W02

Fekete Máté

2020 Április

Tartalom jegyzék

1 Többváltozós lineáris regresszió

Jelölések:
$x_j^{(i)}$ - a j jellemző értéke az i-edik tanuló példában
$x^{(i)}$ - az i-edik tanuló példa bemenete (jellemzők összessége)
m - a tanuló példák száma
n - a jellemzők száma

A hipotézis függvény többváltozós formája a következő:

$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + ... + \theta_n x_n$

Pl. θ_0 egy ház alap ára, θ_1 az ár négyzetméterenként, θ_2 az ár emeletenként, x_1 a ház mérete, x_2 az emeletek száma.

Ha a θ és x értékeket egy-egy vektorként tekintjük (x_0-t mindig 1-nek nézzük, így nem változtat a fenti képleten), akkor vektorizálva a képlet:

$\theta^T x$

2 Gradiant Descent több változóval

A képletünk ugyan az, csak el kell végeznünk minden jellemzőre.

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

Amit ugyan úgy ismétlünk, ameddig nem találunk lokális minimumot.

3 Normalizáció

Gyorsíthatunk a Gradiens módszer konvergenciáján azzal, ha a bemeneti értékeinket egy kisebb intervallumra csökkentjük, mivel nagy intervallumokon és egyenletlen eloszlással sokkal lassabban érjük csak el a minimumot.
Legjobb esetben a célünk valami ilyesmi elérése:

$-1 \leq x_{(j)} \leq 1$

Ez nem egy konkrét érték, a célunk csak az, hogy csökkentsük az intervallum méretét.
Általában a következő függvényt használjuk:

$x_j := \frac{x_j - \mu_j}{s_j}$

Ahol μ_j a j jellemző összes értékének átlaga, s_j pedig az értékek intervallumának hossza (max-min).

4 Polinomiális regresszió

Az adatkörünkre sok esetben nem illeszkedik jól egy szimpla egyenes. Ilyen esetekben hozhatunk létre saját jellemzőket, amik a meglévők kombinációi.

Például ha a jelenlegi hipotézis függvényün a $h_\theta(x) = \theta_0 + \theta_1 x_1$, de egy négyzetes függvény görbéje jobban illeszkedne, vehetjük az x_1 változónk négyzetét, ezzel a hipotézist a következő formára módosítva:

$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_1^2$

Ilyen esetben viszont fokozottan fontossá válik a normalizáció, hiszen ha az x_1 intervalluma [1-1000], a hozzáadott tagban már [1-1000000] lesz.

5 Normál egyenlet

Ugyan úgy a költség függvényt minimalizálja, viszont az iteratív megoldás helyett egyetlen lépésben.

$\theta = (X^T X)^{-1} X^T y$

Nem kell normalizációt használni.

5.1 Gradiens módszer vs normál egyenlet

Gradiens módszer:

- Választanunk kell egy α értéket
- Iteratív
- $O(kn^2)$
- Jól működik sok jellemző esetén is

Normál egyenlet:

- Nincs α
- Invertálni kell $X^T X$-et, ami $O(n^3)$ műveletigényű és lehet hogy nem is invertálható
- Lassú, ha túl sok jellemzőnk van

Ha $X^T X$ nem invertálható általában a következők okozhatják:

- Redundáns jellemzők, lineárisan összefüggnek
- Túl sok jellemző, regularizálni kell