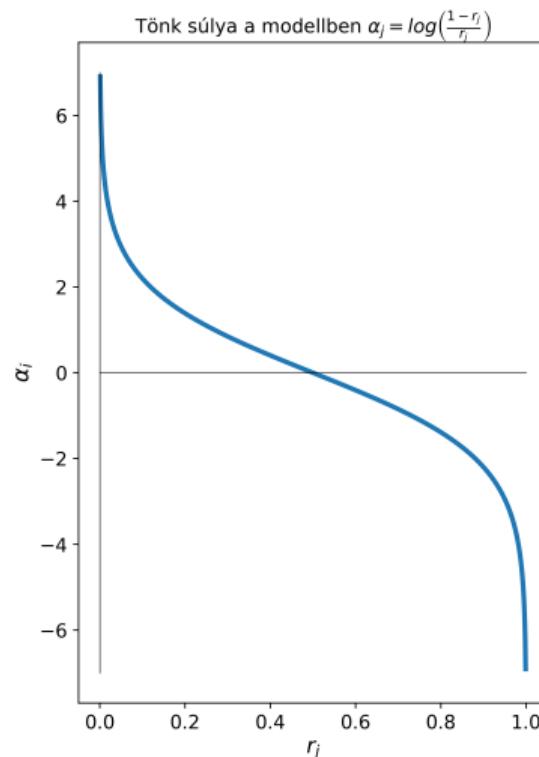
Prediktor teljes hibája

$$r_j = \frac{\sum_{\substack{i=1 \\ \hat{y}_j \neq y_i}}^m w_i}{\sum_{i=1}^m w_i}$$

Tönk súlya a modellben

$$W_j = \alpha \cdot \log \left( \frac{1 - r_j}{r_j} \right)$$

Ahol  $\alpha$  a tanulási sebesség.



# Adaptív turbózás: lépésről lépésre

- ⑤ Az eljárás frissíti a mintaegyed súlyait.

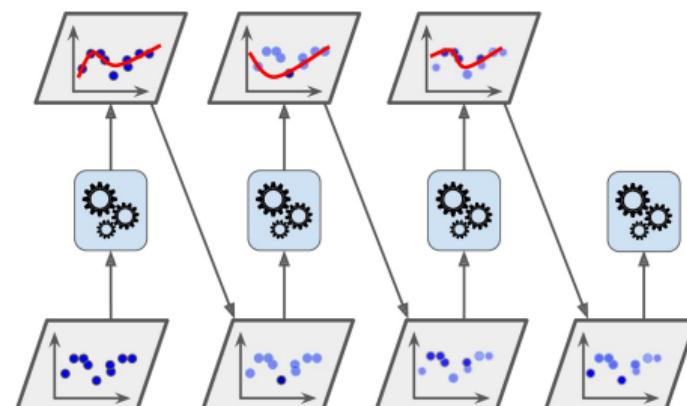
Minden helyesen beosztályozott mintaegyedre:

$$w \leftarrow w \cdot e^{W_j}$$

Minden helyesen beosztályozott mintaegyedre:

$$w \leftarrow w \cdot e^{-W_j}$$

- ⑥ A létrejött egyedsúlyoknak megfelelő valószínűség eloszlást felhasználva az adathalmaz újramintázódik és eszerint kapja meg a következő fa a saját mintáját.
- ⑦ Az algoritmus a folyamatot addig ismétli, amíg el nem éri a kilépési kritériumot.

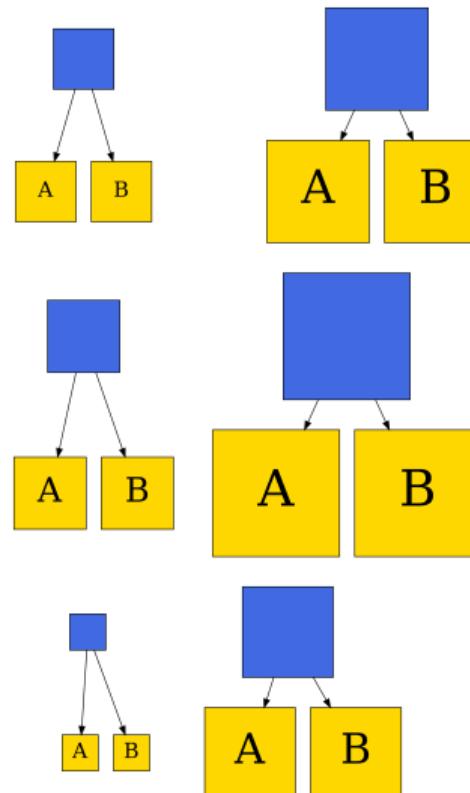


# Az osztályozás eljárása

Osztályozáskor minden létrejött döntési tönk létrehozza a saját predikcióját, és a különböző prediktorok szavazása által dől el, mi legyen a végső predikció.

Az egyed abba az osztályba lesz besorolva, amelyikhez tartozó tönök súlyának összege a legnagyobb.

$$\hat{y} = \operatorname{argmax}_k \sum_{\substack{j=1 \\ \hat{y}_j(x)=k}}^N \alpha_j$$



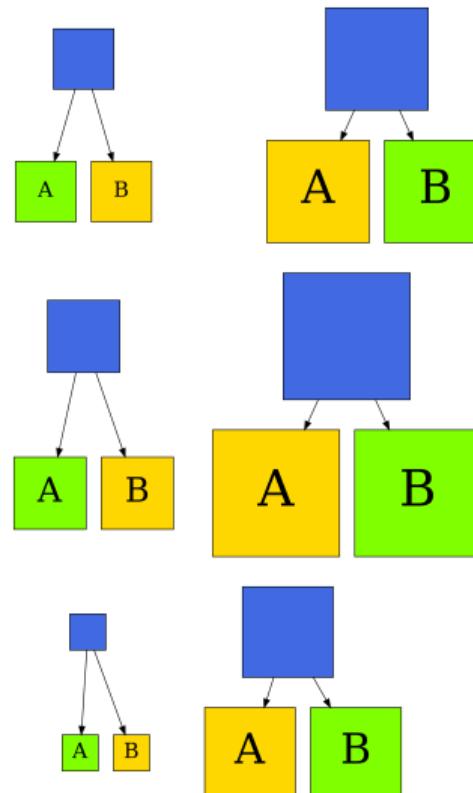
# Az osztályozás eljárása

Osztályozáskor minden létrejött döntési tönk létrehozza a saját predikcióját, és a különböző prediktorok szavazása által dől el, mi legyen a végső predikció.

Az egyed abba az osztályba lesz besorolva, amelyikhez tartozó tönök súlyának összege a legnagyobb.

$$\hat{y} = \operatorname{argmax}_k \sum_{\substack{j=1 \\ \hat{y}_j(x)=k}}^N \alpha_j$$

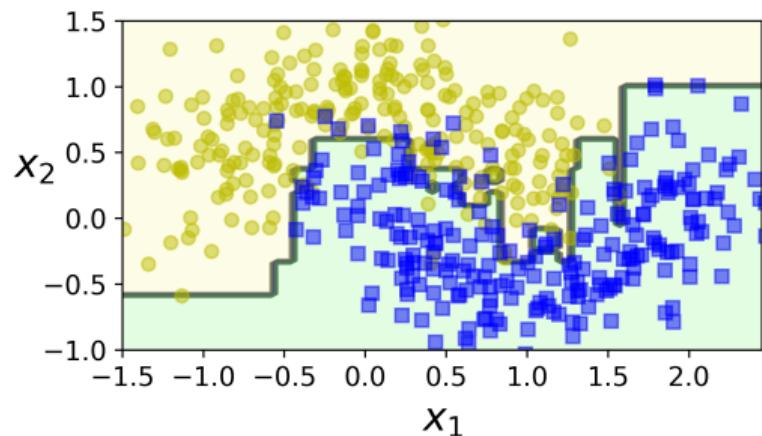
A példában:  $\hat{y} = B$



# Adaptív turbózás a Moons halmazon

A `make-moons` könyvtár egy nemlineárisan szeparálható adathalmazt generál.

A következő ábrán egy döntési fa alapú adaptív turbózó döntési határai láthatók.



1 Bevezetés

2 Adaptív turbózás

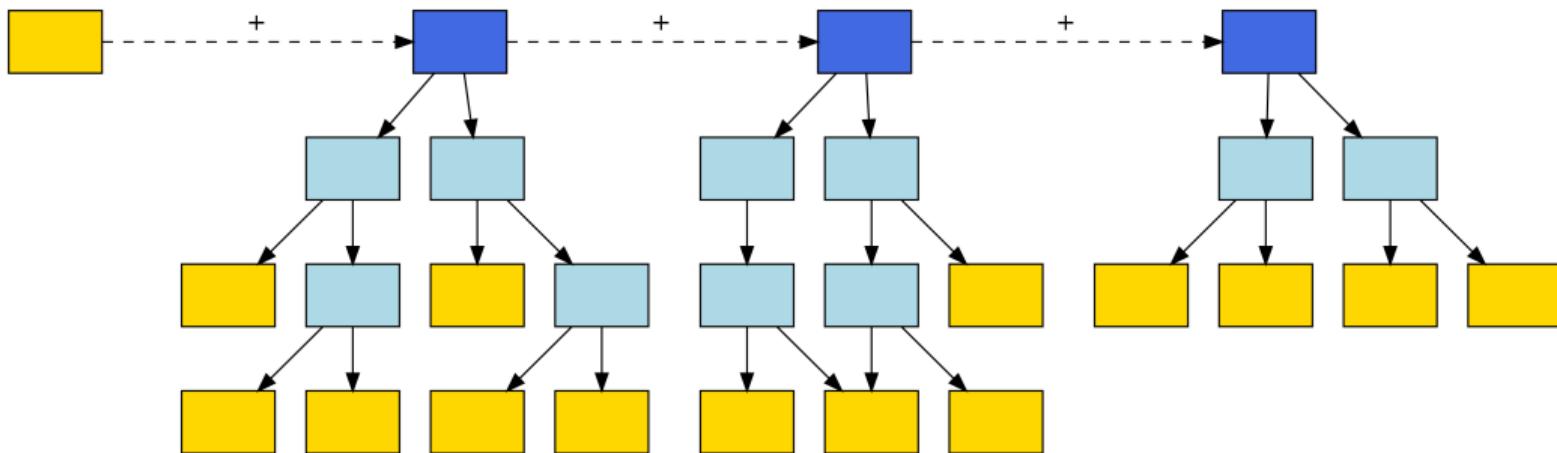
3 Gradiens turbózás

# Gradiens turbózás (GBDT)

Együttes tanuló algoritmus, amely a rezidumok szekvenciális javításával állítja elő a becsült értéket. minden újonnan létrejövő modell javít az elődje hibáján.

A becsült érték előállítása:

$$\text{Becsült érték} = \text{előző predikció} + \text{tanulási sebesség} * \text{reziduum}$$



# Gradiens turbózás (GBDT)

Együttes tanuló algoritmus, amely a rezidumok szekvenciális javításával állítja elő a becsült értéket. minden újonnan létrejövő modell javít az elődje hibáján.

---

## Algoritmus 2: Gradiens turbózás

---

- ① Kezdeti predikció a célváltozóra
  - ② Rezidumok kiszámítása
  - ③ Döntési fa illesztése a rezidumokra
  - ④ Levelek output értékének kiszámítása
  - ⑤ Predikció a minta adathalmazra
  - ⑥ 2-5 lépés iterálása kilépésig
-

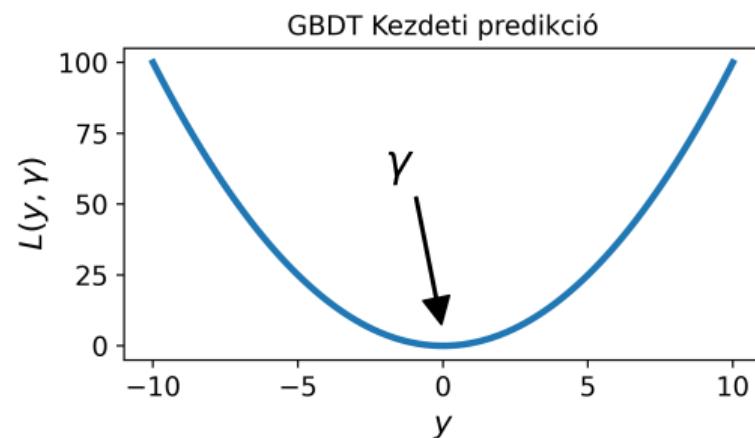
# Gradiens turbózás: lépésről lépésre

- ① A GBDT egy adott  $y$  célváltozóra vonatkozó első predikciója a célváltozó várható értéke:

$$F_0(x) = \underset{\gamma}{\operatorname{argmin}} \sum_{i=1}^n L(y_i, \gamma)$$

Ahol:

- $L(y, \gamma)$ : Deriválható költségfüggvény
- $\gamma$ : Az az érték, ami minimalizálja a mintaegyedekre kiszámolt reziduumok összegét



# Gradiens turbózás: lépésről lépésre

- ② Következőnek az algoritmus kiszámolja a rezidumokat minden mintaegyedre. A rezidum a becsült és valós érték különbsége adott  $L$  költségfüggvény szerint:

## Rezidum

$$r_{i,m} = - \left[ \frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)}$$

Ahol:

- $i$ : A minta indexe
- $m$ : A modell indexe

Magasság	Szemszín	Nem	Súly	Rezidum
1.6	Kék	Férfi	88	16.8
1.6	Barna	Nő	76	4.8
1.5	Kék	Nő	56	-15.2
1.8	Zöld	Férfi	73	1.8
1.5	Barna	Férfi	77	5.8
1.4	Kék	Nő	57	-14.2

$$F_0(x) = \frac{1}{6} (88 + 76 + 56 + 73 + 77 + 57) = 71.2$$

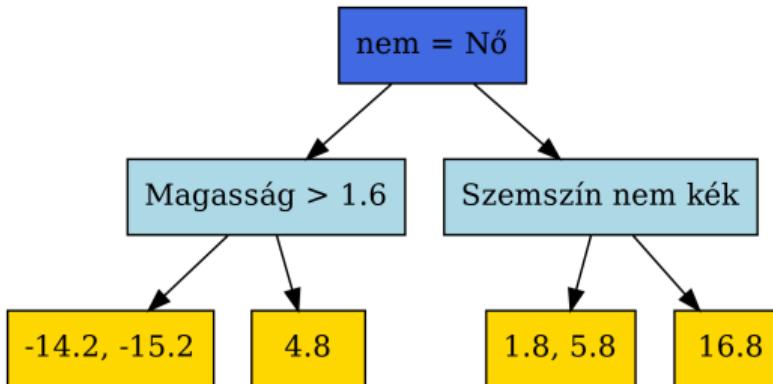
$$r_{1,1} = 88 - 71.2 = 16.8$$

$$r_{1,2} = 76 - 71.2 = 4.8$$

# Gradiens turbózás: lépésről lépésre

## ③ Döntési fa illesztése a rezidumokra.

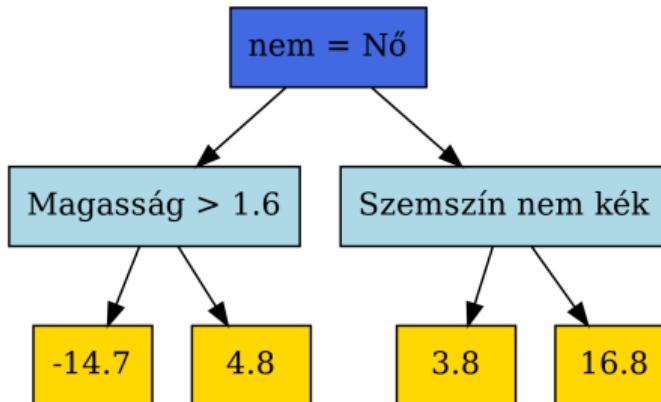
Ebben a lépésben az algoritmus besorolja a rezidumokat egy döntési fa leveleibe.



Magasság	Szemszín	Nem	Súly	Rezidum
1.6	Kék	Férfi	88	16.8
1.6	Barna	Nő	76	4.8
1.5	Kék	Nő	56	-15.2
1.8	Zöld	Férfi	73	1.8
1.5	Barna	Férfi	77	5.8
1.4	Kék	Nő	57	-14.2

# Gradiens turbózás: lépésről lépésre

- ④ Levelek output értékének kiszámítása:  
a levelek outputja az az érték, ami minimalizálja a levélbe bekerült értékekre a költségfüggvényt. Ez az esetek többségében a reziduumok átlaga.



## Levél output

$$\gamma_{j,m} = \underset{\gamma}{\operatorname{argmin}} \sum_{x_i \in R_{i,j}} L(y_i, F_{m-1}(x_i) + \gamma)$$

Ahol:

- $\gamma$ : A levél outputja
- $F_{m-1}x_i$ : A modell előző fája által adott predikció  $x_i$  egyedre.
- $y_i$ : A célváltozó valós értéke

# Gradiens turbózás: lépésről lépésre

- 5 Predikciók készítése a minta adathalmazra

## GBDT Predikció

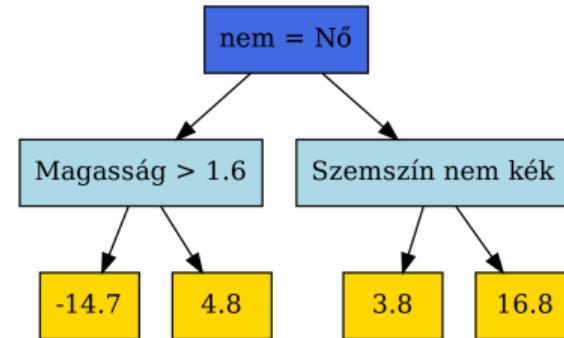
$$F_m(x) = F_{m-1}(x) + \alpha \sum_{j=1}^{J_m} \gamma_{j,m} \quad x \in R_{j,m}$$

Ahol:

- $F_{m-1}(x)$ : Az  $x$  mintaegyedre az előző fa által adott becsült érték
- $\alpha$ : A tanulási sebesség
- $J$ : A levelek száma

A predikció az első mintaegyedre  $\alpha = 0.1$  tanulási sebességgel:

Magasság	Szemszín	Nem	Súly	Rezidum
1.6	Kék	Férfi	88	16.8



$$F_0(x) = 71.2$$

$$F_1(x) = 71.2 + 0.1 \cdot 16.8 = 72.9$$

# Gradiens turbózás: lépésről lépésre

Az  $F_1(x)$  predikciója alapján a következő rezidumok minden mintaegyedre:

Magasság	Szemszín	Nem	Súly	Rezidum $r_{i,1}$	Rezidum $r_{i,2}$
1.6	Kék	Férfi	88	16.8	15.1
1.6	Barna	Nő	76	4.8	4.3
1.5	Kék	Nő	56	-15.2	-13.7
1.8	Zöld	Férfi	73	1.8	1.4
1.5	Barna	Férfi	77	5.8	5.4
1.4	Kék	Nő	57	-14.2	-12.7

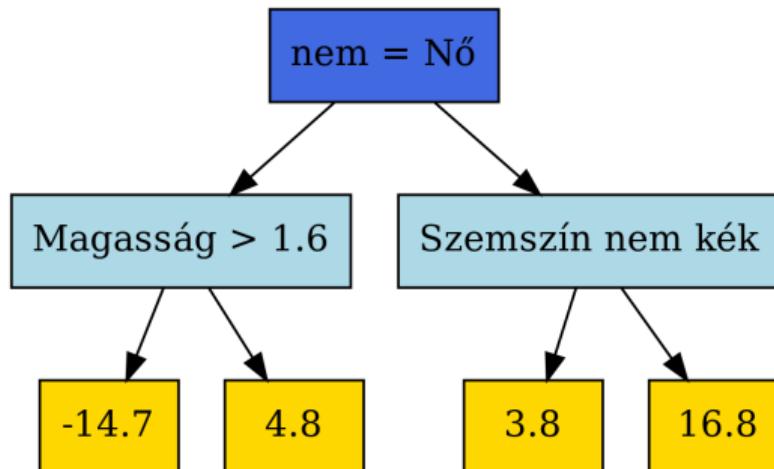
$$r_{1,2} = 88 - (71.2 + 0.1 \cdot 16.8) = 15.1$$

$$r_{2,2} = 76 - (71.2 + 0.1 \cdot 4.8) = 4.3$$

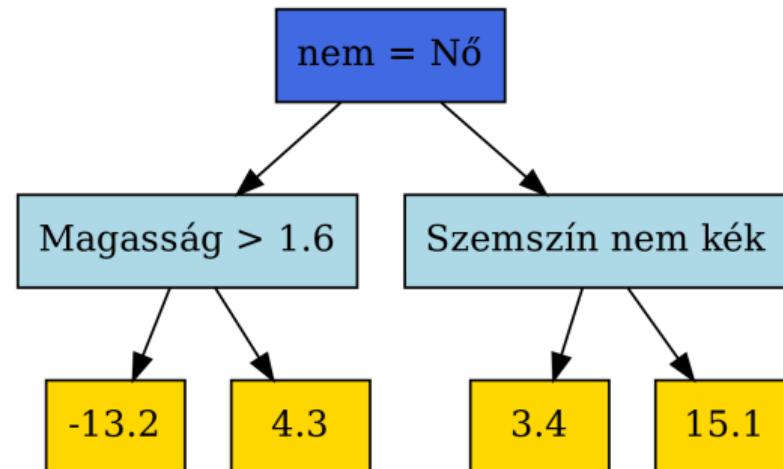
# Gradiens turbózás: lépésről lépésre

Miután létrejöttek az  $r_{i,2}$  reziduumok ismételten besorolódnak egy döntési fába.

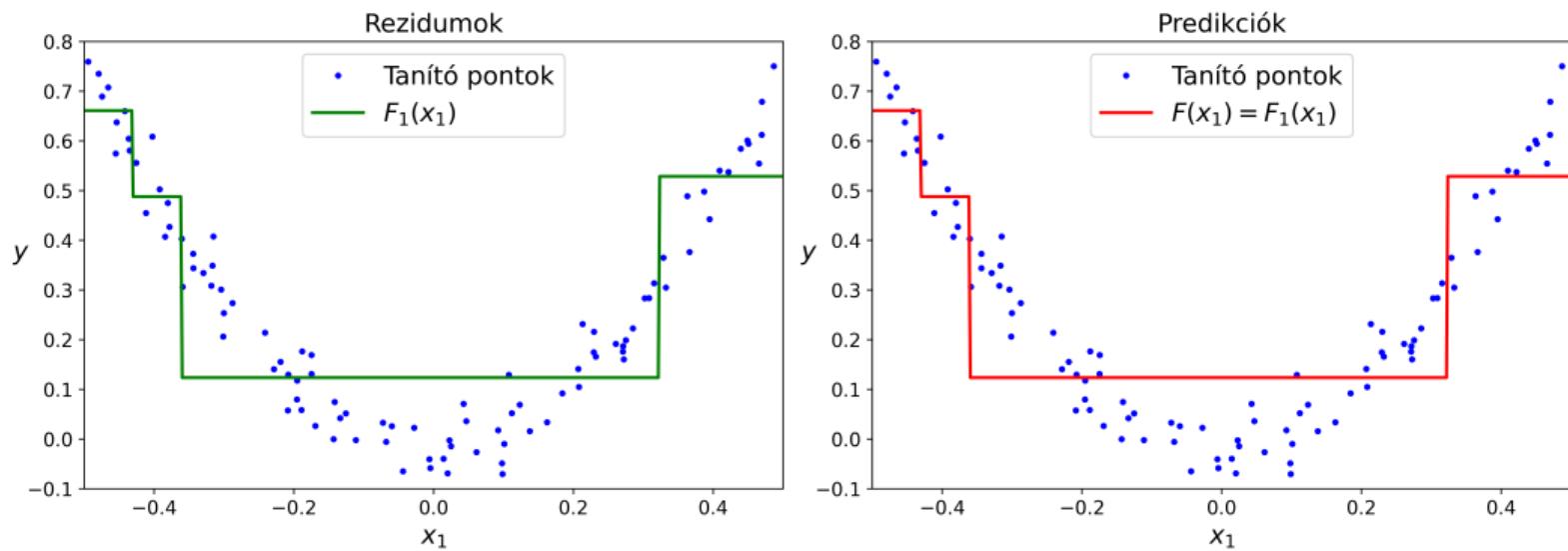
Az első döntési fa:



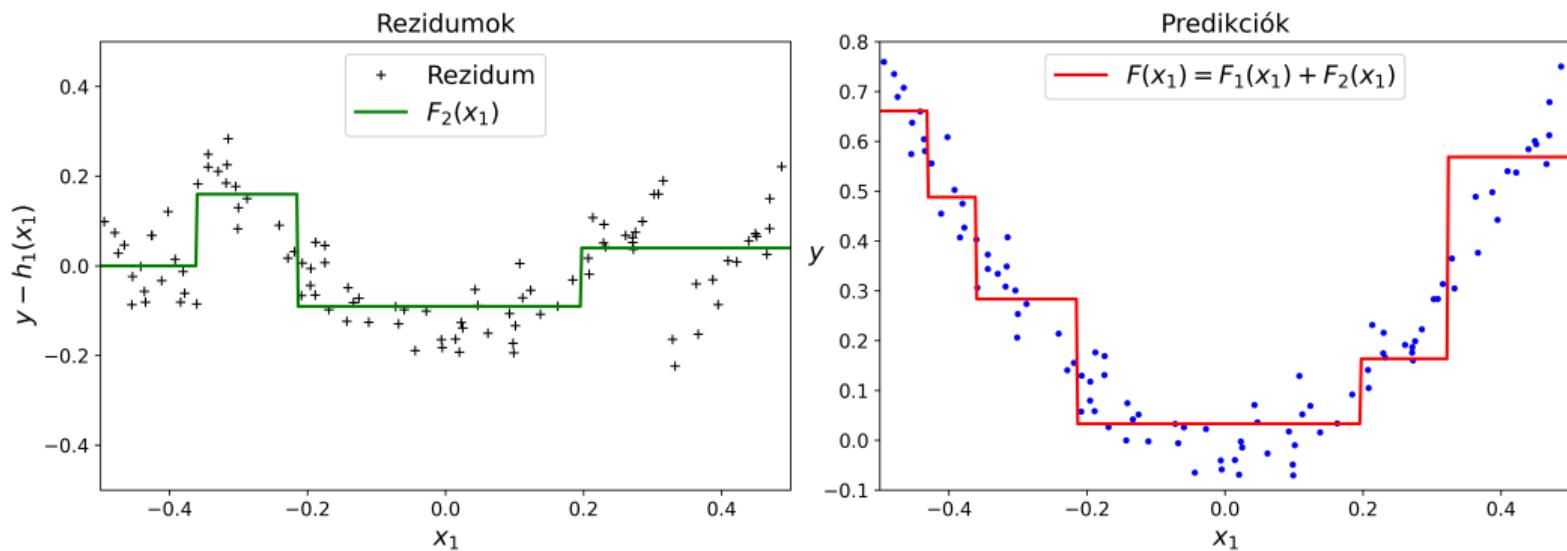
A második döntési fa:



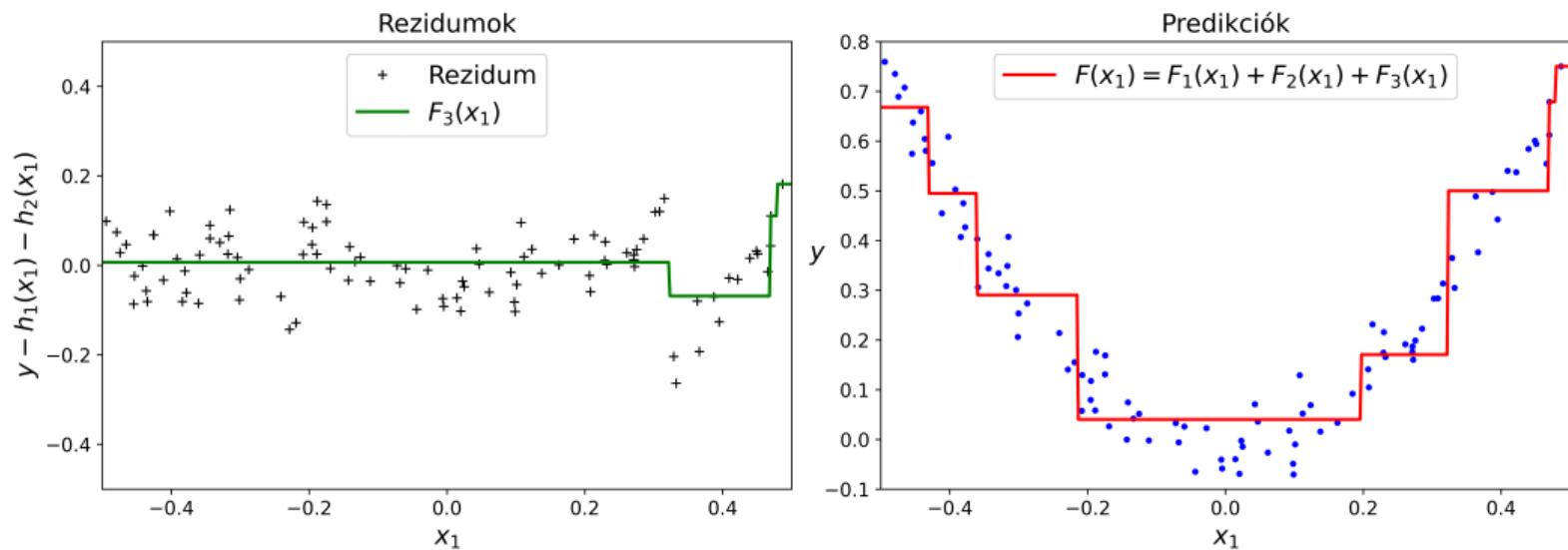
## Tanítás lépései egy minta adathalmazon



## Tanítás lépései egy minta adathalmazon



## Tanítás lépései egy minta adathalmazon

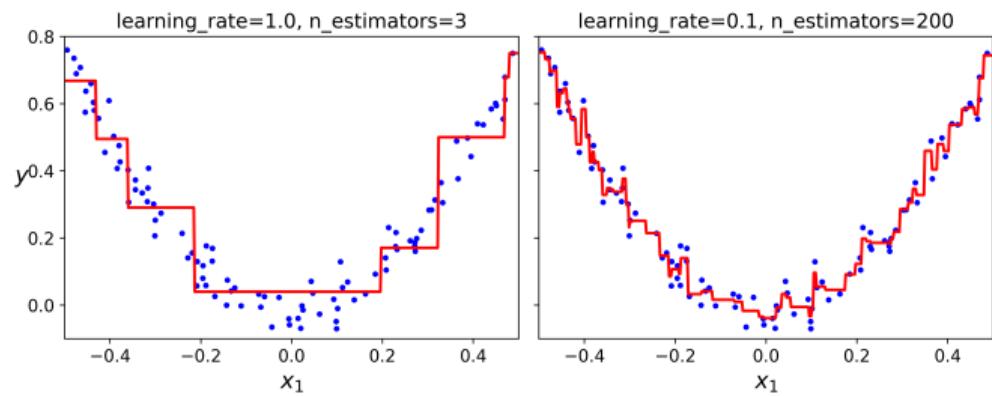


# Alultanulás és túltanulás

A turbózott regresszor fák is hajlamosak a túltanulásra.

A tanító feladata felismerni a túltanulás jelenségét és vagy **az egyéni fák korlátozásával, vagy a tanítási időszakok rövidítésével regularizálni a modellcsoportot.**

A bal oldali ábrán egy alultanult, a jobb oldalon pedig egy túltanult modell predikciói láthatók.



# Korai leállás gradiens turbózással

Korai leállás implementálása segíthet elkerülni a túltanulást. Ilyenkor az a modell lesz a végleges, amelyik a legkisebb validációs hibát érte el. Ha a validációs hiba emelkedőn van, a tanító iteráció kilép.

