W01

Fekete Máté

2020 Április

Tartalomjegyzék

1 Bevezetés

A gépi tanulás nagy vonalakban két kategóriára osztható.
Felügyelt tanulás: Van egy dataset, amire tudjuk, hogy mi a helyes kimenet, és feltételezzük, hogy az input és output között van valami kapcsolat. Az ilyen problémákat két további kategóriára bonthatjuk, a **regressziós**, ahol az inputra egy folytonos függvényt próbálunk illeszteni, ezzel szintén folytonos output-ot képezve (pl. lakás árak jóslása), illetve az **osztályozás**, ahol pedig diszkrét output-ot várunk (pl. egy igen/nem válasz).
Felügyelet nélküli tanulás: Fogalmunk sincs, hogy hogy nézhet ki az output, próbálunk valamilyen kapcsolatot találni.
Clustering: Az input-ot kisebb csoportokra bontja, ahol valamilyen hasonlóságot vagy esetleges kapcsolatot talál a változók között.
Non-clustering: Pl. zajszűrés

2 Modell reprezentáció

Jelölések:
$x^{(i)}$ - bemeneti változók
$y^{(i)}$ - kimeneti vagy cél változók (ezeket akarjuk megjósolni)
$(x^{(i)}, y^{(i)})$ - tanuló példa (training example), ezekből m darab van, i=1,...,m
ezek összessége a tanuló adatkör (training set).
X: bementek halamza
Y: kimenetek halmaza
A célunk egy olyan h hipotézis felépítése, amely az X -¿ Y leképezés függvényét alkotja ($h(x) \approx y$).

3 Költség függvény

A hipotézisünk pontosságának ellenőrzésére használjuk. A kapott és a várt értékek különbségének értékét számítjuk ki,általában a négyzetes hibafüggvényt használva.

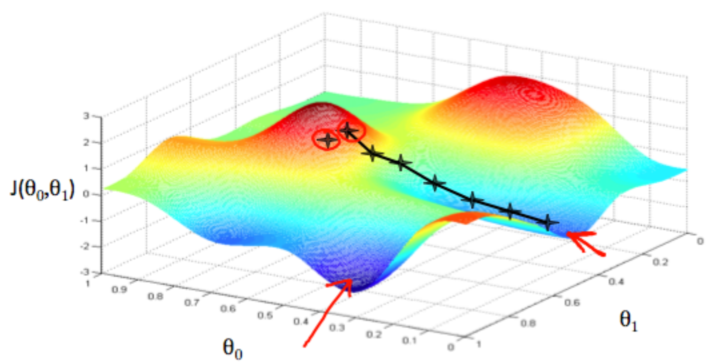
$J(\theta_0, \theta_1) = \frac{1}{2m} \sum\limits_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$
$h_\theta(x^{(i)}) = \theta_0 + \theta_1 x^{(i)}$

Ahol J a költség függvényt jelöli, θ a függvény paramétereit.A hipotézis függvényünk egy szimpla egyenes (jelen esetben).
A célunk ennek a függvénynek a minimalizálása, hiszen ezzel kerül közelebb a jósolt értékünk a valódihoz.
Ha ábrázoljuk a bemenetünket egy grafikonon, ahol az x tengely

egy bemeneti változónk (pl. lakások mérete négyzetméterben), az y tengely pedig a várt kimeneti érték (jelen esetben pl. a lakás ára), akkor a cél, hogy a h hipotézis függvényünk minél jobban illeszkedjen az X és Y által megadott pontokra.

4 Gradiant Descent

Van egy hipotézis függvényünk, illetve le tudjuk ellenőrizni, hogy mennyire illeszkedik jól az adatainkra, viszont a θ paramétereket még nem tudjuk meghatározni.
Ábrázoljuk egy grafikonon a θ_0 és θ_1 értékeket, illetve a velük paraméterett hipotézis függvény költségét.



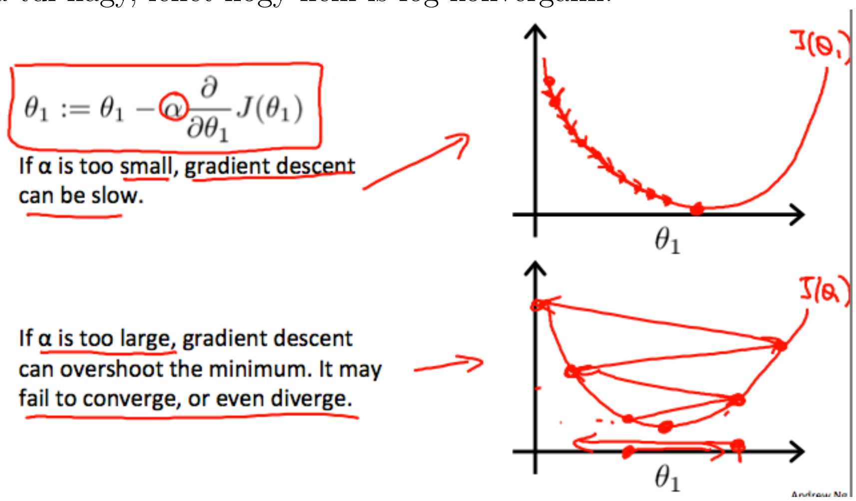
Akkor értük el a célünkat, amikor a költésg függvényünk valamelyik, a piros nyilak által jelölt minimum értékben van.
Ennek elérése érdekében lederiváljuk a költség függvényünket, ami megadja az irányt, amerre csökken a függvény értéke (mindig a legmeredekebb irányt választjuk).
Egy-egy ilyen lépést az X-ek jelülnek az ábrán, ezek távolságát az α, úgynevezett **tanulási ráta** adja meg, amit az algoritmus paraméterként kap.
A kezdőpontunktól függően különböző lokális minimumokat is találhatunk, a fenti ábrán a két bekarikázott X például nem ugyan oda fog konvergálni.

Az algoritmus:

$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$
Amit egészen addig ismétlünk, ameddig el nem érünk egy lokális

minimumot.
j=0,1 a tulajdonságokat jelöli (pl. ha a már említett lakásos példában nem csak a méretet nézzük, hanem a hálószobák számát is, akkor j=0 lehet a méret, j=1 az utóbbi).

FONTOS!
Az összes j értékre egyszerre kell elvégeznünk a θ értékek frissítését, hogy egy iteráción belül ne legyenek egymásra hatással. α értékének megválasztása:
Ha túl kicsit, nagyon sok lépésben fog csak konvergálni az algoritmus.
Ha túl nagy, lehet hogy nem is fog konvergálni.



Mivel a $\frac{\partial}{\partial \theta_1} J(\theta_1)$ derivált érték a 0-ba konvergál ahogy közelítjük a lokális minimumot, illetve a minimumban 0 a derivált, egy jól megválasztott α értékkel egyrészt nem fogunk túllőni a minimumon, másrészt biztosan elérjük azt, anélkül, hogy megváltoztatnánk α értékét.

5 Gradiens módszer lineáris regresszión

Mivel lineáris regresszióhoz már van költség és hipotézis függvényünk, ezért ezeket be tudjuk helyettesíteni a Gradiens módszer képletbe.

$\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum\limits_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})$
$\theta_1 := \theta_1 - \alpha \frac{1}{m} \sum\limits_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x^{(i)}$

Addig ismételjük, amedddig lokális minimumhoz nem érünk.