Szoftver mély neuronhálók alkalmazásához

2. előadás

Kovács Bálint, Varga Viktor
ELTE IK Mesterséges Intelligencia Tanszék

Előző órán

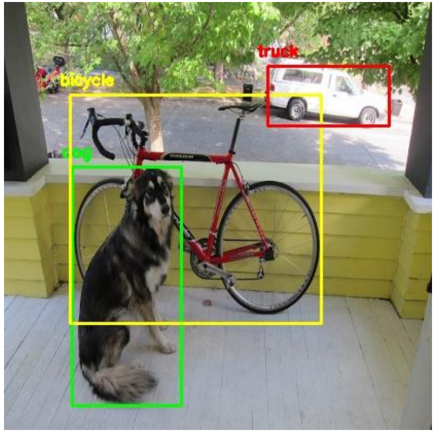
A gépi tanulás lényege

- Egy feladatot nem a megoldó algoritmus leprogramozása segítségével oldunk meg.
- Helyette, egy **paraméteres modellhez keresünk** olyan **paramétereket**, melyekkel a modell megoldja a feladatot.
- Az ideális paramétereket a feladathoz tartozó mintapéldák segítségével találjuk meg.

Előző órán - Gépi tanuláshoz ideális feladatok

- Ha közelítően optimális megoldás is megfelelő.
- Ha nem ismert konkrét algoritmus, ami megoldja a feladatot.
- Ha nem gazdaságos egyedi algoritmust fejleszteni a problémára.
- Ha a feladatot csak mintapéldák segítségével tudjuk formálisan definiálni.

Hogyan definiáljuk azt, hogy “kutya”, vagy “kerékpár” mintapéldák nélkül?



Előző órán - Felügyelt tanulás

Adott: A tanítóminta (training set), input-címke párok halmaza

$$\{(x^{(1)}, y^{(1)}), \ldots, (x^{(m)}, y^{(m)})\}$$

$$x \in \boldsymbol{X} \subset \mathbb{R}^n,\ y \in \boldsymbol{Y} \subset \mathbb{R}^k$$

Feladat: A címke (az elvárt output) minél jobb becslése az inputból.

Azaz, keresünk olyan h_θ függvényt (hipotézisfüggvényt), melyre:

$$h_\theta(x) = \hat{y} \approx y$$

Előző órán - Felügyelt tanulás

a zárójeles felső index nem hatvány, hanem a minta elemeinek indexe

Adott: A tanítóminta (training set), input-címke párok halmaza

$$\{(x^{(1)}, y^{(1)}), \ldots, (x^{(m)}, y^{(m)})\}$$

$$x \in X \subset \mathbb{R}^n, \; y \in Y \subset \mathbb{R}^k$$

Feladat: A címke (az elvárt output) minél jobb becslése az inputból.

Azaz, keresünk olyan h_θ függvényt (hipotézisfüggvényt), melyre:

$$h_\theta(x) = \hat{y} \approx y$$

a hipotézisfüggvény x inputból becsli az y címkét. A becslését “y kalap”-pal jelöljük.

Előző órán - A felügyelt tanulás két fő feladata

Regresszió: A címkehalmaz összefüggő és végtelen

$$|Y| = \infty$$

Példa: Autók számának, vagy életkor becslése

Klasszifikáció: A címkehalmaz diszkrét és számossága véges

$$|Y| < \infty$$

Példa: Mintaelemek kategorizálása

- A lakosság számából eldönteni, hogy város-e, vagy falu egy adott település
- Mi a foglalkozása a képeken szereplő személyeknek?

Ezen az órán - Regresszió

Regresszió: A címkehalmaz összefüggő és végtelen

$$|Y| = \infty$$

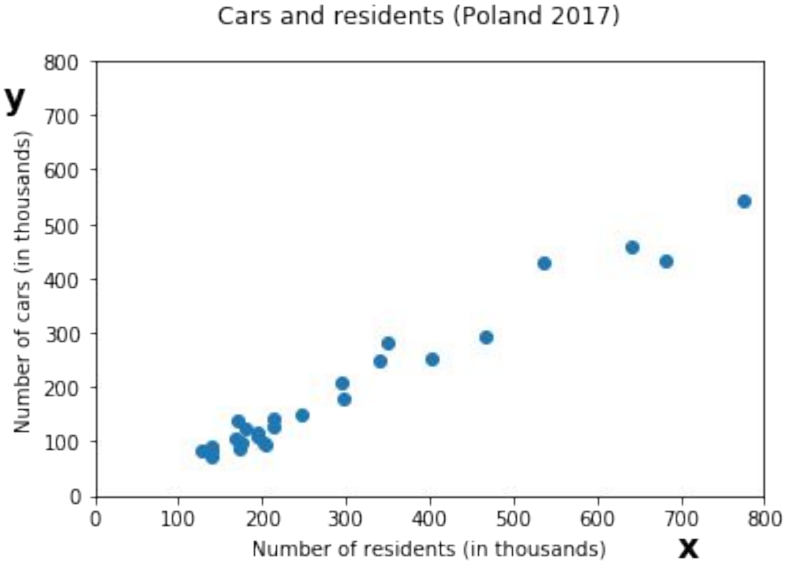
Példa: Autók számának, vagy életkor becslése

Regresszió

Példa: Becsüljük meg az autók számát egy adott városban, ha ismerjük a város lakosságának számát.

x: Egy adott város lakosságának száma

y: Egy adott városban megtalálható autók száma



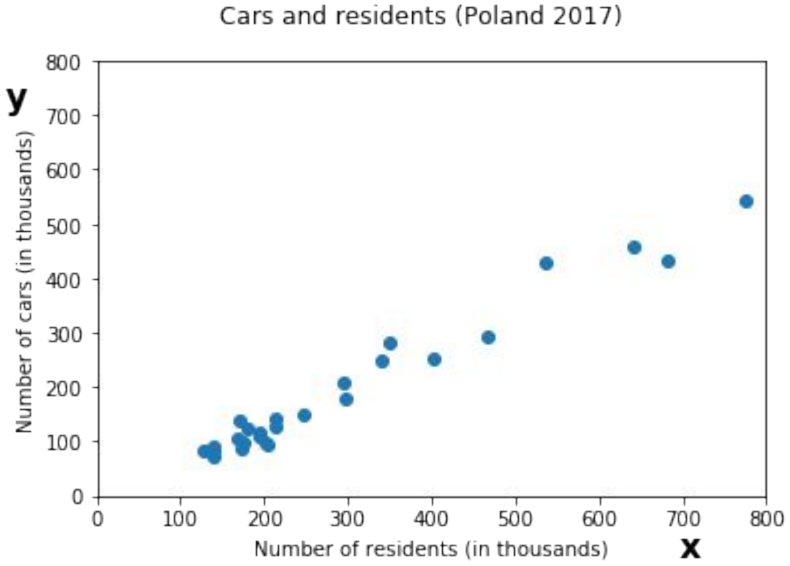
$$x^{(j)}, y^{(j)} \in \mathbb{R}$$

Regresszió

Példa: Becsüljük meg az autók számát egy adott városban, ha ismerjük a város lakosságának számát.

x: Egy adott város lakosságának száma

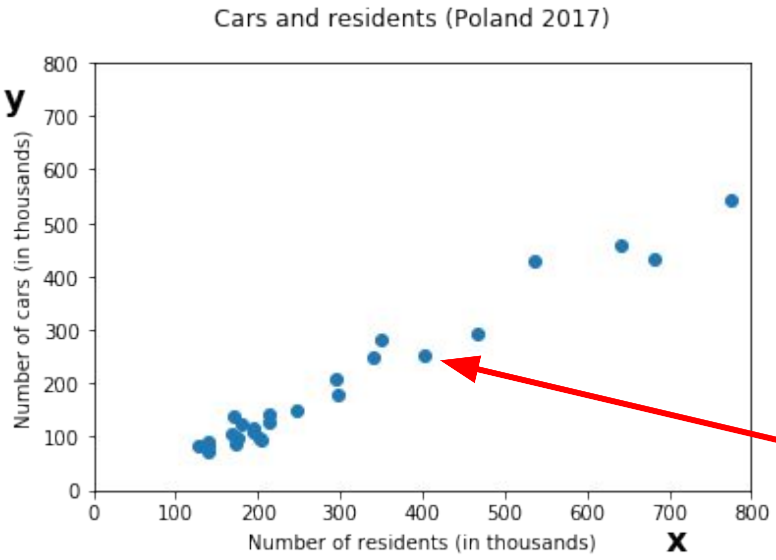
y: Egy adott városban megtalálható autók száma



Egy input változó és **egy címke**: a mintaelemek a címkékkel, kétdimenziós vektortérben (síkon) elhelyezkedő pontok

$$x^{(j)}, y^{(j)} \in \mathbb{R}$$

A minta ábrázolása



$$x^{(j)}, y^{(j)} \in \mathbb{R}$$

	input	címke
	x (egy adott város lakossága, ×1000)	y (autók száma a városban, ×1000)
Varsó (j = 1)	1760	910
Krakkó (j = 2)	770	535
Szczecin (j = 3)	407	263
...	...	...

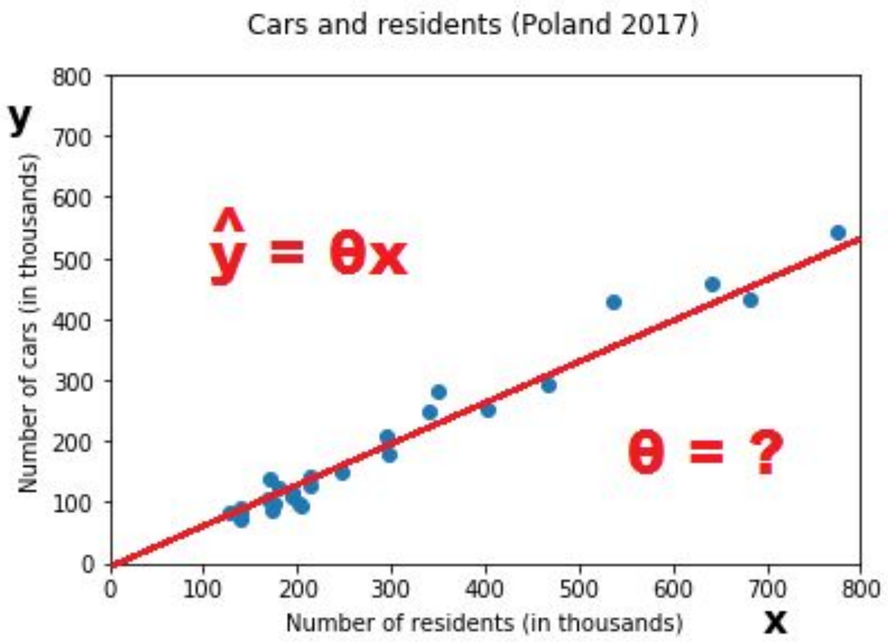
Lineáris regresszió

Hipotézisfüggvény:

Nagyon egyszerű (lineáris) hipotézisfüggvény:

$$y \approx \hat{y} = h_\theta(x) = \theta x$$

$\theta \in \mathbb{R}$ **a hipotézisfüggvény paramétere**, ebben az esetben az egyenes meredeksége lesz...



Lineáris regresszió

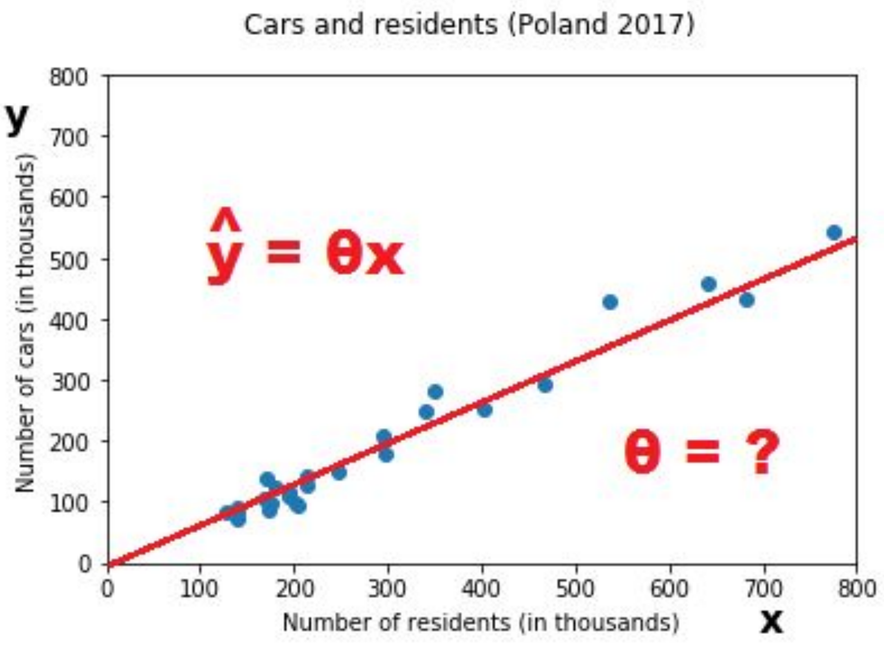
Keresünk egy olyan θ paramétert, mellyel h(x) jól közelíti az igazi y címkéket!
Például a h(x) = 0.65*x hipotézis-függvény erre a konkrét mintára jól illeszkedik, így tehát egy jó paraméter a θ = 0.65.

Hipotézisfüggvény:

Nagyon egyszerű (lineáris) hipotézisfüggvény:

$$y \approx \hat{y} = h_\theta(x) = \theta x$$

$\theta \in \mathbb{R}$ **a hipotézisfüggvény paramétere**, ebben az esetben az egyenes meredeksége lesz...



Lineáris regresszió

Hogyan állapítjuk meg mennyire jó a becslés?

$$h_\theta(x) = \hat{y} \overset{?}{\approx} y$$

Lineáris regresszió

Hogyan állapítjuk meg mennyire jó a becslés?

$$h_\theta(x) = \hat{y} \overset{?}{\approx} y$$

J költségfüggvény segítségével: $J : \theta \to \mathbb{R}_{\geq 0}$

A költségfüggvény megadja, hogy mennyire tér el a valódi címke és a becslésünk adott paraméter értékek esetén.

Lineáris regresszió

Hogyan állapítjuk meg mennyire jó a becslés?

$$h_\theta(x) = \hat{y} \overset{?}{\approx} y$$

J költségfüggvény segítségével: $J : \theta \rightarrow \mathbb{R}_{\geq 0}$

A költségfüggvény megadja, hogy mennyire tér el a valódi címke és a becslésünk adott paraméter értékek esetén.

Akárhány dimenziós a címkénk és akárhány mintaelem van a tanítóhalmazban, **a költség mindenképpen egyetlen szám kell, hogy legyen**. Így egyértelműen össze tudunk hasonlítani két hipotézisfüggvényt / paraméterkombinációt.

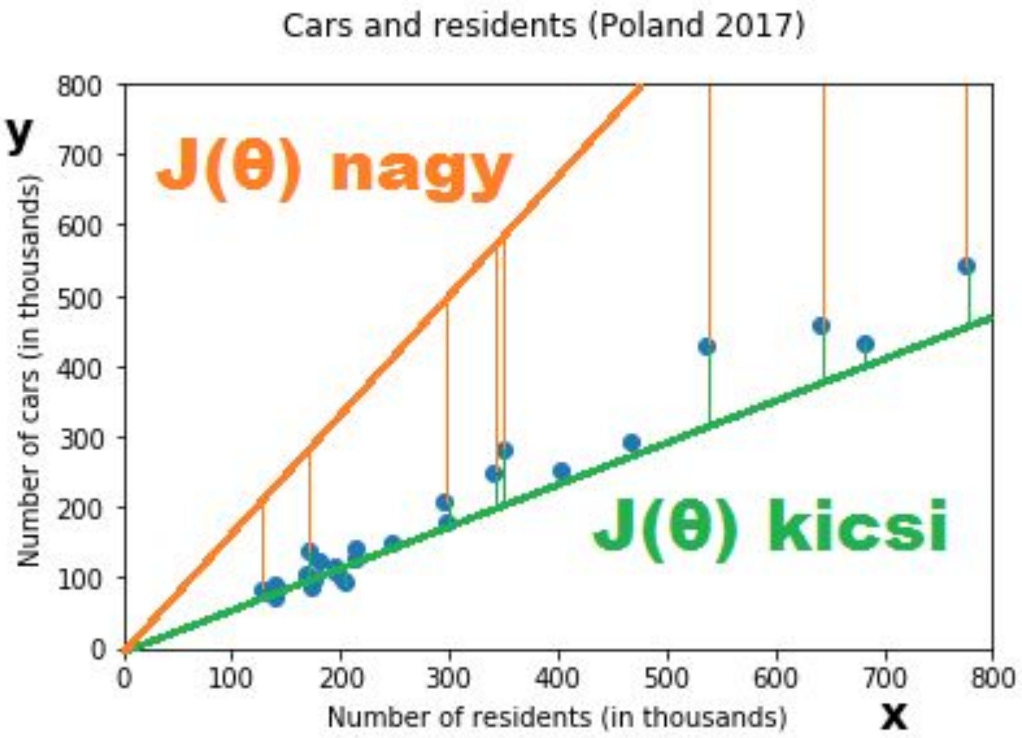
Lineáris regresszió

Lineáris regresszió esetén a költségfüggvény:

$$J(\theta) = \frac{1}{2m} \sum_{j=1}^{m} (\overbrace{h_\theta(x^{(j)})}^{\hat{y}^{(j)}} - y^{(j)})^2$$

… azaz, a hipotézisfüggvény által becsült címkék és az igazi címkék különbségének a négyzete, átlagolva a mintaelemek felett: **MSE (Mean Squared Error**, átlagos négyzetes eltérés)

Lineáris regresszió



Lineáris regresszió

A feladat tehát: $$\theta^* = \underset{\theta}{\operatorname{argmin}}\, J(\theta)$$

→ keressük azt az **optimális paramétert** (**θ***), mellyel minimális a költség, azaz a címke predikciónk átlagos négyzetes eltérése az igazi címkétől.

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \frac{1}{2m} \sum_{j=1}^{m} (h_\theta(x^{(j)}) - y^{(j)})^2 =$$

$$= \underset{\theta}{\operatorname{argmin}} \frac{1}{2m} \sum_{j=1}^{m} (\theta x^{(j)} - y^{(j)})^2$$

Lineáris regresszió - megoldás

A feladat: $$\theta^* = \underset{\theta}{\operatorname{argmin}}\, J(\theta)$$

$$J(\theta) = \frac{1}{2m}\sum_{j=1}^{m}(h_\theta(x^{(j)}) - y^{(j)})^2 = \frac{1}{2m}\sum_{j=1}^{m}(\theta x^{(j)} - y^{(j)})^2$$

Hogy néz ki a költségfüggvény kifejtve?

Lineáris regresszió - megoldás

A feladat: $\theta^* = \underset{\theta}{\operatorname{argmin}} J(\theta)$

$$J(\theta) = \frac{1}{2m}\sum_{j=1}^{m}(h_\theta(x^{(j)}) - y^{(j)})^2 = \frac{1}{2m}\sum_{j=1}^{m}(\theta x^{(j)} - y^{(j)})^2$$

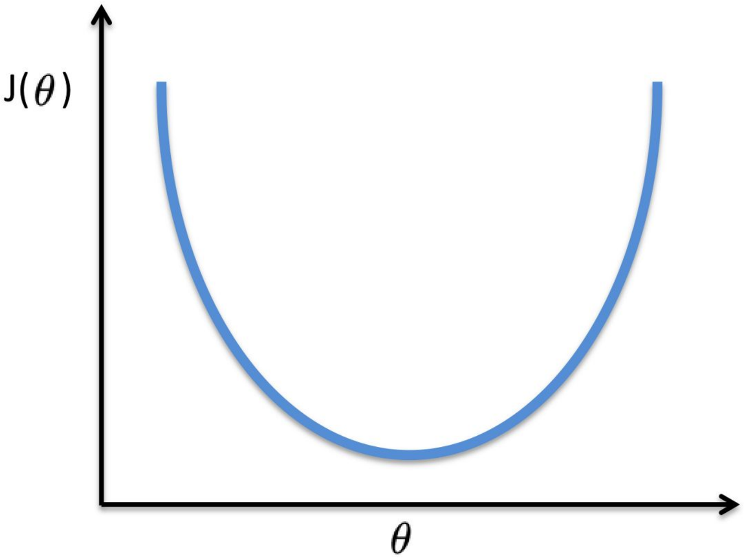
Hogy néz ki a költségfüggvény kifejtve?

x, y ismert konstansok, hiszen a tanítóhalmazban vannak.
θ ismeretlen.
→ **J kvadratikus θ szerint.**

Lineáris regresszió - megoldás

J költségfüggvény kvadratikus.

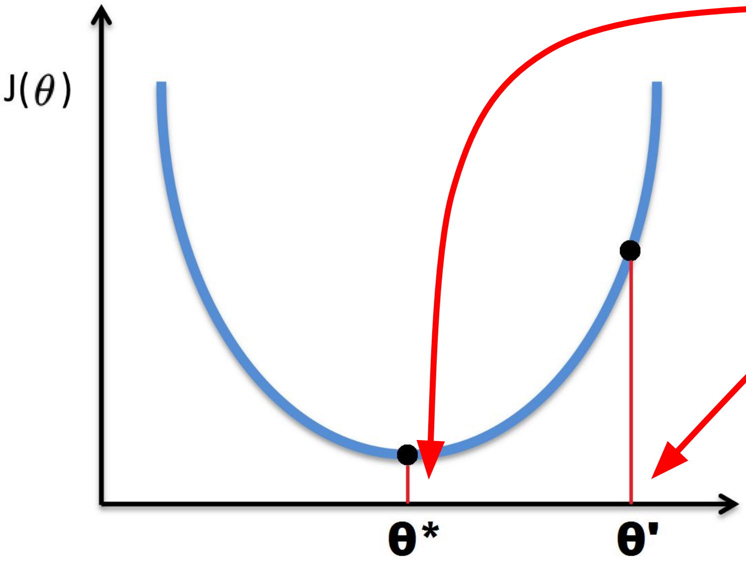
Mivel a θ paraméter most csak egyetlen szám, így J parabola:



$$J(\theta) = \frac{1}{2m} \sum_{j=1}^{m} (\theta x^{(j)} - y^{(j)})^2$$

Lineáris regresszió - megoldás

A feladat: $$\theta^* = \underset{\theta}{\operatorname{argmin}}\, J(\theta)$$



Optimális paraméter: θ*

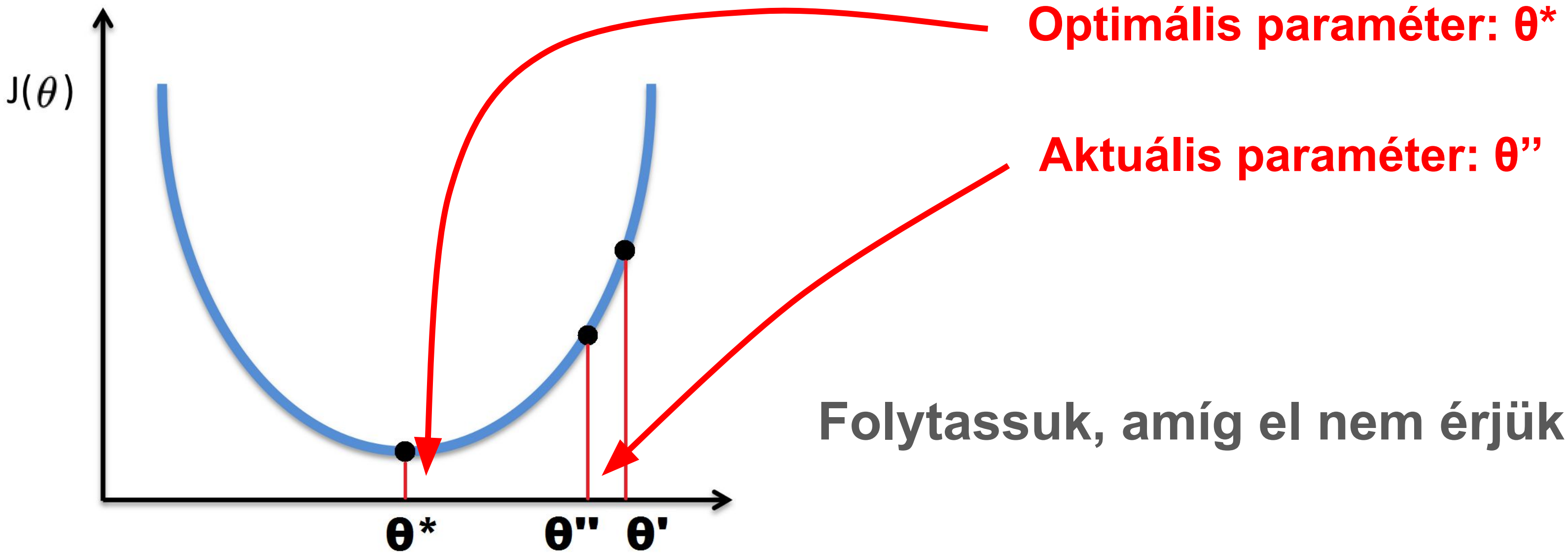
Aktuális paraméter: θ’

Hogyan jutunk el θ’ -ból θ* -ba?

Lineáris regresszió - megoldás

Hogyan jutunk el θ’ -ból θ* -ba?

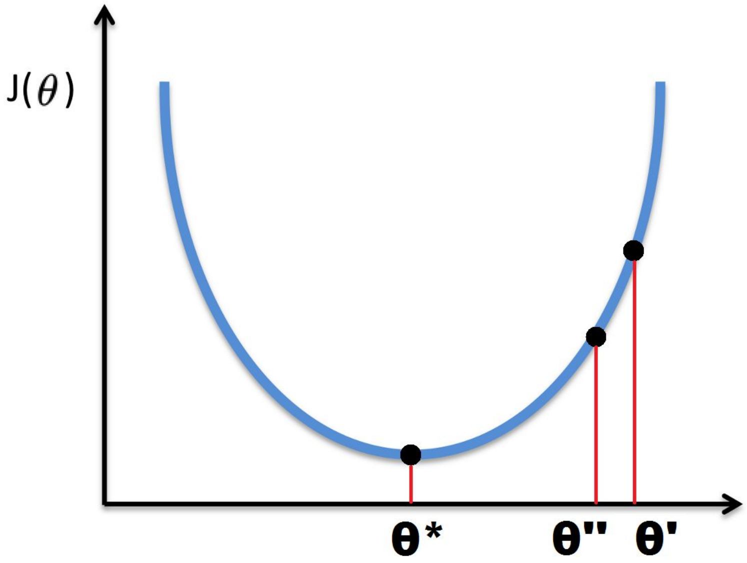
Változtassunk θ’ -n, úgy, hogy J(θ’’) < J(θ’) !



Folytassuk, amíg el nem érjük θ* -t!

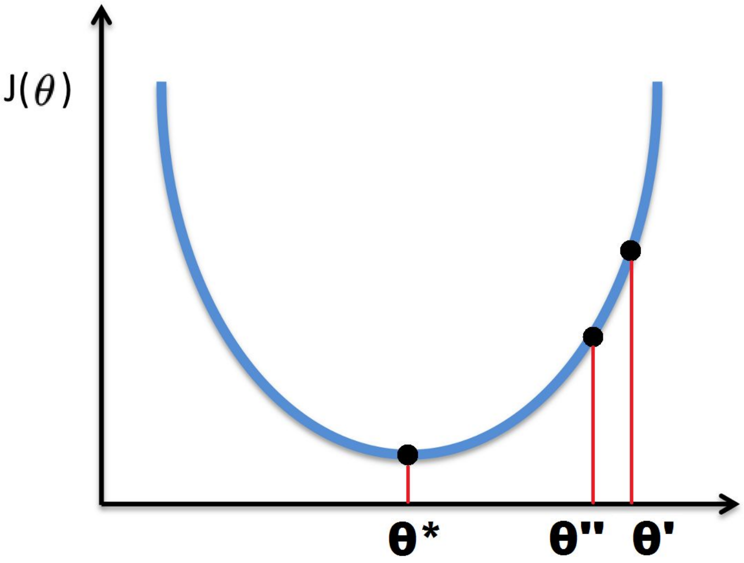
Lineáris regresszió - megoldás

Honnan tudjuk, hogy merre kell lépnünk, hogy a költség csökkenjen?



Lineáris regresszió - megoldás

Honnan tudjuk, hogy merre kell lépnünk, hogy a költség csökkenjen?



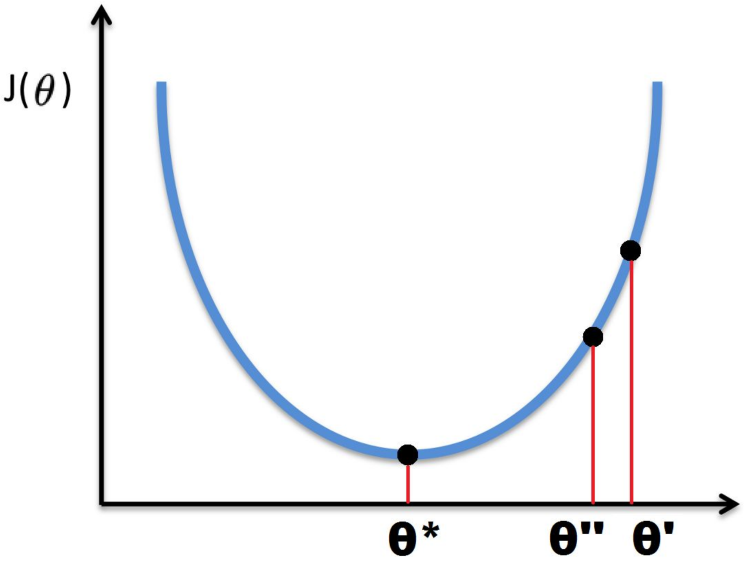
Egyetlen paraméter esetén nem nehéz: pl. megnézzük, hogyan változik a költség, ha θ'-t csökkentjük, vagy növeljük egy kicsivel. Lépjünk arra, amerre csökken a költség!

Azonban, **ha majd több paraméterünk lesz**, a kipróbálandó lépésirányok száma exponenciálisan nő a paraméterek számával…

Használjunk kifinomultabb módszert!

Lineáris regresszió - megoldás

Honnan tudjuk, hogy merre kell lépnünk, hogy a költség csökkenjen?

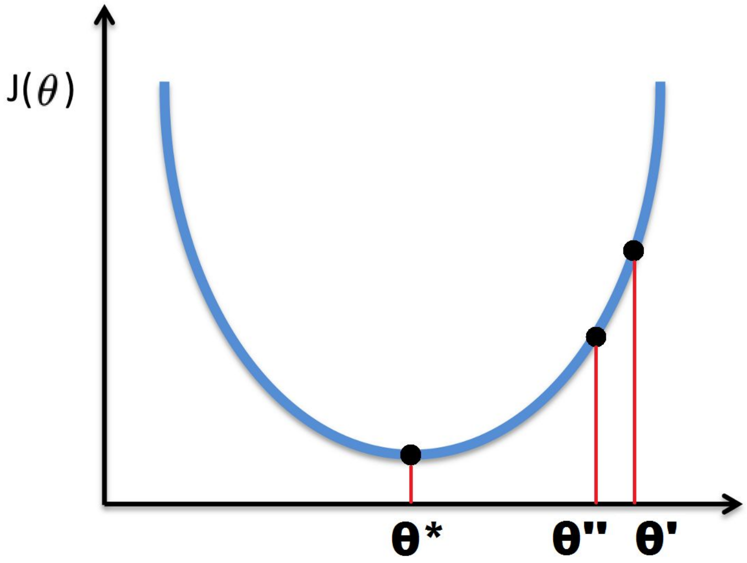


Nézzük meg merre lejt leginkább a költségfüggvény az aktuális paraméterértékben állva!

Mi adja meg J meredekségét egy adott θ pontban?

Lineáris regresszió - megoldás

Honnan tudjuk, hogy merre kell lépnünk, hogy a költség csökkenjen?



Nézzük meg merre lejt leginkább a költségfüggvény az aktuális paraméterértékben állva!

Mi adja meg J meredekségét egy adott θ pontban?

J deriváltja a θ pontban.

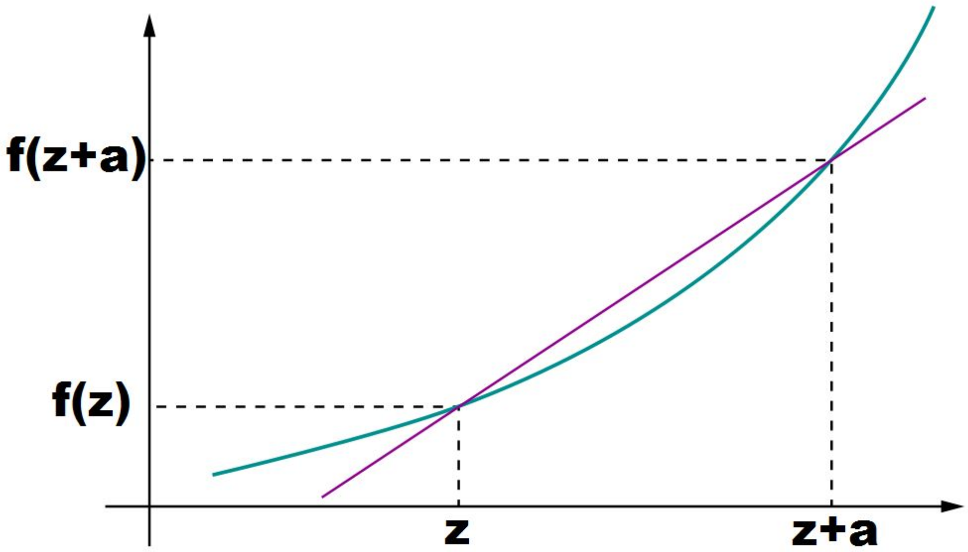
Lineáris regresszió - megoldás

Mi egy függvény deriváltja?

Lineáris regresszió - megoldás

Mi egy függvény deriváltja?

$$f'(z) = \lim_{a \to 0} \frac{f(z+a) - f(z)}{a}$$



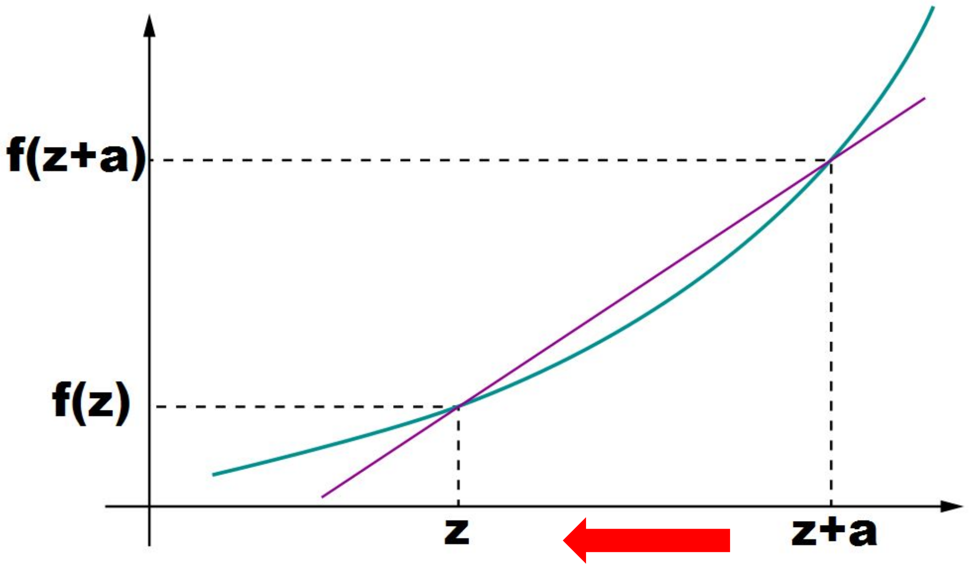
Lineáris regresszió - megoldás

Mi egy függvény deriváltja?

$$f'(z) = \lim_{a \to 0} \frac{f(z+a) - f(z)}{a}$$

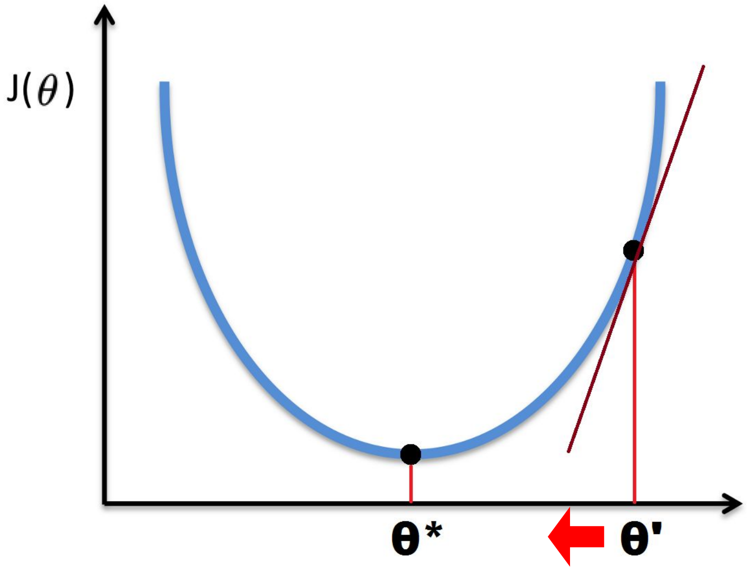
f függvény deriváltja a z pontban
a függvénygörbe z-beli érintőjének meredeksége.

A derivált a differenciahányados határértékével van definiálva, feltéve, ha az létezik és véges.



Lineáris regresszió - megoldás

Honnan tudjuk, hogy merre kell lépnünk, hogy a költség csökkenjen?



Lépjünk θ-val abba az irányba, amerre a legjobban lejt a költségfüggvény felülete.

Ehhez ki kell számolnunk a J költségfüggvény deriváltját.

Lineáris regresszió - megoldás

A költségfüggvény deriváltja:

$$J(\theta) = \frac{1}{2m} \sum_{j=1}^{m} (\theta x^{(j)} - y^{(j)})^2$$

$$\frac{\partial}{\partial \theta} J(\theta) = \ ?$$

Lineáris regresszió - megoldás

A költségfüggvény deriváltja:

$$J(\theta) = \frac{1}{2m} \sum_{j=1}^{m} (\theta x^{(j)} - y^{(j)})^2$$

Ne feledjük: x, y ismert értékek a tanítóhalmazból, egyedül θ az ismeretlen.

$$\frac{\partial}{\partial \theta} J(\theta) = \; ?$$

Lineáris regresszió - megoldás

A költségfüggvény deriváltja:

$$J(\theta) = \frac{1}{2m} \sum_{j=1}^{m} (\theta x^{(j)} - y^{(j)})^2$$

$$\frac{\partial}{\partial \theta} J(\theta) = \frac{1}{2m} \sum_{j=1}^{m} 2 \cdot (\theta x^{(j)} - y^{(j)}) \cdot (\theta x^{(j)} - y^{(j)})' =$$

$$= \frac{1}{m} \sum_{j=1}^{m} (\theta x^{(j)} - y^{(j)}) x^{(j)}$$

Lineáris regresszió - megoldás

A költségfüggvény deriváltja:

Összetett függvények deriválási szabálya (láncszabály):
f(g(z))’ = f’(g(z)) · g’(z)

$$J(\theta) = \frac{1}{2m}\sum_{j=1}^{m}(\theta x^{(j)} - y^{(j)})^2$$

$$\frac{\partial}{\partial\theta}J(\theta) = \frac{1}{2m}\sum_{j=1}^{m}2\cdot(\theta x^{(j)} - y^{(j)})\cdot(\theta x^{(j)} - y^{(j)})' =$$

$$= \frac{1}{m}\sum_{j=1}^{m}(\theta x^{(j)} - y^{(j)})x^{(j)}$$

Leibniz-féle jelölés a parciális deriválásra: J deriváltja θ szerint. Itt még csak egy változónk van (θ), így nem kérdés, hogy mi szerint deriválunk…

Lineáris regresszió - megoldás

Gradiensmódszer (gradient descent)

Iteratívan lépegetünk θ-val, abba az irányba, amerre a legnagyobb a lejtése a költségfüggvénynek az aktuális θ-ban.

$$
\begin{aligned}
&\text{repeat until convergence} \quad \{ \\
&\quad grad := \frac{\partial}{\partial\theta} J(\theta) \\
&\quad \theta := \theta - \alpha \cdot \; grad \\
&\}
\end{aligned}
$$

Lineáris regresszió - megoldás

Gradiensmódszer (gradient descent)

Iteratívan lépegetünk θ-val, abba az irányba, amerre a legnagyobb a lejtése a költségfüggvénynek az aktuális θ-ban.

$$\text{repeat until convergence} \quad \{$$

$$grad := \frac{\partial}{\partial\theta} J(\theta)$$

$$\theta := \theta - \alpha \cdot \; grad$$

$$\}$$

$$\frac{\partial}{\partial\theta} J(\theta) = \frac{1}{m} \sum_{j=1}^{m} (\theta x^{(j)} - y^{(j)}) x^{(j)}$$

Lineáris regresszió - megoldás

Gradiensmódszer (gradient descent)

Iteratívan lépegetünk θ-val, abba az irányba, amerre a legnagyobb a lejtése a költségfüggvénynek az aktuális θ-ban.

$$\begin{aligned}
&\text{repeat until convergence} \quad \{ \\
&\quad grad := \frac{\partial}{\partial\theta} J(\theta) \\
&\quad \theta := \theta - \alpha \cdot \ grad \\
&\}
\end{aligned}$$

J gradiense: J legnagyobb emelkedése irányába mutató vektor; elemei J adott pontbeli parciális deriváltjai (itt még csak egyelemű)

alfa: tanulási ráta (learning rate); a lépések mérete skálázható vele

Lineáris regresszió - megoldás

Gradiensmódszer (gradient descent)

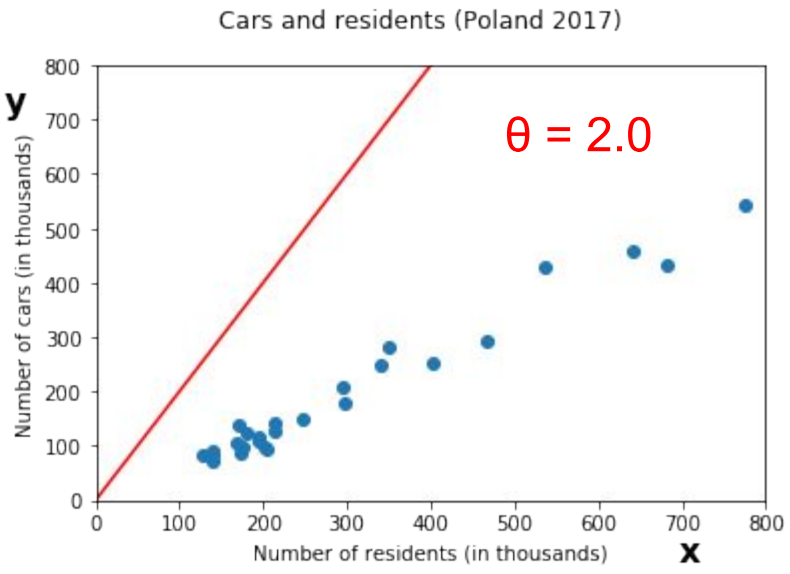
Iteratívan lépegetünk θ-val, abba az irányba, amerre a legnagyobb a lejtése a költségfüggvénynek az aktuális θ-ban.

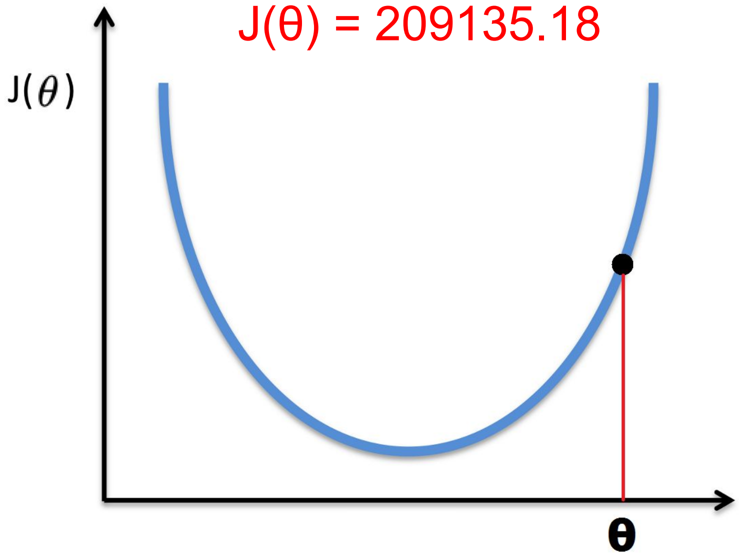
$$\text{repeat until convergence} \quad \{$$
$$\quad grad := \frac{\partial}{\partial \theta} J(\theta)$$
$$\quad \theta := \theta - \alpha \cdot \ grad$$
$$\}$$

A gradienst **kivonjuk** az aktuális paraméterértékből, hiszen a legnagyobb lejtést keressük.

Lineáris regresszió - megoldás

A gradiensmódszer alkalmazása, T = 0 (az első lépés előtt)

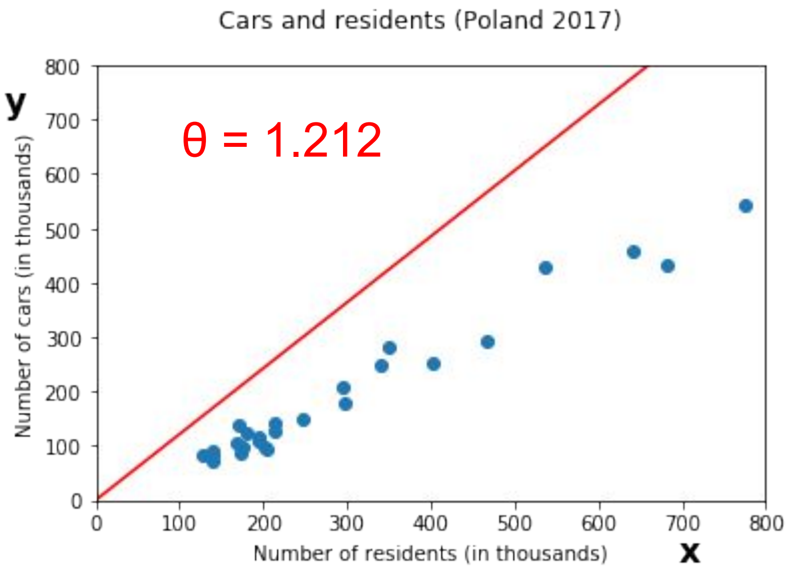


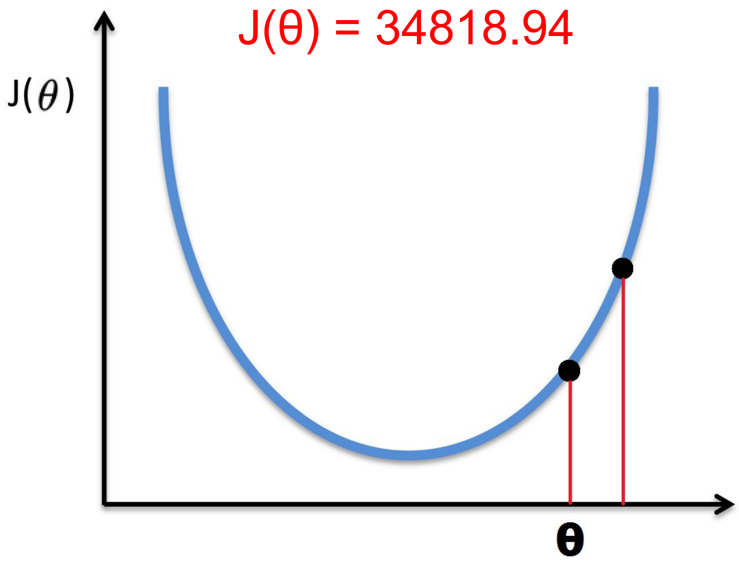


A kezdeti θ -t véletlenszerűen választottuk.

Lineáris regresszió - megoldás

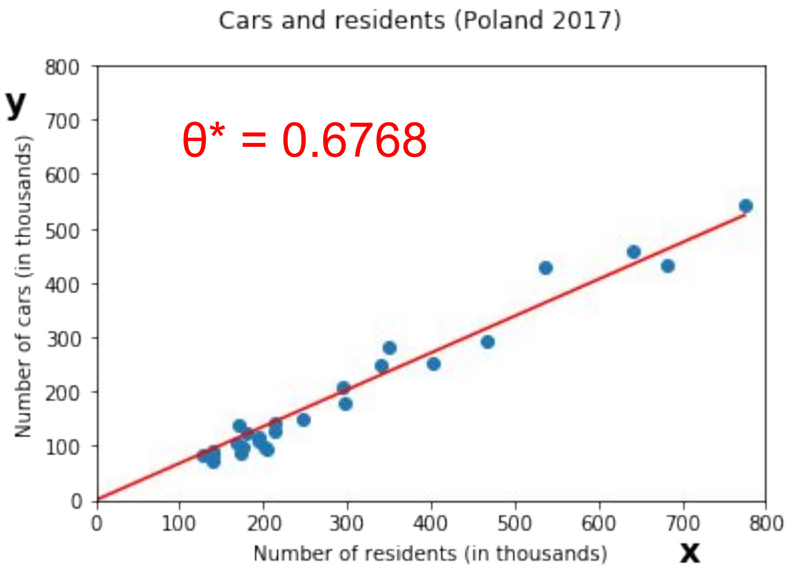
A gradiensmódszer alkalmazása, T = 1

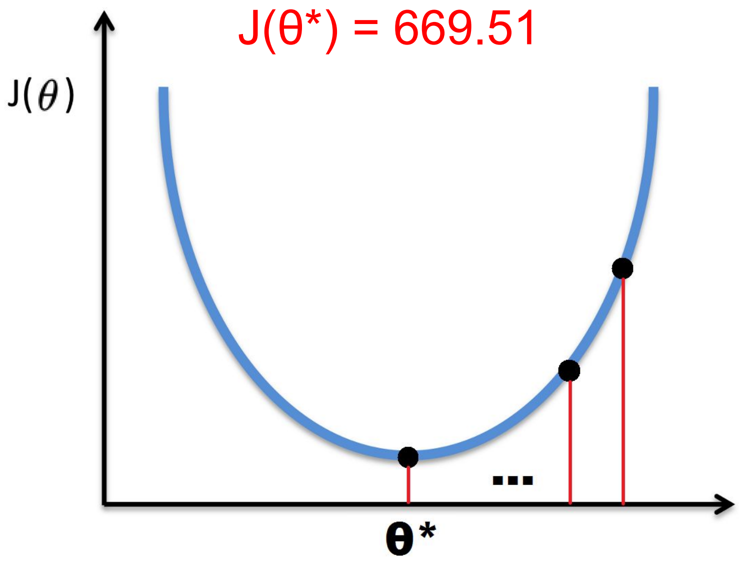




Lineáris regresszió - megoldás

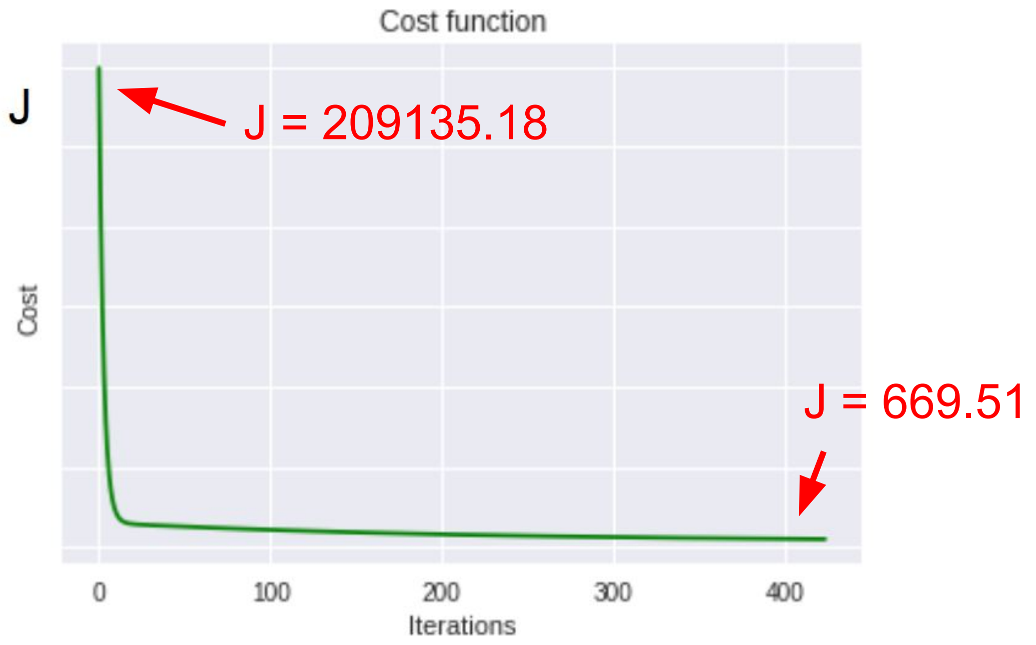
A gradiensmódszer alkalmazása, T = <sok>





Lineáris regresszió - megoldás

A költség alakulása a gradiensmódszer alkalmazása során

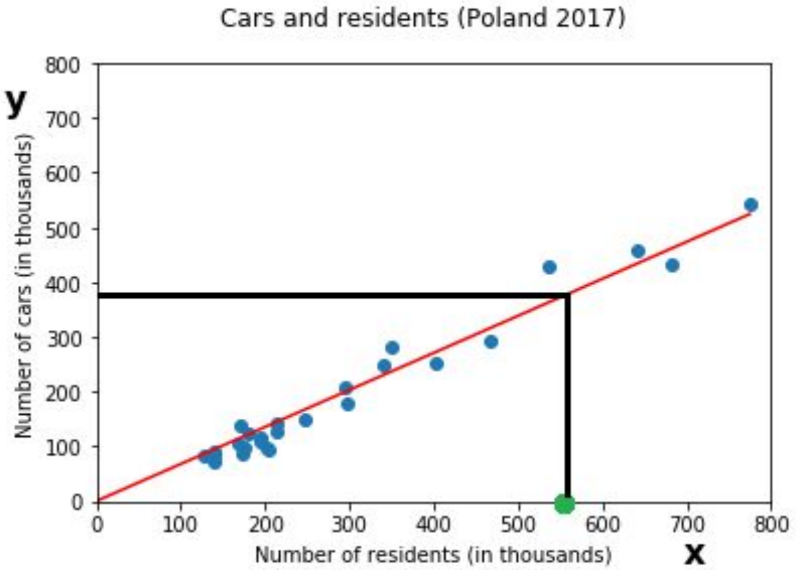


Lineáris regresszió - megoldás

Mit értünk el?

Betanítottunk egy egyszerű (lineáris) regressziós modellt.

Új, címkézetlen mintaelemekre fogunk tudni címkét becsülni.



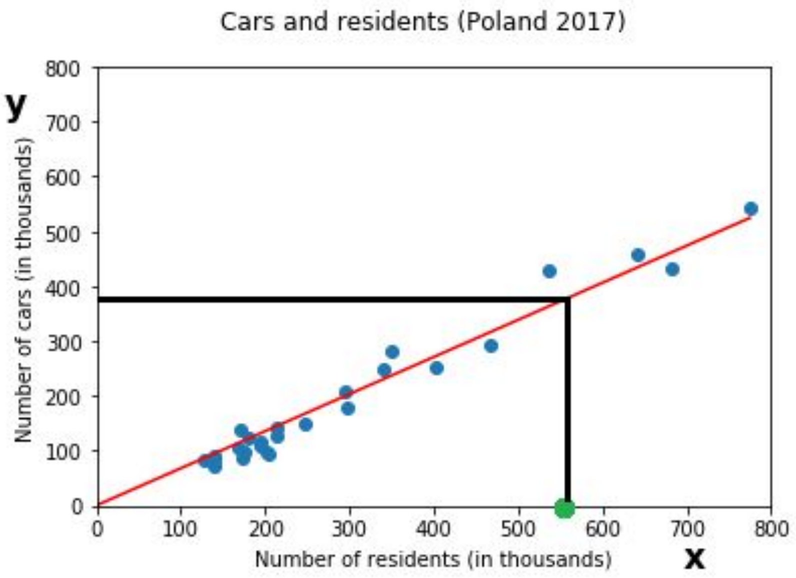
Lineáris regresszió - megoldás

Mit értünk el?

Betanítottunk egy egyszerű (lineáris) regressziós modellt.

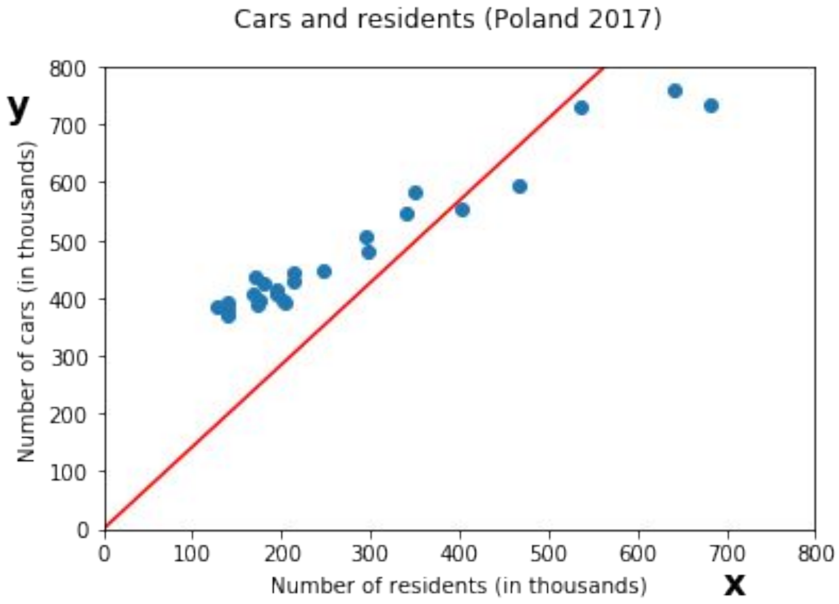
Új, címkézetlen mintaelemekre fogunk tudni címkét becsülni.

Például **egy 550 ezer lakosú városra**
0.6768 * 550 000 = **372 240 autót becslünk.**



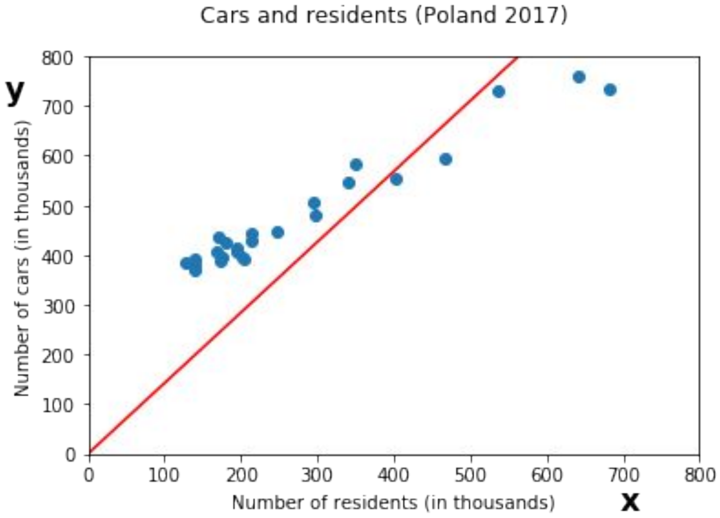
Lineáris regresszió - megoldás

Mi hiányzik?



Lineáris regresszió - megoldás

Mi hiányzik?



Tegyük fel, hogy minden város, függetlenül a lakosságának számától beszerez 300 ezer autót tartalékba…

Erre a mintára nem illeszkednek jól az origón átmenő egyenesek.

Lineáris regresszió - megoldás

Mi hiányzik?

Az eddigi modellünk túlzottan limitált. A hipotézisfüggvény egy olyan egyenes volt, amely az origón kellett, hogy átmenjen…

A meredekség mellé egy konstans (bias / intercept) paramétert is bevezetünk.

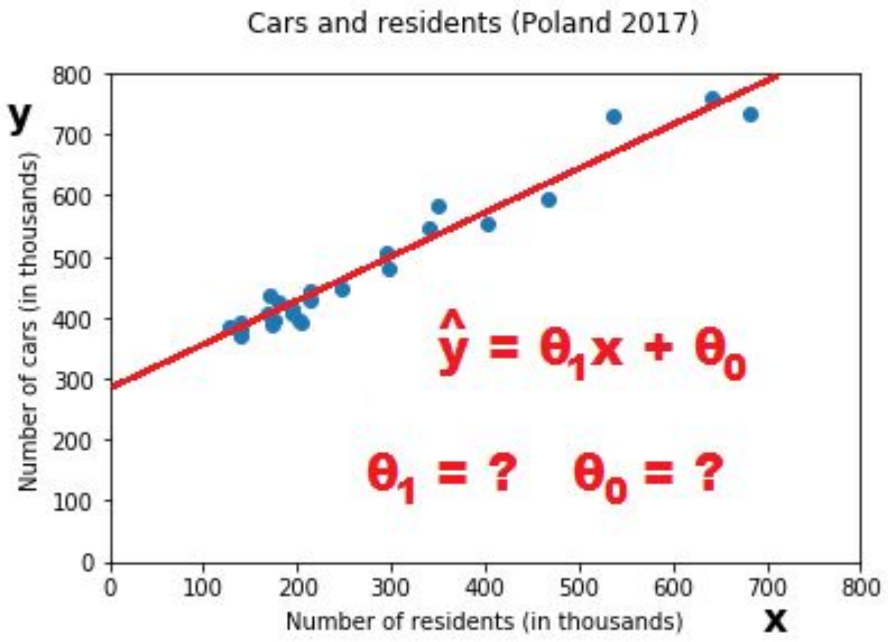
Eddigi hipotézis: $h(x) = \theta x = \hat{y} \approx y$

Új hipotézis: $h(x) = \theta_1 x + \theta_0 = \hat{y} \approx y$

Lineáris regresszió

Az új hipotézisfüggvény: $$y \approx \hat{y} = h(x) = \theta_1 x + \theta_0$$

$$\theta_1, \theta_0 \in \mathbb{R}$$

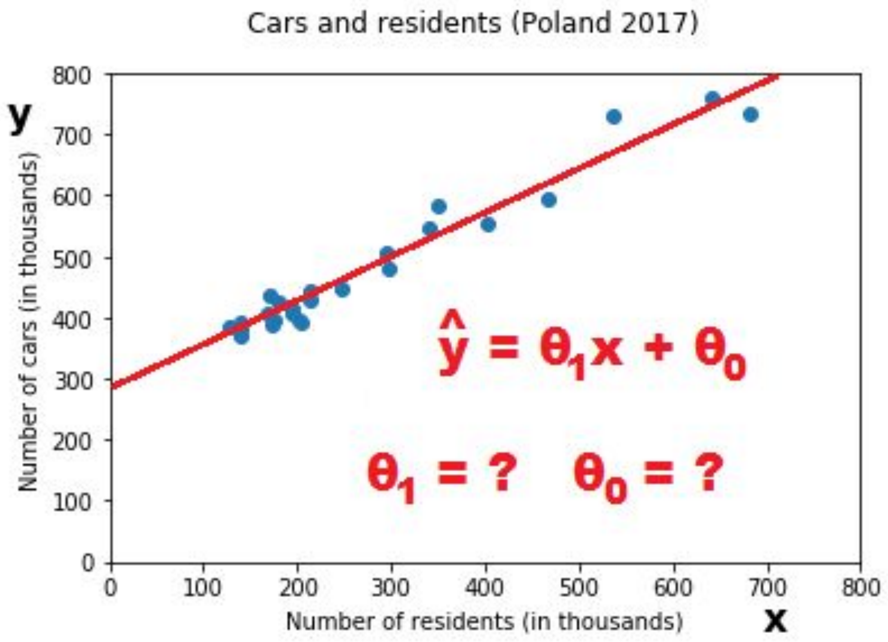


Lineáris regresszió

Az új hipotézisfüggvény: $$y \approx \hat{y} = h(x) = \theta_1 x + \theta_0$$

$$\theta_1, \theta_0 \in \mathbb{R}$$

θ_1 az egyenes meredeksége, **θ_0** pedig az az érték, ahol az egyenes metszi az y tengelyt (bias / intercept).



Lineáris regresszió - megoldás

A költségfüggvény továbbra is **MSE** (Mean Squared Error):

$$J(\theta) = \frac{1}{2m} \sum_{j=1}^{m} (\overbrace{h_\theta(x^{(j)})}^{\hat{y}^{(j)}} - y^{(j)})^2$$

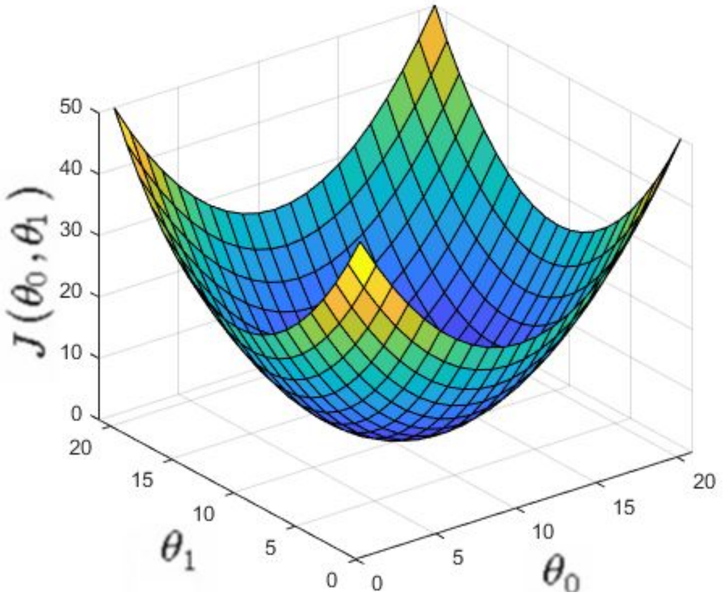
A hipotézisfüggvény viszont megváltozott.

Hogy fog kinézni a költségfüggvény grafikonja?

Lineáris regresszió - megoldás

J költségfüggvény továbbra is kvadratikus.

Mivel már két paraméterünk van, J egy elliptikus paraboloid:



$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{j=1}^{m} (\overbrace{\theta_1 x^{(j)} + \theta_0}^{\hat{y}^{(j)}} - y^{(j)})^2$$

Lineáris regresszió - megoldás

Hogyan változtassuk a paramétereket, hogy a költség csökkenjen?

Használjuk a gradiens módszert: lépegessünk a legnagyobb meredekségű lejtés irányába.

Ezt az irányt egy adott pontban a **gradiens** vektor fogja megadni, melynek elemei a J **költségfüggvény** parciális deriváltjai az egyes paraméterek szerint.

Lineáris regresszió - megoldás

A költségfüggvény parciális deriváltjai:

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{j=1}^{m} (\theta_1 x^{(j)} + \theta_0 - y^{(j)})^2$$

$$\frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) = \ ?$$

$$\frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1) = \ ?$$

Lineáris regresszió - megoldás

A költségfüggvény parciális deriváltjai:

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{j=1}^{m} (\theta_1 x^{(j)} + \theta_0 - y^{(j)})^2$$

$$\frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) = \frac{1}{m} \sum_{j=1}^{m} (\theta_1 x^{(j)} + \theta_0 - y^{(j)})$$

$$\frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1) = \frac{1}{m} \sum_{j=1}^{m} (\theta_1 x^{(j)} + \theta_0 - y^{(j)}) \cdot x^{(j)}$$

Lineáris regresszió - megoldás

A költségfüggvény parciális deriváltjai:

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{j=1}^{m} (\theta_1 x^{(j)} + \theta_0 - y^{(j)})^2$$

A parciális deriváláskor a függvényt az egyik változó szerint deriváljuk. **Ekkor a többi változót konstansként kezeljük.**

$$\frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) = \frac{1}{m} \sum_{j=1}^{m} (\theta_1 x^{(j)} + \theta_0 - y^{(j)})$$

érintő meredeksége **θ_0** irányban

$$\frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1) = \frac{1}{m} \sum_{j=1}^{m} (\theta_1 x^{(j)} + \theta_0 - y^{(j)}) \cdot x^{(j)}$$

érintő meredeksége **θ_1** irányban

Lineáris regresszió - megoldás

Gradiens módszer algoritmusa két paraméter esetén:

$$
\begin{aligned}
&\text{repeat until convergence} \quad \{ \\
&\quad grad_0 := \frac{\partial}{\partial\theta_0} J(\theta_0, \theta_1) \\
&\quad grad_1 := \frac{\partial}{\partial\theta_1} J(\theta_0, \theta_1) \\
&\quad \theta_0 := \theta_0 - \alpha \cdot \ grad_0 \\
&\quad \theta_1 := \theta_1 - \alpha \cdot \ grad_1 \\
&\}
\end{aligned}
$$

Lineáris regresszió - megoldás

Gradiens módszer algoritmusa két paraméter esetén:

$$
\begin{aligned}
&\text{repeat until convergence} \quad \{ \\
&\quad grad_0 := \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) \\
&\quad grad_1 := \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1) \\
&\quad \theta_0 := \theta_0 - \alpha \cdot grad_0 \\
&\quad \theta_1 := \theta_1 - \alpha \cdot grad_1 \\
&\}
\end{aligned}
$$

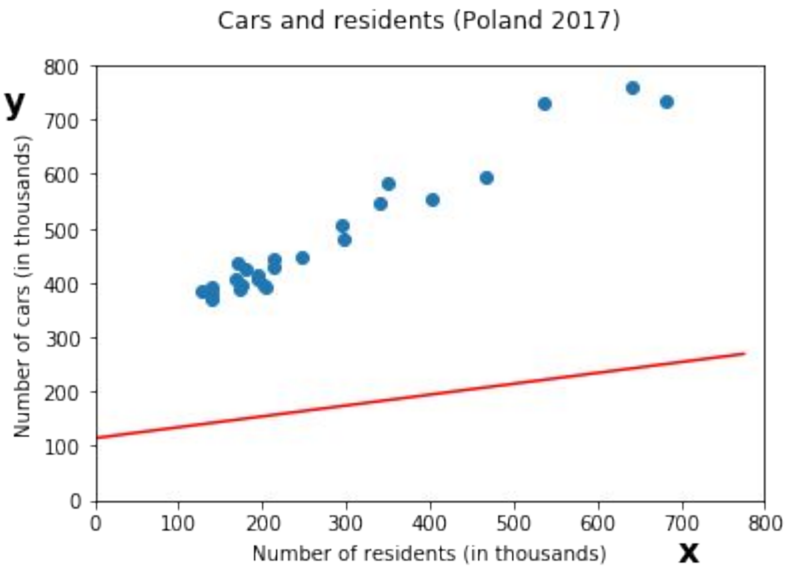
$$\frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) = \frac{1}{m} \sum_{j=1}^{m} (\theta_1 x^{(j)} + \theta_0 - y^{(j)})$$

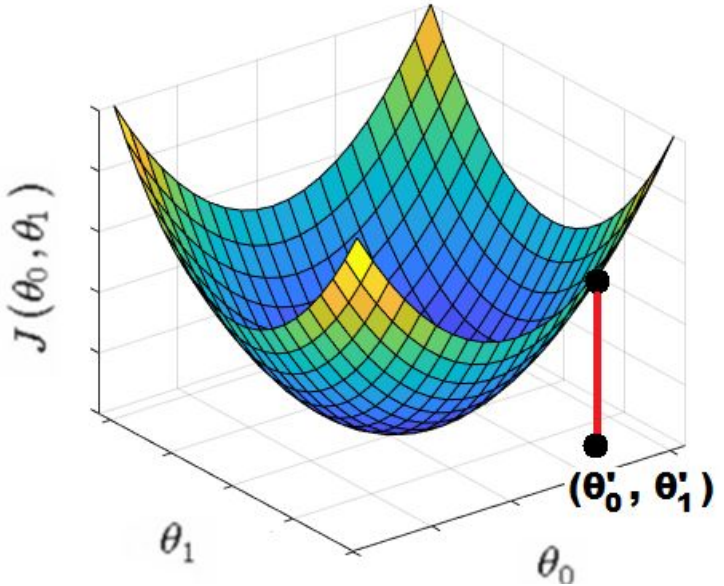
$$\frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1) = \frac{1}{m} \sum_{j=1}^{m} (\theta_1 x^{(j)} + \theta_0 - y^{(j)}) \cdot x^{(j)}$$

alfa: tanulási ráta (learning rate); a lépések mérete skálázható vele

Lineáris regresszió - megoldás

A gradiens módszer alkalmazása, T = 0 (az első lépés előtt)

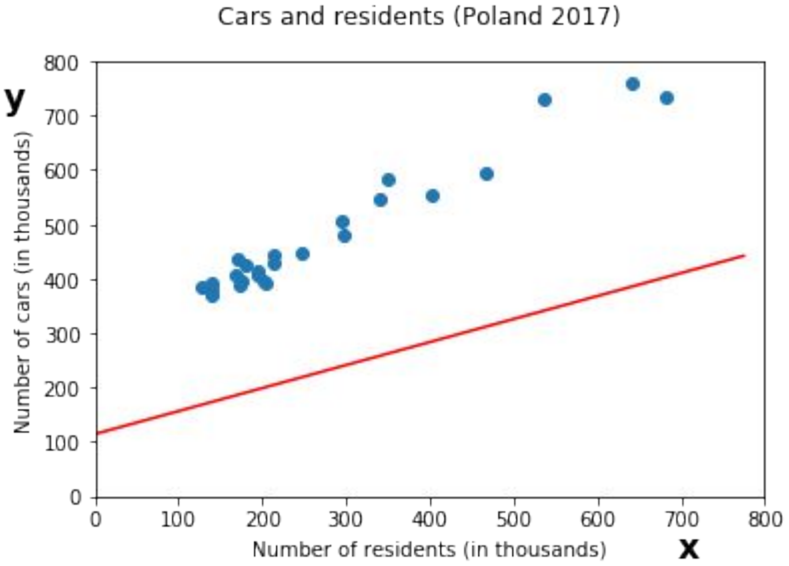


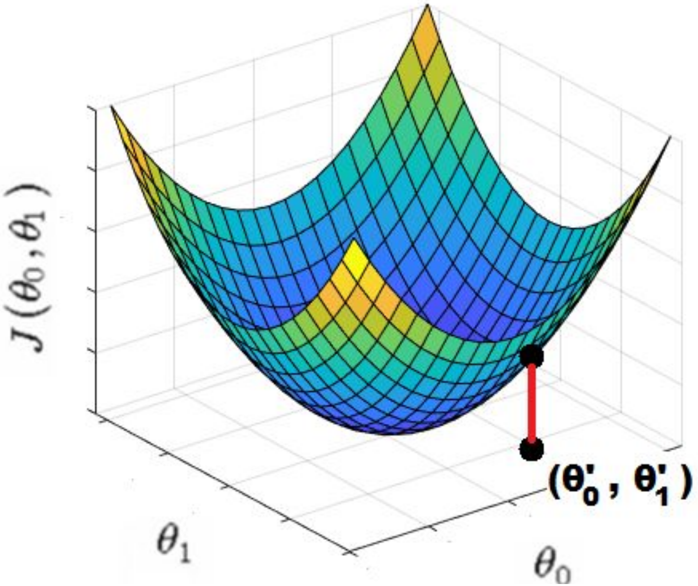


A kezdeti **(θ_0, θ_1)** -t véletlenszerűen válaszottuk.

Lineáris regresszió - megoldás

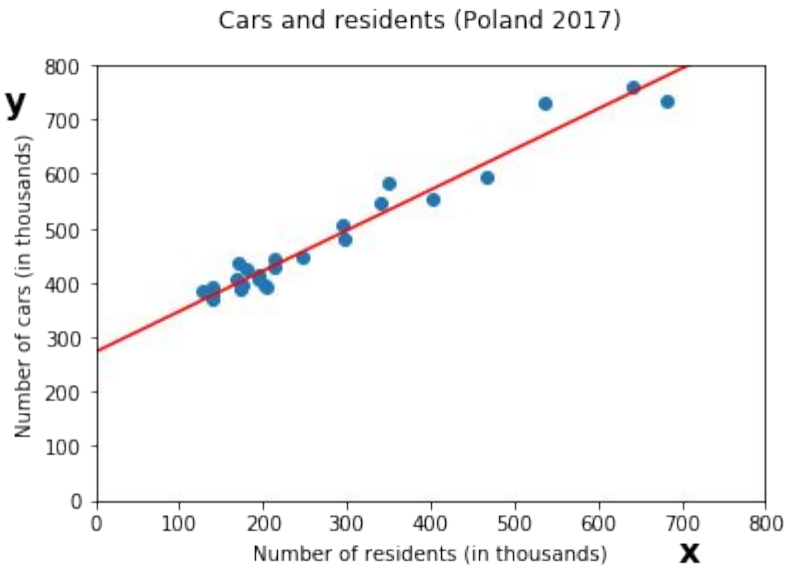
A gradiens módszer alkalmazása, T = 1

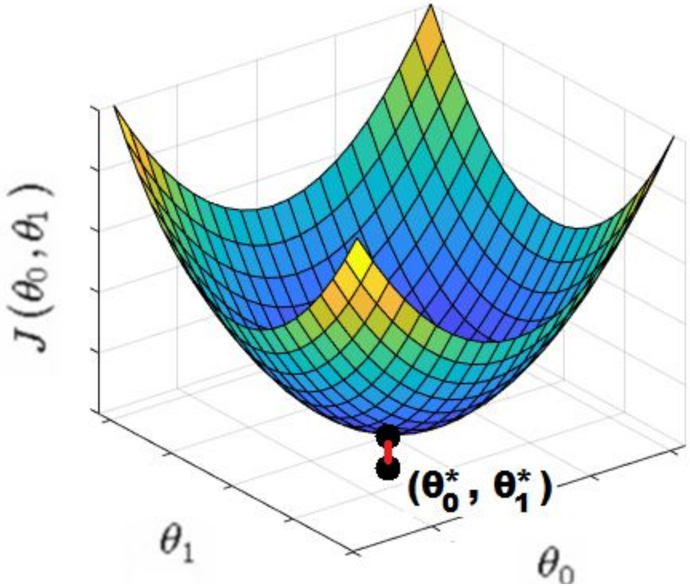




Lineáris regresszió - megoldás

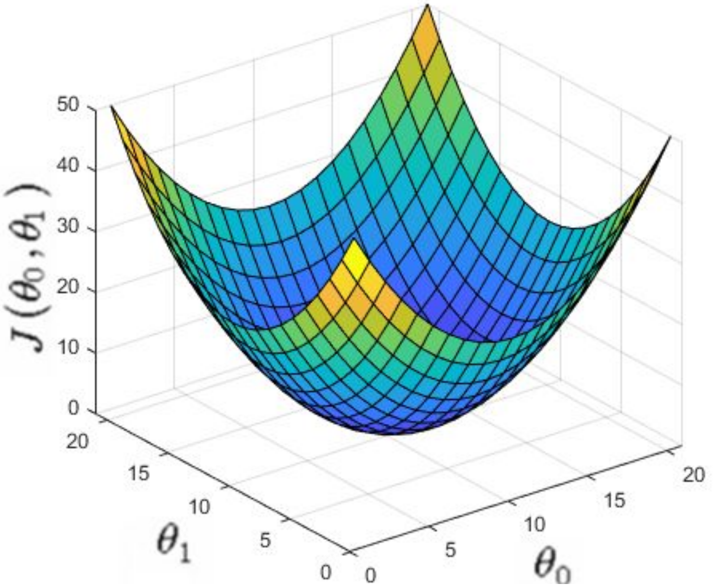
A gradiens módszer alkalmazása, T = <sok>





Lineáris regresszió - megoldás

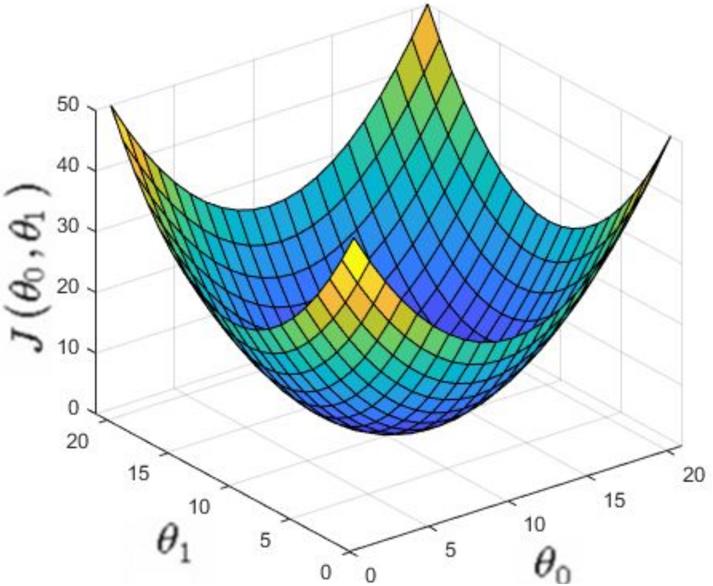
Garantált-e, hogy gradiens módszerrel megtaláljuk a lineáris regresszió optimális (minimális költségű) **megoldását?**



Lineáris regresszió - megoldás

Garantált-e, hogy gradiens módszerrel megtaláljuk a lineáris regresszió optimális (minimális költségű) **megoldását?**

Megfelelően kicsi lépésméret (alfa) esetén igen.



Lineáris regresszió - megoldás

Garantált-e, hogy gradiens módszerrel megtaláljuk **tetszőleges költségfüggvény** globális minimumát?

Lineáris regresszió - megoldás

Garantált-e, hogy gradiens módszerrel megtaláljuk **tetszőleges költségfüggvény** globális minimumát?

Nem. Eljuthatunk az egyik lokális minimum pontba, de amennyiben a költségfüggvény nem konvex, akkor nem garantált, hogy ez a globális minimum lesz.

Lineáris regresszió - megoldás

Garantált-e, hogy gradiens módszerrel megtaláljuk **tetszőleges költségfüggvény** globális minimumát?

Nem. Eljuthatunk az egyik lokális minimum pontba, de amennyiben a költségfüggvény nem konvex, akkor nem garantált, hogy ez a globális minimum lesz.

Vak hegymászó: Le szeretne jutni a völgy legmélyebb pontjába, de csak azt érzi, hogy milyen irányban lejt leginkább a terep a talpa alatt...

Lineáris regresszió - megoldás

Garantált-e, hogy gradiens módszerrel megtaláljuk **tetszőleges költségfüggvény** globális minimumát?

Nem. Eljuthatunk az egyik lokális minimum pontba, de amennyiben a költségfüggvény nem konvex, akkor nem garantált, hogy ez a globális minimum lesz.

A lineáris regresszió nem csak iteratív, de egzakt módon is megoldható. Lásd őszi félév 1. óra, vagy NumMód2...

Lineáris regresszió - megoldás

Miért pont négyzetes hibát használtunk költségként?

(Miért nem mondjuk abszolútértékes hibát?)

Összefoglalás

- **Gépi tanulásra ideális feladatok:** Ismeretlen (hatékony) megoldó-algoritmus, esetleg maga a feladat sem definiálható formálisan (helyette címkézett adathoz igazítunk egy paraméteres modellt).
- **Felügyelt tanulás:** Input-címke párok, az inputból tanuljuk megbecsülni a címkét.
- **Regresszió:** Mintaelemek címkéje folytonos.
- **Lineáris regresszió:** Megpróbáljuk a címkéket az input lineáris függvényeként felírni, majd keressük az egyenes azon paramétereit melyekkel a legjobban közelíthetők az igazi címkék.