MAT 1080

3.3

3.3. Partiell derivasjon og gradient

Her kommer viktige begreper!

For $f: \mathbb{R}^n \to \mathbb{R}$, en funksjon i n variable, kan vi se på den deriverte av én av variablene, mens de øvrige holdes fast

Def. 3.3.1 $f(x_1, x_2, \dots, x_n)$, $f: \mathbb{R}^n \to \mathbb{R}$

Den partielt deriverte av f m.h.t. x_i (den i'te partielt deriverte) er

$$\frac{\partial f}{\partial x_i} = \lim_{h \to 0} \frac{f(x_1, \dots, x_i + h, \dots x_n) - f(x_1, \dots, x_n)}{h}$$

—

NB: • $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$ holdes fast

• Men $\frac{\partial f}{\partial x_i}$ avhenger av alle variablene

Made with Goodnotes

$\frac{\partial f}{\partial x_i}$ lesas "del f del x_i"

Ehs: 1.) $f(x,y) = 3x^2 + x + \sin y$

$\frac{\partial f}{\partial x} = 6x + 1$, $\frac{\partial f}{\partial y} = \cos y$

2.) $f(x,y) = x \cos y$

$\frac{\partial f}{\partial x} = \cos y$, $\frac{\partial f}{\partial y} = -x \sin y$

Så: • $\frac{\partial f}{\partial x}$ sier oss hvor fort f vokser i retning langs x-aksen, altså x øker

• $\frac{\partial f}{\partial y}$ --- langs y-aksen

Made with Goodnotes

Tillbake til $f: \mathbb{R}^n \to \mathbb{R}$, $f(x_1, x_2, \dots, x_n)$

Da har vi de partielle deriverte

$$\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n}$$

i et punkt $x = (x_1, x_2, \dots, x_n)$. Lager en vektor med disse

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right)$$

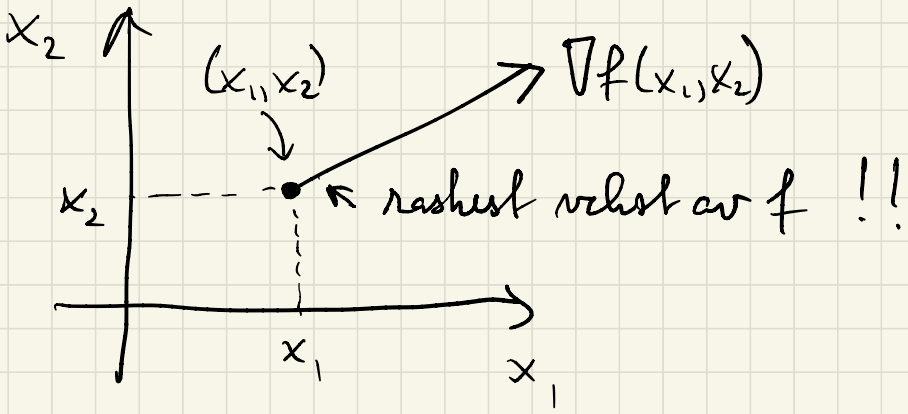
Denne kalles <u>gradienten</u> til f i punktet x

Med $n = 2$ har vi

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} \right).$$

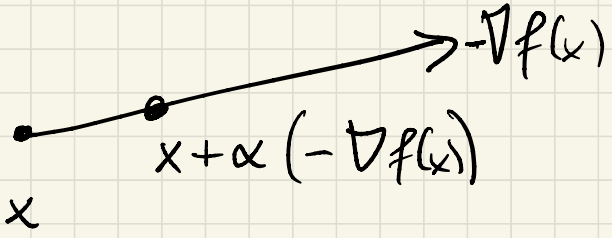
<u>For optimering er gradienten svært viktig fordi gradienten peker i retningen der f vokser raskest lokalt</u>

Made with Goodnotes



Tilsvarende: retningen $-\nabla f(x)$ gir største reduksjon (lavere verdi) for f

Så: Skal vi minimere f så går vi (litt/passende langt) i retning $-\nabla f(x)$.



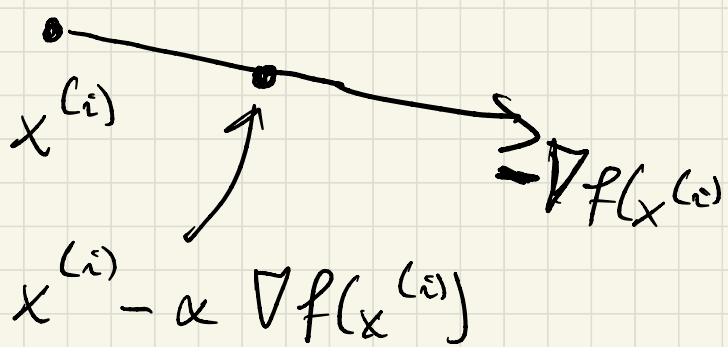
Made with Goodnotes

I Maskinlæring og optimering vil vi minimere en viss funksjon f i mange variable. Metoden er å lage en følge (skvens) av punkter (i $\mathbb{R}^n$):

$$x^{(1)}, x^{(2)}, x^{(3)}, \dots, x^{(k)}$$

der $f(x^{(1)}) > f(x^{(2)}) > f(x^{(3)}) > \dots > f(x^{(k)})$

(noen ganger må vi nøye oss med $=$ i noen iterasjoner). Og i hver iterasjon:



Made with Goodnotes

Metoden kalles steepest descent

og brukes svært mye, gjerne i form av ulike varianter.

Hvorfor gir retning $\nabla f(x)$ mest vekst?

Kan vise Taylors formel (første kens)

$$f(x+h) \approx f(x) + \nabla f(x) \cdot h$$

der $\cdot$ er prikkprodukt:

$$(x_1, \dots, x_n) \cdot (y_1, \dots, y_n) = x_1 y_1 + \dots + x_n y_n$$

Vi vet $\nabla f(x) \cdot h \leq \|\nabla f(x)\| \, \|h\|$

som er Cauchy-Schwarz ulikhet; likhet hvis og bare hvis $h \parallel \nabla f(x)$ parallelle.

Så med $\|h\| = 1$ får vi størst økning $f(x+h) - f(x)$ ved å velge $h \parallel \nabla f(x)$

d.v.s. $h = \nabla f(x) / \|\nabla f(x)\|$

Made with Goodnotes

Tilbake til partielt deriverte. Man kan gjøre dette gjentatte ganger:

$\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)$ betyr først part. der. mhp x, så ta part. der. mhp y.

Under svake betingelser (om kontinuitet) vil

$$\frac{\partial}{\partial x_i}\frac{\partial f}{\partial x_j} = \frac{\partial}{\partial x_j}\frac{\partial f}{\partial x_i}$$

så rekkefølgen ikke spiller noen rolle.

Eks

$$f(x,y) = x\cos y$$

$$\frac{\partial f}{\partial x} = \cos y \quad , \quad \frac{\partial f}{\partial y} = -x\sin y$$

$$\frac{\partial}{\partial y}\frac{\partial f}{\partial x} = -\sin y \quad , \quad \frac{\partial}{\partial x}\frac{\partial f}{\partial y} = -\sin y$$

Made with Goodnotes

Funksjoner fra $\mathbb{R}^n$ til $\mathbb{R}^m$

Hittil sett på $f: \mathbb{R}^n \to \mathbb{R}$.

Men vi kan også funksjoner der funksjonsverdien er, ikke et tall, men en vektor

Skriver da $F: \mathbb{R}^n \to \mathbb{R}^m$

og

$$F(x) = (f_1(x), f_2(x), \dots, f_m(x))$$

der komponentfunksjonene $f_1, f_2, \dots, f_m$ er $f_i: \mathbb{R}^n \to \mathbb{R}$. Så $F = (f_1, f_2, \dots, f_m)$

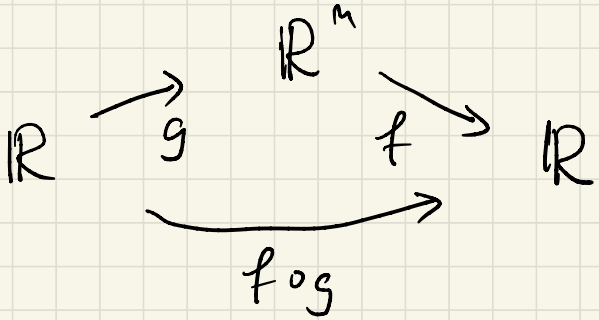
Eks: $F(x,y) = (x+y, x^2+y^2)$

$F(x_1,x_2,x_3) = (x_1+2x_2, x_1 \cdot x_2, x_1^2 - x_2^2 \cdot x_3)$

$\vdots$

Made with Goodnotes

Kjerneregel. Finnes også for funksjoner i flere variable. Vi ser på et spesialtilfelle



Teorem 3.3.3 $f : \mathbb{R}^n \to \mathbb{R}$

$g = (g_1, \dots, g_n) : \mathbb{R} \to \mathbb{R}^n$

$h = f \circ g$ d.v.s. $h(t) = f(g(t))$

Da er

$$h'(t) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \cdot \frac{dg_i}{dt}$$

—

Bevis: Se kompendium.

Made with Goodnotes

Ehs:

$$g(t) = (t, t^2+t), \quad f(x,y) = x^2 - y^2$$

$$g : \mathbb{R} \to \mathbb{R}^2, \quad f : \mathbb{R}^2 \to \mathbb{R}$$

$$h = f \circ g : \mathbb{R} \to \mathbb{R}$$

$$h(t) = x^2 - y^2 = t^2 - (t^2+t)^2$$

$$g = (g_1, g_2)$$

Kjæreregel gir:

$$h'(t) = \frac{\partial f}{\partial x} \cdot \frac{dg_1}{dt} + \frac{\partial f}{\partial y} \cdot \frac{dg_2}{dt}$$

$$= 2t \cdot 1 + (-2)(t^2+t) \cdot (2t+1)$$

$$= 2t - 4t^3 - 4t^2 - 2t^2 - 2t$$

$$= -4t^3 - 6t^2$$

Dette stemmer med h' over, da $h(t) = t^2 - (t^2+t)^2$ hvis vi deriverer direkte.

Made with Goodnotes