Közelítő és szimbolikus számítások

Gradiens módszerek, legkisebb négyzetek

SZTE-TTIK, Számítógépes Optimalizálás Tanszék

A deriváltak használata

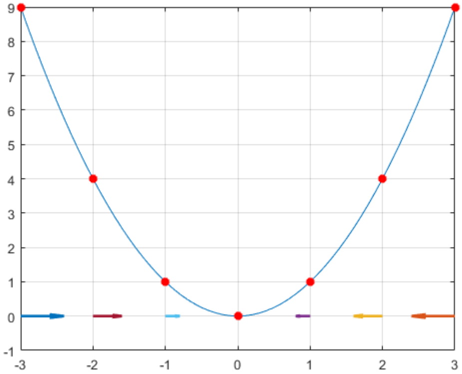
- Kalkusuból tanultuk, hogy a derivált használható a függvény kritikus pontjainak meghatározására
- Az előző órán láttuk, hogy alkalmazható függvények gyökeinek keresésére is (Newton-módszerben)

Most a deriváltnak azon tulajdonságát fogjuk kihasználni, hogy megmutatja, melyik irányba növekszik a függvény.

1D példa

- Nézzük az x^2 függvényt, és a deriváltját ($2x$).
- Minden x pontban a derivált **ellentettje** megmutatja hogy merre csökken a függvény.
- A derivált csak az irányt mutatja, nem azt, hogy pontosan milyen messze vagyunk a minimumtól (az abszolútértéke azonban általában utal a minimumtól vett távolságra)
- pl.: $x = -3$ pontban, a derivált $2 * -3 = -6$, aminek az ellentettje pozitív, tehát a -3-tól pozitív irányba csökken a függvény.

Grafikon



Az ábrán a deriváltak ellentettjeinek 0.1-szeresét ábrázoltuk vektorokként. Pl. 1-ben a derivált 2, így a lila vektor a 0.8-be mutat 1-ből. (1+ -1*2*0.1)

A gradiens

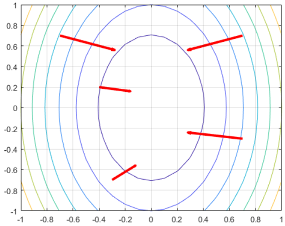
A fenti 1D példa általánosítható többdimenziós függvényekre (amikor több változó van a függvényben pl.: $3x^2 + y$), a következőképpen:

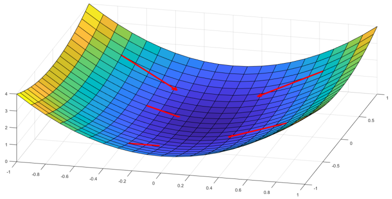
- A parciális deriváltakból képezhető egy vektor, amit **gradiens vektor**nak nevezünk. Jele: ∇
- Ezen vektor minden pontban megmutatja, hogy merre növekszik leginkább a függvény
- Ennek ellentettje a csökkenés irányába mutat

Példa gradiensre

- Legyen a függvény: $f(x, y) = 3x^2 + y^2$
- A parciális deriváltak: $\frac{\partial f(x,y)}{\partial x} = 6x$, $\frac{\partial f(x,y)}{\partial y} = 2y$
  Tehát a gradiens vektor $\nabla f(x, y) = (6x, 2y)$
  pl.: az $(-0.4, 0.2)$ pontban a gradiens vektor
  $\nabla f(-0.4, 0.2) = (-2.4, 0.4)$.

Néhány példa 3D-ben





5 pontban kirajzoltuk, hogy az adott pontból hova mutat a gradiensvektor ellentettjének 1/10-e. Bal oldalt az (x,y) síkon láthatóak, jobb oldalt kivetítve a függvényre

A gradiens módszer

Cél: egy függvény minimumpontjának meghatározása
Az eljárás

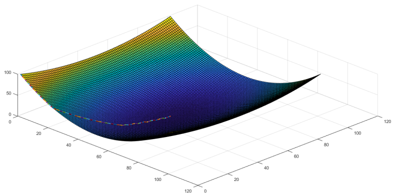
1. Induljunk ki egy tetszőleges x_0 pontból.
2. Lépjünk a negatív gradiens irányába $(-\nabla)$ valamekkorát (α):

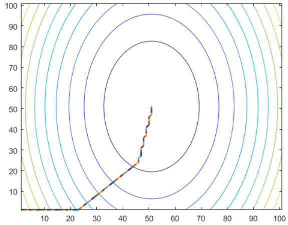
$$x_{k+1} = x_k - \nabla f(x_k) * \alpha$$

Tulajdonságok:

- Iterációs módszer
- Konvex függvények esetén a globális minimumot találjuk meg
- Az α meghatározása nem egyszerű feladat. A példákban úgy választottuk, hogy konstansnyit lépjünk az adott irányba

Gradiens módszer ábrán



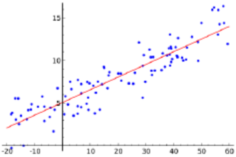


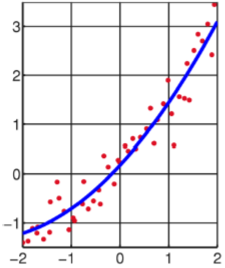
Az α lépésköz konstans értékre volt állítva (i.e. mindig ugyanakkorát léptünk az adott irányba) Lásd 'animálva' a DemonstrateGradientDescent.m script-ben.

A legkisebb négyzetek módszere (LNM)

- A polinomos órán láttuk, hogy sok adapontra a Lagrange interpoláció használhatatlanul rossz eredményt ad.
- Ehelyett használtuk a `polyfit` utasítást, amely a legjobban illeszkedő adott fokú polinomot illesztette a pontokra.
- A gradiens módszer megismerése után választ adhatunk rá, hogy hogyan működik ez az illesztés.

LNM általában





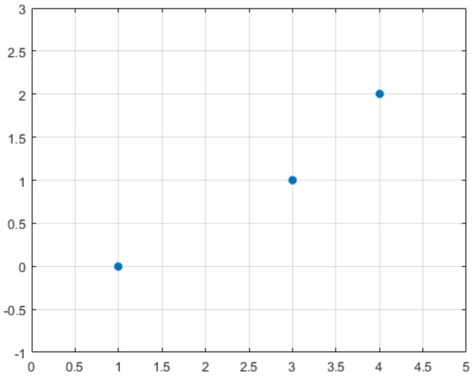
- Adott n db alappont (x_i, y_i)
- Ezekre szeretnénk illeszteni egy adott típusú függvényt, $f(x, \beta)$, ahol β a függvény paramétere
- Az illeszkedés hibája egy pontban (négyzetes eltérés):

$$r_i = y_i - f(x_i, \beta)$$

- Cél: $\sum_{i=1}^{n} r_i^2 \rightarrow min$

Egy konkrét példa

Vegyük az alábbi adathalmazt: (x_i, y_i) párok:
$(1, 0), (3, 1), (4, 2)$

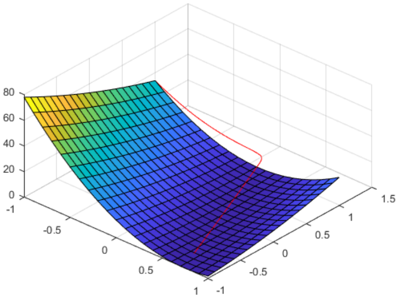


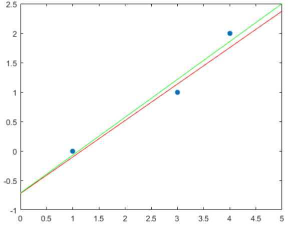
Adathalmaz: $(1,0),(3,1),(4,2)$
Illesszünk erre egy egyenest: $f(x) = a * x + b$

A korábban jelölt β függvényparaméter itt az a és a b

Keressük azt az a és b-t amire a következő minimális:
$\sum_{i=1}^{3} r_i^2 = \sum_{i=1}^{3}(y_i - (a * x_i + b))^2$
$= (0 - (a * 1 + b))^2 + (1 - (a * 3 + b))^2 + (2 - (a * 4 + b))^2 =$
$= ... = 26a^2 + 3b^2 - 22a - 6b + 16ab + 5$
Ez pedig szimplán egy kétváltozós függvény, amit minimalizálhatunk pl. gradiens módszerrel.
A függvény gradiense:

$$\nabla f(a,b) = (52a - 22 + 16b, 6b - 6 + 16a)$$





Az illesztett egyenes paraméterei: $a = 0.6190, y = -0.7218$ (piros)

A `polyfit(x,y,1)` eredménye pedig $a = 0.6429, y = -0.7143$ (zöld)

A konjugált gradiens módszer

- A gradiens módszer egy változata, amely egy speciális feladat: szimmetrikus és pozitív definit mátrixú lineáris egyenletrendszerek ($Ax = b$) megoldására alkalmas
- Definiálható egy célfüggvény, amelynek a minimumhelye éppen az egyenletrendszer megoldása

$$q(x) = \frac{1}{2}x^TAx - x^Tb$$

- Ebben a speciális esetben sok általános számítás leegyszerűsödik:
  - A negatív gradiens épp az ún. reziduális vektor: $-\nabla q(x) = b - Ax$
  - Az optimális α képlettel meghatározható
  - Az egymást követő gradiensek merőlegesek egymásra (innen a konjugált név)