Gépi tanulás a gyakorlatban

Lineáris regresszió

Lineáris Regresszió

- Legyen adott egy tanuló adatbázis:

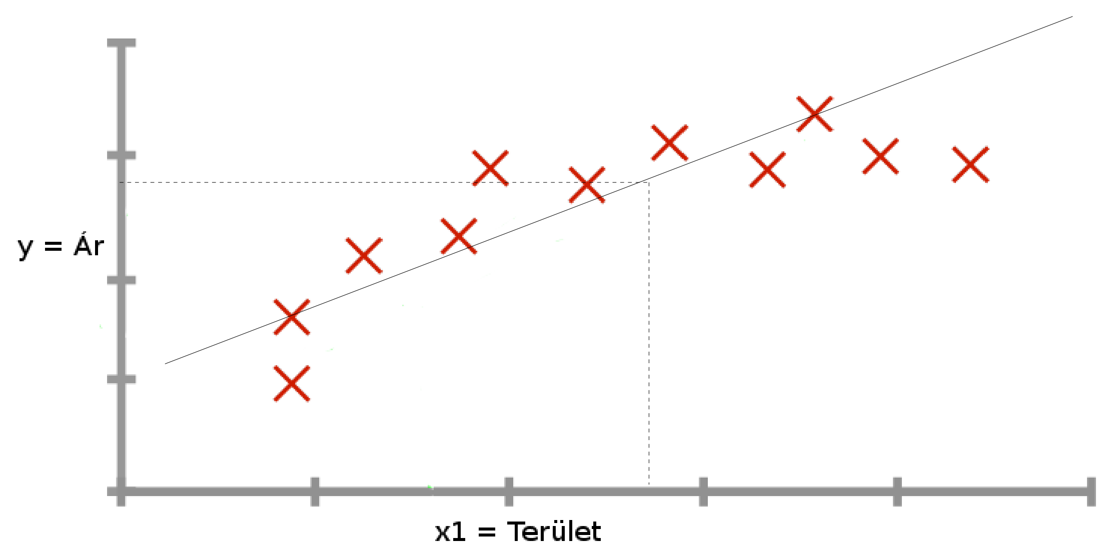
  Rendelkezésünkre áll egy olyan előfeldolgozott adathalmaz, aminek sorai az egyes ingatlanokat írják le. Az adathalmaznak két oszlopa van.

  - Az első az *ingatlanok területét*,
  - a második az *ingatlanok árait tartalmazza*.

  Szeretnénk *előrejelezni ingatlanok árait*.

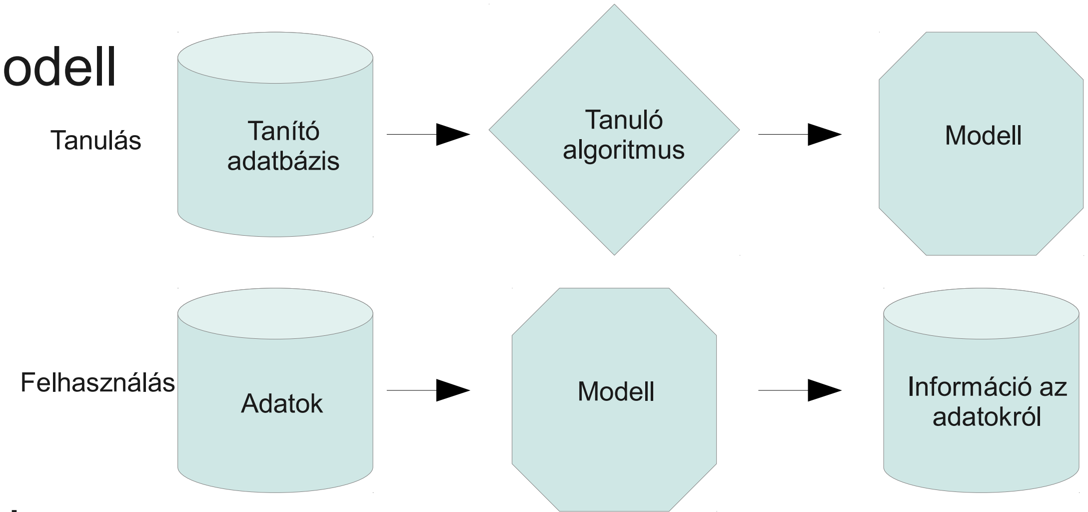
Felügyelt módszer, *valós* segéd információval → regressziós feladat

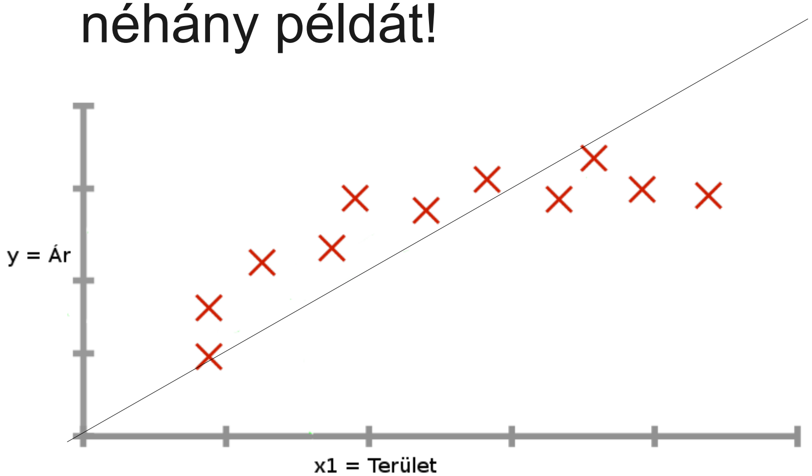
- Tanuló példák száma: m=11
- Dimenzió: d=1 (bemeneti pontok dimenziója)
- Jellemzők: 1 db, terület (x1)
- Címkék: valós érték (ár) → valós érték → felügyelt regresszió (y)
- Modell: egyenes (h)



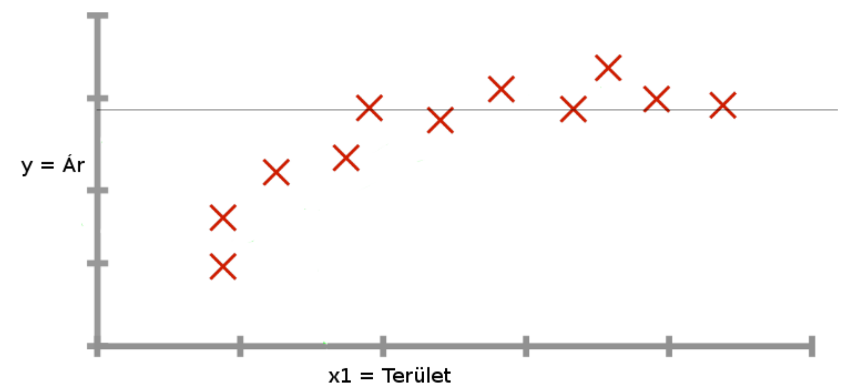
Lineáris regresszió

- Tanulás során: tanuló adatbázis → modell (h)
- Hogyan reprezentáljuk h-t?
- h(x)=a*x+b   ← *a* és *b* paraméterek
- Tanulás során: *a* és *b* paraméterek meghatározása
- Hogyan válasszuk *a*-t és *b*-t? Nézzünk néhány példát!





a=1, b=0

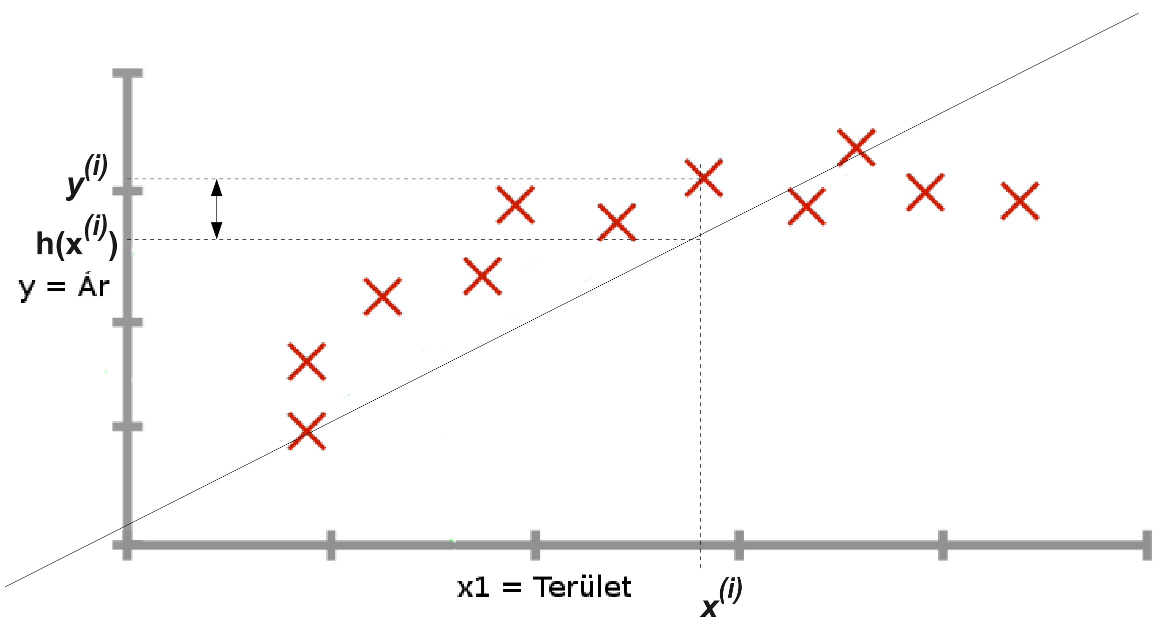


a=0, b=1.2m

Lineáris regresszió

- Hogyan válasszuk *a*-t és *b*-t (h modell paramétereit)?
- <u>Ötlet:</u> válasszuk úgy *a*-t és *b*-t, hogy a tanuló példákon h(*x*) *közel kerüljön* *y*-hoz, minden (*x*,*y*) tanító példára
- <u>Költség függvény:</u> $J(a,b) = \frac{1}{2m}\sum_{i=1}^{m}(h_{a,b}(x^{(i)}) - y^{(i)})^2$, ahol

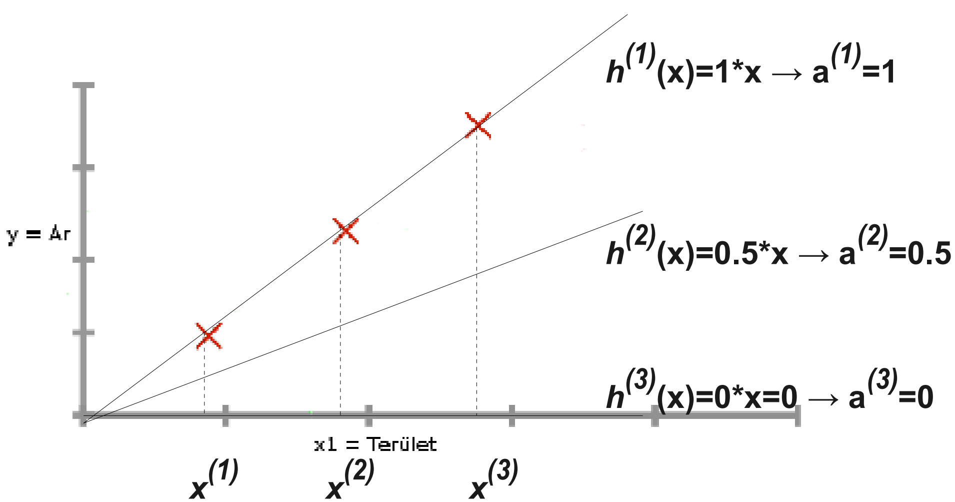
$$h_{a,b}(x^{(i)}) = h(x^{(i)}) = a \cdot x^{(i)} + b$$

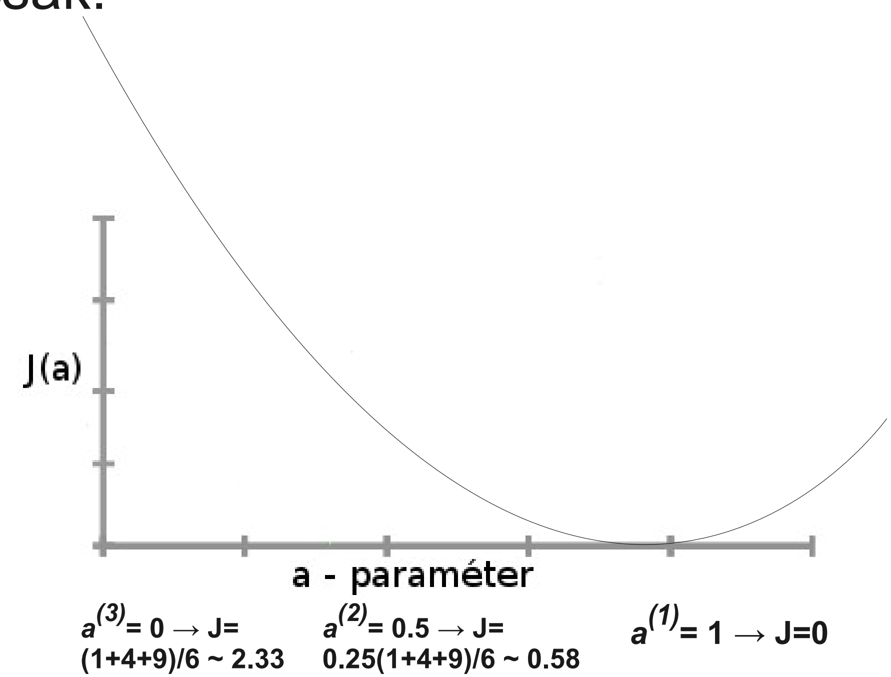


<u>Tanulás:</u> *minimalizáljuk J(a,b)*-t a tanuló adatbázison és a tanult modell legyen a minimum hely

Lineáris regresszió

- Értsük meg a költség függvényt!
  - a modell: $h_{a,b}(x)$ a *példák* (nem csak tanító példák!) függvénye, adott paraméterek esetén → területhez rendel becsült eladási árat
  - a költség függvény: $J(a,b)$ a *paraméterek* függvénye, adott tanuló halmaz esetén → megmondja, hogy az adott paraméterű modell mennyire jól illeszkedik a tanuló halmazra (hiba függvény)
- Egyszerűsítés: legyen $h(x) = a*x$ ← (hagyjuk el a b paramétert egy időre) → Origón átmenő egyeneseket tekintünk csak!





Lineáris regresszió

- Értsük meg a költség függvényt!
- Előzőek alapján sejtés: *J(a)* parabolikus függvény
- <u>Valóban:</u>

$$J(a) = \frac{1}{2m}\sum_{i=1}^{m}(h_a(x^{(i)}) - y^{(i)})^2 = \frac{1}{2m}\sum_{i=1}^{m}(a \cdot x^{(i)} - y^{(i)})^2 =$$

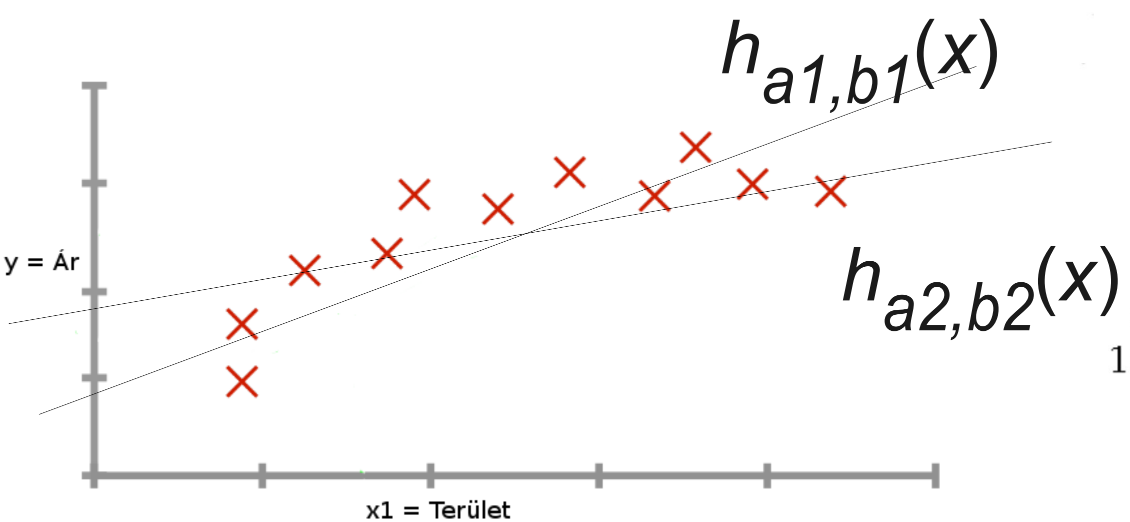
$$= \frac{1}{2m}\sum_{i=1}^{m}((x^{(i)})^2 \cdot a^2 - 2x^{(i)}y^{(i)} \cdot a + (y^{(i)})^2) =$$

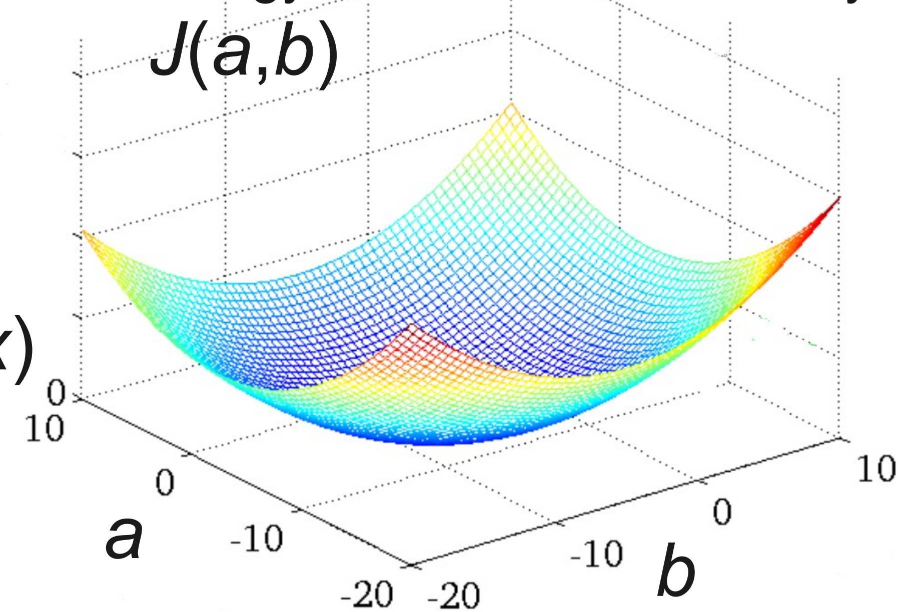
$$= \underbrace{\frac{1}{2m}\sum_{i=1}^{m}(x^{(i)})^2}_{A} \cdot a^2 - \underbrace{\frac{1}{m}\sum_{i=1}^{m}x^{(i)}y^{(i)}}_{B} \cdot a + \underbrace{\frac{1}{2m}\sum_{i=1}^{m}(y^{(i)})^2}_{C} =$$

$$= A \cdot a^2 - B \cdot a + C$$

Lineáris regresszió

- Az egyszerűsítés elvetésével is hasonló a helyzet:
  - <u>a modell:</u> $h_{a,b}(x)=a*x+b$ → nem csak az origón átmenő egyenesek → két paraméter →
  - <u>a költség függvény:</u> $J(a,b)$ *két változós* függvény → szintén parabolikus
- A költség függvény:
  - minden egyes alappontja meghatároz egy modellt
  - a hozzá tartozó függvény érték meghatározza, hogy az adott modell mennyire „jó” a tanuló halmazon

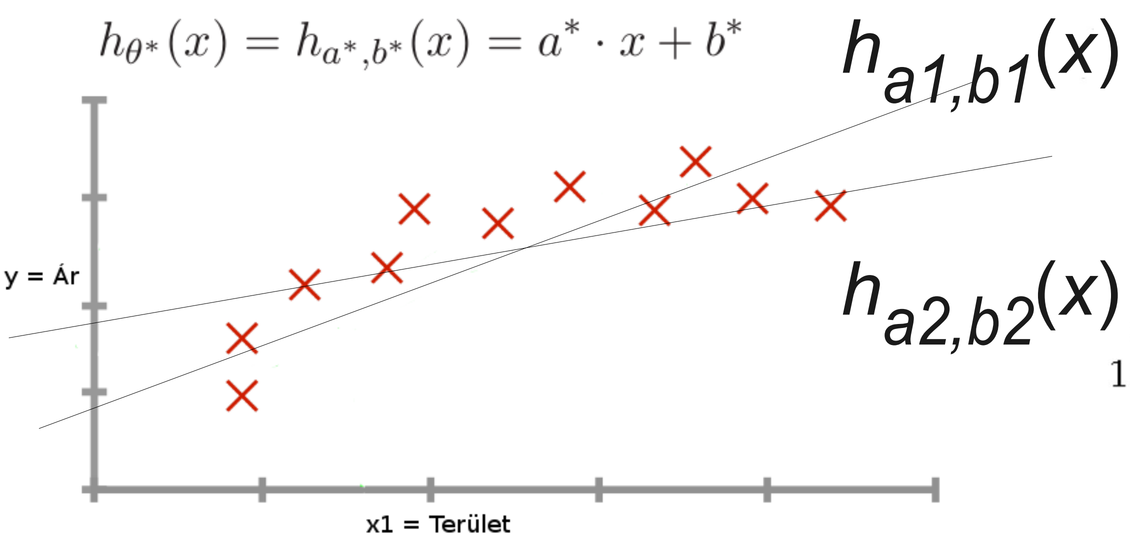


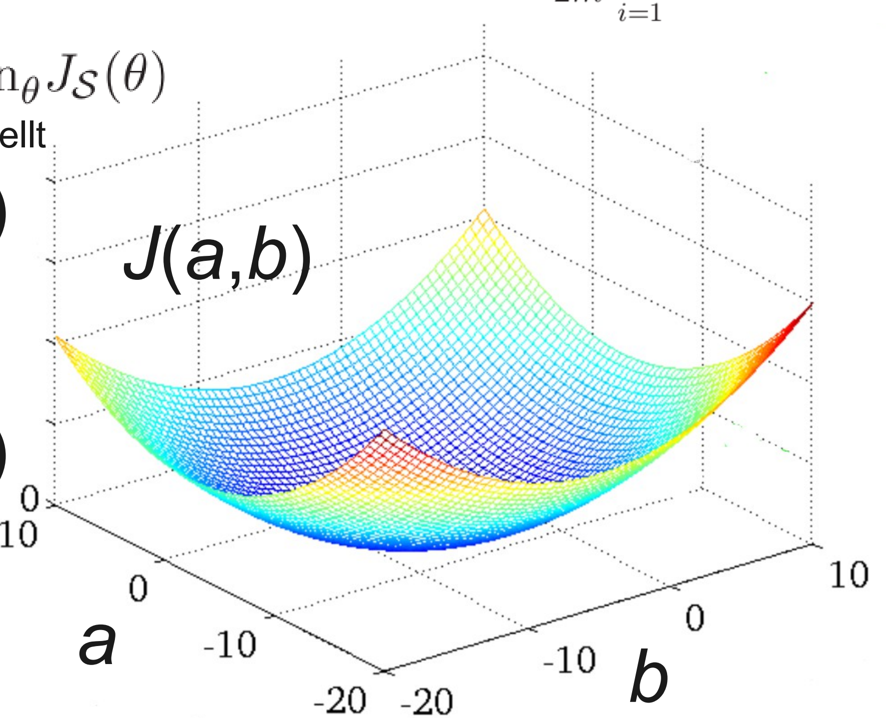


Lineáris regresszió

- A költség függvény:
  - minden egyes alappontja meghatároz egy modellt
  - a hozzá tartozó függvény érték meghatározza, hogy az adott modell mennyire „jó” a tanuló halmazon
- <u>Tanulás-Modell felhasználás:</u>
  - Adott egy tanuló adatbázis: $\mathcal{S} = \{(x_1, y_1), \ldots, (x_m, y_m)\} \subseteq \mathbb{R} \times \mathbb{R}$
  - és néhány paraméter: $\theta = (a, b)$
  - a modell a példák függvénye: $h_\theta(x) = h_{a,b}(x) = a \cdot x + b$
  - definiáljunk egy költség függvényt a paraméterek függvényében: $J_\mathcal{S}(\theta) = J_\mathcal{S}(a,b) = J(a,b) = \frac{1}{2m}\sum_{i=1}^{m}(h_{a,b}(x^{(i)}) - y^{(i)})^2$
  - <u>Tanulás:</u> optimalizáljuk a költség függvényt: $\theta^* = \mathrm{argmin}_\theta J_\mathcal{S}(\theta)$
  - <u>Felhasználás:</u> használjuk az optimális paraméterekkel a modellt

$$h_{\theta^*}(x) = h_{a^*,b^*}(x) = a^* \cdot x + b^*$$





Lineáris regresszió

1. Kvíz: Lehet-e negatív a lineáris regresszió költségfüggvényének az értéke?
    1. Igen, ha a tanító halmaz minden (*x*,*y*) elemére $h(x)<y$
    2. Igen, ha feltesszük, hogy felügyelt tanulásról van szó
    3. Nem
2. Kvíz: Lehet-e olyan paraméter párt mondani, amely mellett az egy dimenziós lineáris regresszió költségfüggvény értéke 0 (adott tanító halmazon)?
    1. Igen, (a=1,b=1)
    2. Nem
    3. Igen, (a=1, b=0)
    4. Igen, (a=0, b=0)

Lineáris regresszió

- <u>Tanulás:</u> opt. a költség függvényt: $\theta^* = \mathrm{argmin}_\theta J_{\mathcal{S}}(\theta)$
- Hogyan határozzuk meg a költség függvény minimumát?
  - <u>Kimerítő keresés</u> → nagy műveletigény főleg nagyobb dimenzióban
  - <u>Hegymászó algoritmus</u> → a lineáris regresszió esetén túl általános
    - Algoritmus:
      - Theta_0 = véletlen, i=0
      - Amíg nem konvergált
        - Theta_i+1 := Theta_i környezetéből egy olyan véletlen pont
        - Ha J(Theta_i+1) >= J(Theta_i) akkor Theta_i+1=Theta_i
  - <u>Gradiens módszer (Gradient descent)</u> → jó választás, hiszen a költség függvény parabolikus → Nézzük meg részletesebben!

Gradiens módszer

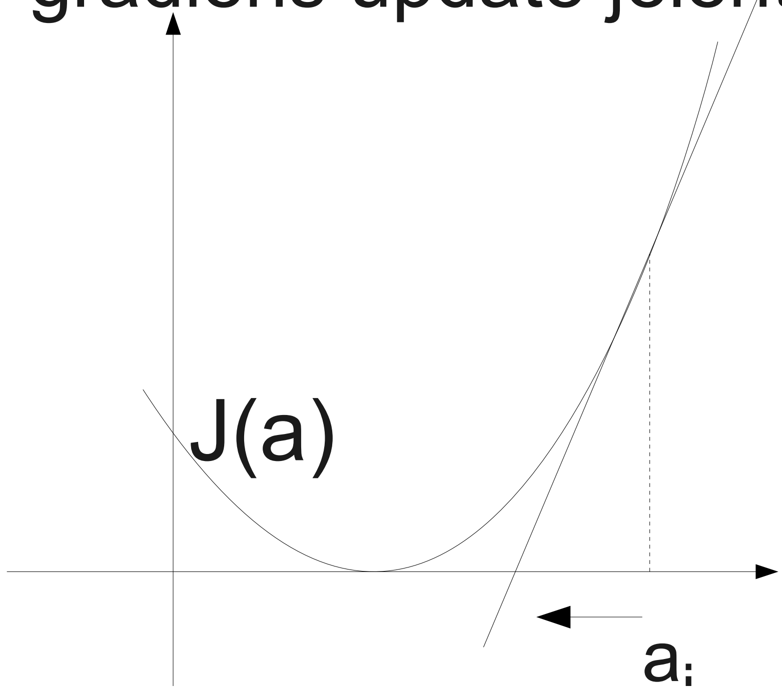
- Hasonlít a Hegymászó algoritmushoz, csak a szomszéd választás irányított.
- <u>Algoritmus:</u>

$$
\begin{array}{l}
\theta_0 := \text{véletlen},\ i = 0 \\
\textbf{repeat} \text{ until convergence } \{ \\
\quad \theta_{i+1} = \theta_i - \alpha \frac{\partial}{\partial \theta_i} J(\theta_i) \\
\}
\end{array}
$$

- Azaz a gradiens irányával ellentétes irányba (kivonás) lépünk el
- A lépés hosszát Alpha határozza meg

Gradiens módszer

- <u>Interpretáció:</u> Tegyük fel ismét, hogy *h*(*x*)=*a***x*, és így $J(a) = \frac{1}{2m}\sum_{i=1}^{m}(a \cdot x^{(i)} - y^{(i)})^2$ egy dimenziós parabola
- Tegyük fel továbbá, hogy eljutottunk a_i pontba, ekkor a gradiens update jelentése a következő

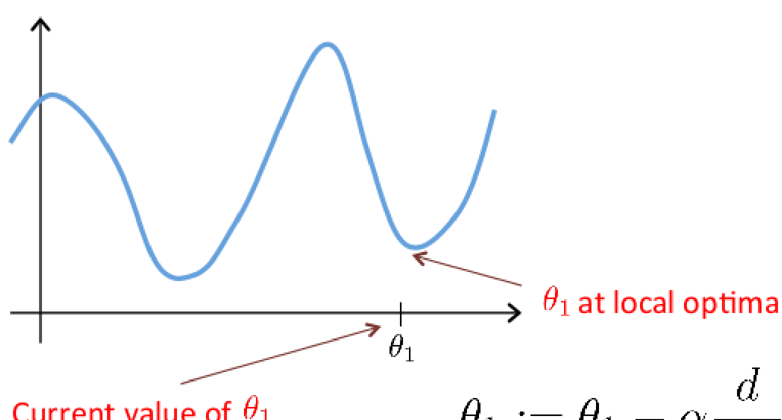


$$a_{i+1} = a_i - \alpha \frac{\partial}{\partial \theta_i} J(a_i)$$

- A derivált függvény értéke (gradiens) az a_i helyen megadja az a_i pontba húzott érintő meredekségét
- A gradiens update ebben az esetben a minimum felé tolja a_{i+1}-et (Alpha > 0)

Gradiens módszer

- Gradiens módszer tulajdonságai:
  - Általánosságban – megfelelő Alpha választásával – lokális optimum megtalálását biztosítja
  - Parabolikus (több dimenzióban kvadratikus függvények esetén) egy optimum van → lokális = globális
- Tanuló konstans (learning rate, Alpha szerepe)
  - Lépés méretét határozza meg
    - Túl kicsi → lassú konvergencia
    - Túl nagy → túl léphet a az optimumon (akár divergálhat is)



$$\theta_1 := \theta_1 - \alpha \frac{d}{d\theta_1} J(\theta_1)$$

Gradiens módszer

3. <u>Kvíz:</u> Parabolikus (kvadratikus) költségfüggvény esetén az gradiens módszer megtalálja a globális optimumot

1. Igen, feltéve, hogy a lépésköz (Alpha) paraméter megfelelően van beállítva
2. Nem, divergál
3. Igen, mindig
4. Igen, feltéve, hogy a költségfüggvény folytonosan differenciálható

Gradiens módszer – Lineáris regresszió esetén

- Tegyük fel ismét, hogy a a két változós optimalizálandó költségfüggvényünk

$$J(a,b) = \frac{1}{2m}\sum_{i=1}^{m}(a \cdot x^{(i)} + b - y^{(i)})^2$$

- Ahhoz, hogy a Gradiens módszer alkalmazható legyen le kell azt deriválnunk minden változója szerint (parciális deriválás)

$$\begin{aligned}
&\theta_0 := \text{véletlen},\ i = 0\\
&\textbf{repeat}\ \text{until convergence}\ \{\\
&\quad \theta_{i+1} = \theta_i - \alpha \tfrac{\partial}{\partial \theta_i} J(\theta_i)\\
&\}
\end{aligned}$$

$$\begin{aligned}
&(a_0, b_0) := \text{véletlen},\ i = 0\\
&\textbf{repeat}\ \text{until convergence}\ \{\\
&\quad a_{i+1} = a_i - \alpha \tfrac{1}{m} \textstyle\sum_{i=1}^{m}(a \cdot x^{(i)} + b - y^{(i)}) \cdot x^{(i)}\\
&\quad b_{i+1} = b_i - \alpha \tfrac{1}{m} \textstyle\sum_{i=1}^{m}(a \cdot x^{(i)} + b - y^{(i)})\\
&\}
\end{aligned}$$

Gradiens módszer

- Annak függvényében, hogy milyen költség függvényt alkalmazunk eltérő módszerekről beszélhetünk:
  - Eddig a teljes tanító halmazt használtuk hiba (költség mérésére) → batch mód $J_{\mathcal{S}}(a,b) = \frac{1}{2m}\sum_{i=1}^{m}(h_{a,b}(x^{(i)}) - y^{(i)})^2$
  - Tanító halmaz egy véletlen részhalmazán történő mérés → mini batch mód $\mathcal{Q} \subseteq \mathcal{S} \quad J_{\mathcal{Q}}(a,b) = \frac{1}{2|\mathcal{Q}|}\sum_{(x^{(i)},y^{(i)})\in\mathcal{Q}}(h_{a,b}(x^{(i)}) - y^{(i)})^2$
  - Egy példán történő mérés → stochastic mód

$$(x^{(i)}, y^{(i)} \in \mathcal{S} \quad J_{(x^{(i)},y^{(i)})}(a,b) = \frac{1}{2}(h_{a,b}(x^{(i)}) - y^{(i)})^2$$