Assignment 3 — MASD

Schmidt, Victor Alexander, `rqc908`
Ibsen, Helga Rykov, `mcv462`
Barchager, Thomas Haulik, `jxg170`

Wednesday 22:00, September 28th

Opgave 1

Vi skal finde alle de stationære punkter for fire forskellige funktioner, samt identificere hvorvidt de er et lokalt minimum, et lokalt maksimum, et vendetangent for funktioner af én variabel eller et seddelpunkt for funktioner af to variable.

a)

$$f(x) = x^3 - 2x^2 + 11x + 7$$

Først differentiere vi funktionen

$$f'(x) = 3x^2 - 4x + 11$$

For at finde minimum og eller maksimum skal vi sætter funktionen lig med 0.

$$3x^2 - 4x + 11 = 0$$

Vi regner nu diskriminanten ud:

$$\begin{aligned} d &= 16 - 4 \cdot (3 \cdot 11) \\ d &= -116 \end{aligned}$$

Da $d < 0$ har ligningen ingen løsninger. Det betyder at funktionen f har ingen minimum eller maksimum, hvilket også tydeligt kan ses på Figur 1, hvor den røde graf for f' ligger over x-aksen.

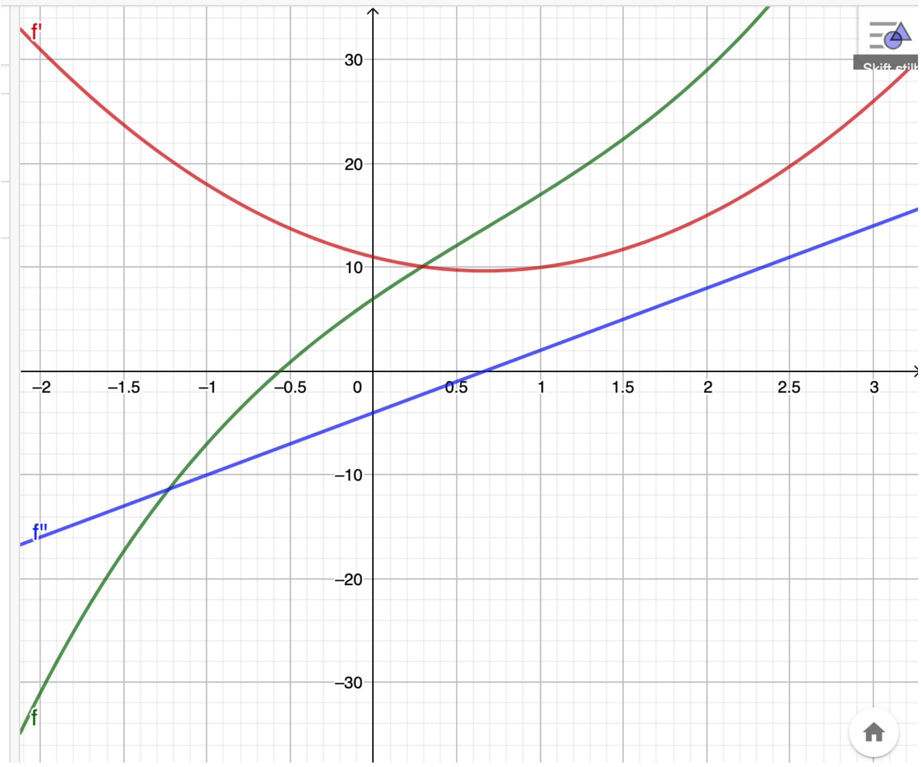
Vi undersøger nu om funktionen har en vendetangent, hvilket vi gør ved at differentiere den afledede funktion f':

$$\begin{aligned} f'(x) &= 3x^2 - 4x + 11 \\ f''(x) &= 6x - 4 \end{aligned}$$

Vi skal nu tjekke om f har et vendetangent og løser derfor ligningen:

$$\begin{aligned} 6x - 4 &= 0 \\ 6x &= 4 \\ x &= \frac{4}{6} = \frac{2}{3} \approx 0,67 \end{aligned}$$

Det kan man også se på Figur 1, hvor den grønne graf skifter fra at krumme nedad til at krumme opad omkring der, hvor x = 0.67.



Figur 1: Funktion a

b)

$$f(x) = x^n \; ln \; x, n \; \epsilon \; \mathbb{N} \backslash \{0\}$$

Vi starter med at anvende produktreglen:

$$\begin{aligned} f(x)' = (x^2 \cdot ln \; x)' &= (x^n)' \cdot ln \; x + x^n \cdot (ln \; x)' \\ &= n \cdot x^{n-1} \cdot ln \; x + x^n \cdot x^{n-1} \\ f(x)' &= x^{n-1} \cdot (n \cdot ln \; x + 1) \end{aligned}$$

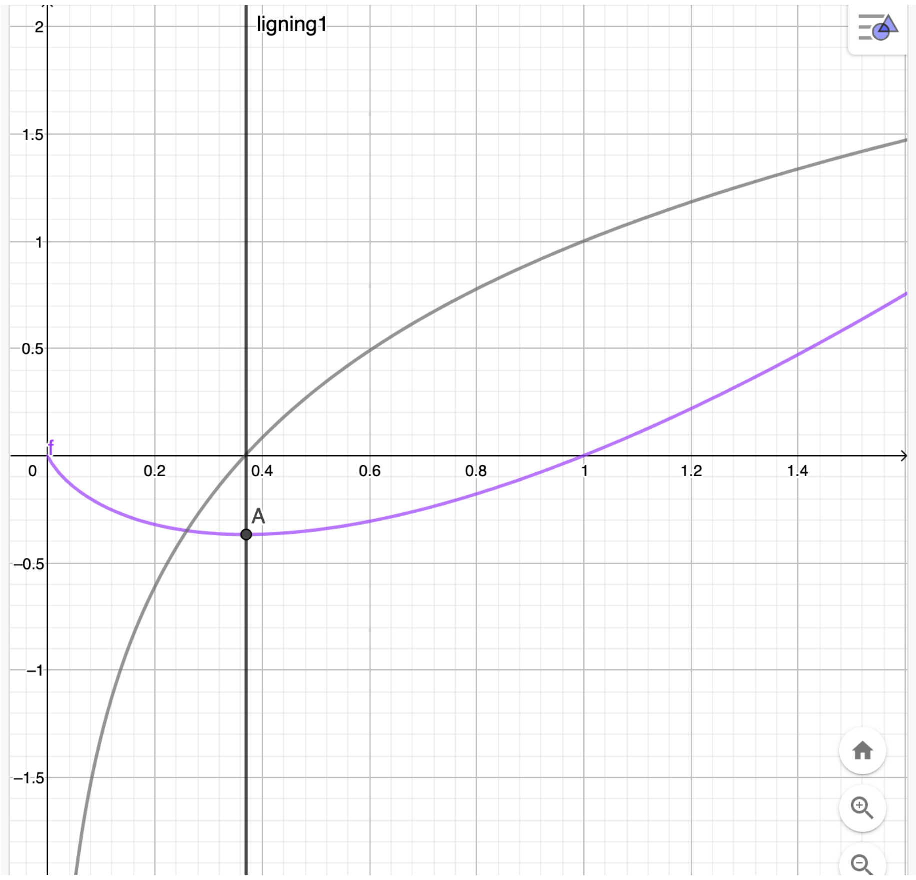
Vi undersøger nu om funktionen har kritiske punkter (vi ved at $n \neq 0$):

$$f(x)' = x^{n-1} \cdot (n \cdot ln \; x + 1) = 0$$

Vi anvender nulreglen og eftersom x^{n-1} er en potensfunktion, ved vi at den ikke kan være lig med 0. Vi løser derfor følgende ligning:

$$\begin{aligned} n \cdot ln \; x + 1 &= 0 \\ n \cdot ln \; x &= -1 \\ ln \; x &= \frac{-1}{n} \\ x &= e^{\frac{-1}{n}} \end{aligned}$$

Ved hjælp af GeoGebra kan vi tydeligt se, at funktionen f har et minimum i det punkt hvor f' krydser x-aksen (x $= 0.37$).
Vha. GeoGebra har vi også fundet ud af at ligningen $f(x)'' = 0$ har ingen løsninger, hvilket betyder at funktionen f har ingen vendetangent.



Figur 2: Funktion b

c)

$$f(x_1, x_2) = 3x_1^2 + 2x_1x_2 + 2x_2^2 - 6$$

Da f er en funktion af to variable, starter vi med at differentiere den mht. x_1 og x_2:

$$\frac{\partial f}{\partial x_1} = 3x_1 + 2x_2$$
$$\frac{\partial f}{\partial x_2} = 2x_1 + 4x_2$$

Vi isolerer x_1 i den første ligning:

$$\begin{cases} 3x_1 + 2x_2 = 0 \\ 2x_1 + 4x_2 = 0 \end{cases}$$

$$\begin{cases} 6x_1 + 4x_2 = 0 \\ -2x_1 - 4x_2 = 0 \end{cases}$$

$$4x_1 = 0$$
$$x_1 = 0$$

Vi kan nu indsætte den fundet variable x_1 ind i den anden ligning, og isolere x_2.

$$\begin{aligned} 2x_1 + 4x_2 &= 0 \\ 2 \cdot (0) + 4x_2 &= 0 \\ 4x_2 &= 0 \\ x_2 &= 0 \end{aligned}$$

Ved at differentiere x_1 og x_2 igen, kan vi bestemme arten af det kritiske punkt $f'(0,0) = 0$. For læsbarhedens skyld, indfører vi notationen for hver af de partielle afledede:

$$\begin{aligned} \frac{\partial^2 f}{\partial^2 x_1} &= (3x_1 + 2x_2)' = 3 = r \\ \frac{\partial^2 f}{\partial^2 x_2} &= (2x_1 + 4x_2)' = 4 = t \\ \frac{\partial^2 f}{\partial^2 x_1 \partial^2 x_2} &= 0 = s \end{aligned}$$

$$\begin{aligned} r \cdot t - s^2 = 3 \cdot 4 - 0 &= 12 \\ r = 3 &> 0 \end{aligned}$$

Det vil sige at i punktet $f'(0,0) = 0$ har vores funktion f altså et lokalt minimum.

d)

$$f(x_1, x_2) = 3x_1^2 + 4x_1x_2 + x_2^2$$

Vi starte med at differentiere hver variabel:

$$\begin{aligned} \frac{\partial f}{\partial x_1} &= 6x_1 + 4x_2 \\ \frac{\partial f}{\partial x_2} &= 4x_1 + 2x_2 \end{aligned}$$

Vi isolere x_1 i den første ligning:

$$\begin{cases} 6x_1 + 4x_2 \\ 4x_1 + 2x_2 \end{cases}$$

$$\begin{cases} 6x_1 + 4x_2 \\ -8x_1 - 4x_2 \end{cases}$$

$$-2x_1 = 0$$
$$x_1 = 0$$

Vi kan nu indsætte den fundet variable x_1 ind i den anden ligning, og isolere x_2.

$$4x_1 + 2x_2 = 0$$
$$4 \cdot (0) + 2x_2 = 0$$
$$2x_2 = 0$$
$$x_2 = 0$$

Ved at differentiere x_1 og x_2 igen, kan vi bestemme arten af det kritiske punkt $f'(0,0) = 0$. For læsbarhedens skyld, indfører vi notationen for hver af de partielle afledede:

$$\frac{\partial^2 f}{\partial^2 x_1} = (6x_1 + x_2)' = 6 = r$$
$$\frac{\partial^2 f}{\partial^2 x_2} = (4x_1 + 2x_2)' = 2 = t$$
$$\frac{\partial^2 f}{\partial^2 x_1 \partial^2 x_2} = 0 = s$$

$$r \cdot t - s^2 = 6 \cdot 2 - 0 = 12$$
$$r = 6 > 0$$

Ligesom i opgave c, har vi en funktion hvor i punktet $f'(0,0) = 0$ har den et lokalt minimum.

Opgave 2

a)

Hvis vi antager at matrix A indeholder de seks personers ratings fra 1 til 10 mht. 2 filmgenre — gyser og romantiske film — og matrix B indeholder ratings fra 1 til 10 af de 10 film mht. hvor meget de svarer til de to filmgenre, og hvis vi så tager de to matricer A og B og ganger dem sammen, så skulle det gerne være at M matrix indeholdte de sandsynlige ratings på de manglende pladser. Altså, hvis Fabio fx havde set "Halloween", så ville han sandsynligt have givet en sandsynlig rating for den.

Med andre ord kan modellen bruges til at guide folk til de film de formindelig vil kunne lide.

b)

Eftersom der står i opgavebeskrivelsen at error-funktion $E(A, B)$ giver et norm af en vektor, hvor I matrix indeholder kun et og nul taller på de pladser, hvor matrix M har henholdsvis et tal fra 1 til 10 og ingenting ”-”, så ville det ikke gøre nogen forskel, at vi så løfter I op i 2. eller ej. Det er derfor kun interessent at kigge på differencen mellem M og produktet af de to matricer A og B mht. koordinaterne.
Vi starter derfor med at fremstille produktet de to matricer A og B mht. deres koordinater.

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \\ a_{51} & a_{52} \\ a_{61} & a_{62} \end{bmatrix}, B = \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} & b_{16} & b_{17} & b_{18} & b_{19} & b_{110} \\ b_{21} & b_{22} & b_{23} & b_{24} & b_{25} & b_{26} & b_{27} & b_{28} & b_{29} & b_{210} \end{bmatrix}$$

$$A \cdot B \begin{bmatrix} a_{11} \cdot b_{11} + a_{12} \cdot b_{21} & ... & a_{11} \cdot b_{110} + a_{12} \cdot b_{210} \\ a_{21} \cdot b_{11} + a_{22} \cdot b_{21} & ... & a_{21} \cdot b_{110} + a_{22} \cdot b_{210} \\ a_{31} \cdot b_{11} + a_{32} \cdot b_{21} & ... & a_{31} \cdot b_{110} + a_{32} \cdot b_{210} \\ a_{41} \cdot b_{11} + a_{42} \cdot b_{21} & ... & a_{41} \cdot b_{110} + a_{42} \cdot b_{210} \\ a_{51} \cdot b_{11} + a_{52} \cdot b_{21} & ... & a_{51} \cdot b_{110} + a_{52} \cdot b_{210} \\ a_{61} \cdot b_{11} + a_{62} \cdot b_{21} & ... & a_{61} \cdot b_{110} + a_{62} \cdot b_{210} \end{bmatrix}$$

Vi tager også 6 x 10 matrix M for at fremstille differencen:

$$M = \begin{bmatrix} M_{11} & M_{12} & M_{13} & M_{14} & M_{15} & M_{16} & M_{17} & M_{18} & M_{19} & M_{110} \\ M_{21} & M_{22} & M_{23} & M_{24} & M_{25} & M_{26} & M_{27} & M_{28} & M_{29} & M_{210} \\ M_{31} & M_{32} & M_{33} & M_{34} & M_{35} & M_{36} & M_{37} & M_{38} & M_{39} & M_{310} \\ M_{41} & M_{42} & M_{43} & M_{44} & M_{45} & M_{46} & M_{47} & M_{48} & M_{49} & M_{410} \\ M_{51} & M_{52} & M_{53} & M_{54} & M_{55} & M_{56} & M_{57} & M_{58} & M_{59} & M_{510} \\ M_{61} & M_{62} & M_{63} & M_{64} & M_{65} & M_{66} & M_{67} & M_{68} & M_{69} & M_{610} \end{bmatrix}$$

Hvis vi fx tager række 3 og søjle 1 og beregner differencen, så får vi:

$$(M_{31} - (a_{31} \cdot b_{11} + a_{32} \cdot b_{21})$$

Og så løfter vi udtrykket op i 2. og får:

$$(M_{31} - (a_{31} \cdot b_{11} + a_{32} \cdot b_{21}))^2$$

hvilket også passer med udtrykket inde i parentesen i error funktion $E(A, B)$:

$$(M_{ij} - (a_{i1}b_{1j} + a_{i2}b_{2j}))^2$$

Opgave 3

Vi skal vise at følgende partielt afledte gælder:

$$\frac{\partial E}{\partial a_{km}} = 2\sum_{j=1}^{10} I_{kj}(-M_{kj}b_{mj} + a_{k1}b_{1j}b_{mj} + a_{k2}b_{2j}b_{mj})$$

$$\frac{\partial E}{\partial b_{ml}} = 2\sum_{i=1}^{6} I_{il}(-M_{il}a_{im} + a_{i1}b_{im}b_{1l} + a_{i2}b_{im}b_{2l})$$

I denne opgave gøres der brug af regler til differentiation. Som genopfriskning:

Kædereglen:

$$f'(g(x)) = f'(g(x)) \cdot g'(x) \tag{1}$$

Produktreglen:

$$(f(x) \cdot g(x))' = f'(x) \cdot g(x) + f(x) \cdot g'(x) \tag{2}$$

Brøkreglen:

$$\left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x) \cdot g(x) - f(x) \cdot g'(x)}{(g(x))^2} \tag{3}$$

Vi løser for $\frac{\partial E}{\partial a_{km}}$

Vi starter fra sidste opgave, koordinaterne som skrevet i "the error function" $E(A, B)$:

$$\sum_{i=1}^{6}\sum_{j=1}^{10} I_{ij}(M_{ij} - (a_{i1}b_{1j} + a_{i2}b_{2j}))^2$$

$$\frac{\partial E}{\partial a_{km}} = \frac{\partial E}{\partial a_{km}} \sum_{i=1}^{6}\sum_{j=1}^{10} I_{ij}(M_{ij} - (a_{i1}b_{1j} + a_{i2}b_{2j}))^2$$

Da vi ved at $\sum_{i=1}^{6}$ kun indeholder a_{km} i leddet når $i = k$, vil alle summe når $i \neq k$ blive konstante - og da vi er igang med at differentiere vil disse konstante gå bort. Derfor vil alle i'er i udtrykket blive til k'er og summen kan bortskrives:

$$\frac{\partial E}{\partial a_{km}} \sum_{j=1}^{10} I_{kj}(M_{kj} - (a_{k1}b_{1j} + a_{k2}b_{2j}))^2$$

Vi kan nu begynde at aflede:

$$\frac{\partial E}{\partial a_{km}} \sum_{j=1}^{10} I_{kj}(M_{kj} - (a_{k1}b_{1j} + a_{k2}b_{2j}))^2$$

Der gøres brug af kædereglen:

$$2\sum_{j=1}^{10} I_{kj}(M_{kj} - (a_{k1}b_{1j} + a_{k2})) \cdot (M_{kj} - (a_{k1}b_{1j} + a_{k2}b_{2j}))'$$

$$2\sum_{j=1}^{10} I_{kj}(M_{kj} - (a_{k1}b_{1j} + a_{k2})) \cdot (-a_{k1}b_{1j} - a_{k2}b_{2j})'$$

Det sidste led der skal differentieres, vil afhænge af om $m = 1$ eller $m = 2$, da enten a_{k1} eller a_{k2} vil være konstant. Det vil sige at vi kan sige at det led der ikke ikke forsvinder vil være $a_{km}b_{mj}$, da den afledte er afhængig af m's værdi:

$$2\sum_{j=1}^{10} I_{kj}(M_{kj} - (a_{k1}b_{1j} + a_{k2})) \cdot (-a_{km}b_{mj} - k)'$$

... hvor k er det konstante led. Vi fortsætter:

$$2\sum_{j=1}^{10} I_{kj}(M_{kj} - (a_{k1}b_{1j} + a_{k2})) \cdot (-b_{mj})$$

$$\frac{\partial E}{\partial a_{km}} = 2\sum_{j=1}^{10} I_{kj}(-M_{kj}b_{mj} + a_{k1}b_{1j}b_{mj} + a_{k2}b_{2j}b_{mj})$$

□

Vi løser for $\frac{\partial E}{\partial b_{ml}}$

Hvis vi gør brug af samme metode og samme logik som da vi løste for $\frac{\partial E}{\partial a_{km}}$, kan vi hurtigt lave følgende udregninger (i dette tilfælde vil l erstatte j, istedet for at k erstatter i):

$$\frac{\partial E}{\partial b_{ml}} = \frac{\partial E}{\partial b_{ml}} \sum_{i=1}^{6}\sum_{j=1}^{10} I_{ij}(M_{ij} - (a_{i1}b_{1j} + a_{i2}b_{2j}))^2$$

$$\frac{\partial E}{\partial b_{ml}} \sum_{i=1}^{6} I_{il}(M_{il} - (a_{i1}b_{1l} + a_{i2}b_{2l}))^2$$

$$2\sum_{i=1}^{6} I_{il}(M_{il} - (a_{i1}b_{1l} + a_{i2}b_{2l})) \cdot (M_{il} - (a_{i1}b_{1l} + a_{i2}b_{2l}))'$$

Igen er vi i gang med at finde den afledte ift. b_{ml}, som afhænger af om $m = 1$ eller om $m = 2$. Vi bruger samme argumentation som i udledningen af $\frac{\partial E}{\partial a_{km}}$. Derfor vil følgende gælde:

$$2\sum_{i=1}^{6} I_{il}(M_{il} - (a_{i1}b_{1l} + a_{i2}b_{2l})) \cdot (-a_{im}b_{ml} - k)'$$

$$2\sum_{i=1}^{6} I_{il}(M_{il} - (a_{i1}b_{1l} + a_{i2}b_{2l})) \cdot (-a_{im})$$

$$\frac{\partial E}{\partial b_{ml}} = 2\sum_{i=1}^{6} I_{il}(-M_{il}a_{im} + a_{i1}b_{im}b_{1l} + a_{i2}b_{im}b_{2l})$$

□

Opgave 4

a)

Program implementation

`netflix_gradient(M,A,B)` finder gradienten $\nabla f = \begin{pmatrix} f_a(a,b) \\ f_b(a,b) \end{pmatrix}$ med respekt til a og b og returnere tuplen $(f_a(a,b), f_b(a,b))$

`netflix(M)` udregner gradienten af en tilfældig A og B, hvorefter den påfører gradient descent på A og B. Derefter tjekker funktionen om gradienten er konvergeret, altså om størrelse af gradienten er mindre end den givne tolerance, hvorefter funktionen bestemmer om den er færdig eller ej (om den er konvergeret eller er kørt så længe at der ikke er noget resultat).

Programmets resultater

... er givet i programmet/**.ipynb** filen.

Fortolkning af M'

... forekommer i delopgave b).

b)

Vi skønner at én af fordele ved denne model kunne være at ved at kende de enkelte personers holdning/ratings af de givne fillmgenre er det muligt at beregne en "optimeret"matrix M', som ville indeholde de sandsynlige ratings, som de samme personer ville have givet til film af samme genre, hvis de havde set den. På den måde ville denne model gøre det muligt for at komme med nogen filmanbefalinger til den enkelte person.

Hvad ulemperne ved modellen angår, så er det svært at sige noget bestemt hvorvidt de konkrete tal vi får ud i matrix M' er meningsfulde. Som udgangspunkt, forventer vi at i den optimerede matrix M' indeholder tal i intervallet [1, 10] de steder, hvor i den oprindelige matrix M står "-".

Efter at have sammenlignet matricerne M og M' kan vi skønne at M' indeholder nogle sandsynlige ratings i intervallet [1, 10] som de enkelte personer ville have givet til de film de ikke havde set.