MASD Assignment 3

Group 56: wgh762, wqm216, hfc635

September 2021

Exercise 1

a)

først udregnes afledningen:

$$f'(x) = 4x + 3(x-4)^2 = 4x + 3(x^2 - 8x + 16) = 3x^2 - 20x + 48$$

Hvis denne afledte funktion sættes lig 0, så kan ligningen $\frac{-b \pm \sqrt{b^2-4ac}}{2a}$ bruges:

$$3x^2 - 20x + 48 = 0 \Rightarrow x = \frac{20 \pm \sqrt{(-20)^2 - 4 \cdot 3 \cdot 48}}{2 \cdot 3} = \frac{20 \pm \sqrt{400 - 576} - 6}{}$$

Hvilket giver to løsninger der er komplekse værdier. Siden funktionen er defineret til kun at være en funktion af reelle tal, så har denne funktion ikke nogen kritiske punkter.

b)

Den afledte funktion findes:

$$f'(x) = 2x \ln x + x^2 \cdot \frac{1}{2} = 2x \ln x + x = x(2 \ln x + 1)$$

Denne ligning sættes lig 0 for at finde de kritiske punkter:

$$f'(x) = 0 \Rightarrow x(2 \ln x + 1) = 0$$

Dette ville tyde på at enten $x = 0$ eller $2 \ln x + 1 = 0$. Siden funktionen ikke er defineret når $x = 0$, da $\ln 0$ ikke er defineret. Må det eneste kritiske punkt være når $2 \ln x + 1 = 0$, eller sagt på en anden måde: når $\ln x = -\frac{1}{2}$, hvilket giver $x = e^{-\frac{1}{2}}$. Man afleder den afledte funktion for at finde hvilken slags kritisk punkt det er:

$$f''(x) = x(\frac{2}{x}) + (2 \ln x + 1) = 3 + 2 \ln x$$

Med punktet der før blev fundet:

$$f''(e^{-\frac{1}{2}} = 3 + 2\ln e^{-\frac{1}{2}} = 3 + 2(-\frac{1}{2}) = 2 > 0$$

Derfor er det kritiske punkt $x = e^{-\frac{1}{2}}$ et minimum.

c)

Da det vides at et kritisk punkt for en funktion med to variabler er et punkt hvor $\nabla f(x_1, x_2) = 0$ Dette fører til følgende:

$$\nabla f(x_1, x_2) = \begin{pmatrix} 2\text{x}_1 + 2x_2 \\ 2\text{x}_1 + 6x_2 \end{pmatrix} = \begin{pmatrix} 2 & 2 \\ 2 & 6 \end{pmatrix} x$$

Det kan ses at udtrykket $\nabla f(x_1, x_2) = 0$ kun er sandt ved følgende:

$$2x_1 + 2x_2 = 0 \iff x_1 = -x_2 \; og$$
$$2x_1 + 6x_2 = 0 \iff x_1 = -3x_2$$

Det eneste tidspunkt dette er sandt, er når begge punkter er lig 0, derfor er $x = 0$ det eneste kritiske punkt. Den hessiske matrix findes for at afgøre hvilken slags kritisk punkt det er.

$$Hf(x_1, x_2) = \begin{pmatrix} 2 & 2 \\ 2 & 6 \end{pmatrix}$$

Eigenværdierne, fundet igennem Wolframalpha ved at søge "((2, 2), (2, 6)) eigenvalues" bliver til:

$$2(2 + sqrt(2)) = 6,83$$
$$-2(-2 + sqrt(2)) = 1,17$$

Siden begge værdier er positive, så er punktet $x = 0$ et minimum.

d)

Endnu engang findes $\nabla f(x_1, x_2)$:

$$\nabla f(x_1, x_2) = \begin{pmatrix} 2\text{x}_1 + 2x_2 \\ \text{-2x}_1 + 2x_2 \end{pmatrix} = \begin{pmatrix} 2 & \text{-2} \\ \text{-2} & 2 \end{pmatrix} x$$

Hvis dette sættes lig nul fås:

$$2x_1 - 2x_2 = 0 \iff x_1 = x_2 \; og$$
$$-2x_1 + 2x_2 = 0 \iff x_1 = x_2$$

På denne måde ses det at der er kritiske punkter for f langs hele linjen hvor $x_1 = x_2$. Den hessiske matrix findes:

$$Hf(x_1, x_2) = \begin{pmatrix} 2 & \text{-2} \\ \text{-2} & 2 \end{pmatrix}$$

som, ligesom forrige opgave (gennem søgning af wolframalpha med "((2, -2), (-2, 2)) eigenvalues"), finder eigenværdierne 0 og 4. Da en af eigenværdierne er 0, er andengradsudledningen ukonklusiv. Dog, kan man med lidt tænkning nå frem til følgende:
Ved ethvert punkt langs linjen $x_1 = x_2$, så har funktionen returværdien 0. Dog, siden at funktionsværdien er i andenpotens, så vides det at ethvert andet punkt end det kritiske, ville være positivt. Derfor er hvert punkt langs linjen $x_1 = x_2$ et minimum

Exercise 2

a

Kort sagt kan vi sige at den matematiske formel beskriver hvorledes sammenhængen mellem de forskellige brugeres smag indenfor film kan forudses. Mere udførligt kan vi sige at argmin har matricerne A og B, som når de bliver ganget giver (næsten) matricen M. Forskellen er at M vil indeholde flere tomme celler eller 0'er.

Grunden til at matricen M er opdelt i A og B matricer er for at overskueligegøre datasætte og gøre det nemmere at lave vurderinger på trods af de ukendte værdier. Ved at opdele matricen reduceres mængden af ukendte. Opdelingen er optaget ved at kigge på antallet af film og person og derefter er de blevet delt op i to matricer af henholdsvis 6(personer)x2 og 2x10(film). Vi ved ikke hvorfor at det netop er 2 de er delt med udover at resultatet af A x B vil give M. De resulterende værdier i A og B matricen vil give et indblik i hvorledes sammenhængen mellem personerne/film samt den ukendte faktor.

Formlen/modelen udregner længden på normaliseret matrice, som vi forventer vil være fyldt med meget små værdier. Matricerne **A** og **B** er defineret som være i stand til at *næsten* at give matricen **M**, når de ganges samme. Grunden til at dette er at den oprindelige M matrice indeholder tomme værdier/ 0'er for film som brugerne ikke har set hvorimod den "nye" M vil være i stand til at udfylde disse tomme felter. Disse værdier - hvor M har nul eller tomme felter mens A x B har værdier - er vores værdi som ligger til grund for vores forudsigelser for hvad der skulle stå i de tilsvarende rækker/koloner i M. Når værdierne ganges med I, bliver de negeret. Da AB og M næsten er ens vil resultatet af M - AB derfor være en matrice med ganske små værdier. Når den resulterende matricen normaliseres, så er det i bund og grund en summering af alle differenceren for at tjekke om de ligger indenfor en grænse vi vælger. Ved at summe dem alle op og derefter sætte dem i anden gør det nemmere for algoritmen at tjekke for denne gennemsnitlige fejl ratio.

Ved at gange en persons tilbøjeligheder til at kunne lide en bestemt genre med dennes score i den pågældende genre, fortsætte med dette for alle genre og derefter lægge dem sammen giver et guesstimat for hvor tilbøjelig en person er for at se en film, som de ikke kender til.

b

I denne opgave skal vi vise at følgende er sandt:

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

Vi løser denne opgave ved at dele udsagnet op i mindre stykker. Da vi ved at ved at gange to matricer, som er på formen m x n og n x m med hinanden kan skrives således:

$$C_{j1} = a_{i1}b_{j1} + a_{i2}b_{2j} + ... + a_{ni}b_{nj} \quad (2)$$

Ud fra dette kan vi se at to dele af formelen passer, nemlige $a_{i1}b_{j1} + a_{i2}b_{2j}$ og M_{ij}. Det sidste giver mening når C_{ij} i den ovenstående formele erstattes med M_{ij} og at man holder i mente at A x B er defineret som at være næsten lige med M.

Næste trin er ved at kigge på hvorledes en matrice ganges med en anden og kan også udledes ud fra formel 2. Nemlige at hvis I ganges elementvis på (M-AB) og M, A og B har række og kolonnerne i,j vil I også havde række og kolonerne i,j.

Vi har nu vist:

$$I_{ij}(M_{ij} - (a_{i1}b_{j1} + a_{i2}b_{2j})) \quad (3)$$

De sidste led vi mangler er de to summaertionstegn samt normaliseringen af denne ($\|$). Her kan vi sige at $\|V - U\|$ er et udtryk for afstanden mellem vektor V og U. Som også kan skrives ud som $\sqrt{(v_1 - u_2)^2 + ... + (v_n - u_n)^2}$ derfor er $||V - U||^2 = (v_1 - u_1) + ... + (v_n - u_n)$ Dette kunne også udtrykkes som $\sum_{i=1}^{n}(v_i - u_i)^2$. Det samme gør sig gældende for vores udtryk som blot arbejder i flere dimensioner (i og j).

Exercise 3

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

Dette kan vi omkskrive til følgende, da vi tager den afledte i forhold til a_{km} kan vi erstatte alle vores i med K og vores 1 og 2 med M. På den måde kan vi fjerne i sumationen: $E(A, B) = \sum_{j=1}^{10} I_{kj}(M_{kj} - (a_{k1}b_{1j} + a_{k2}b_{2j})^2$

Vi valgte for overskueligheden at dele denne op i forskellige underfunktioner:

$$\begin{aligned} v &= a_{k1}b_{1j} + a_{k2}b_{2j} \\ g &= M_{kj} - v \\ c &= g^2 \\ d &= I_{kj} * c \end{aligned}$$

$$\frac{\delta v}{\delta a_{km}} = 1 * b_{1j} + 1 * b_{2j} = b_{mj} \qquad \text{Sumregel}$$

$$\frac{\delta g}{\delta a_{km}} = 0 - \frac{\delta v}{\delta a_{km}} = \frac{\delta v}{\delta a_{km}} \qquad \text{Sumregel}$$

$$\begin{aligned} \frac{\delta c}{\delta a_{km}} &= 2(g)\frac{\delta g}{\delta a_{km}} \qquad \text{Kæderegel} \\ &= 2(M_{kj} - a_{k1}b_{1j} + a_{k2}b_{2j})(b_{mj}) \\ &= 2(M_{kj}b_{mj} - a_{k1}b_{1j}b_{mj} + a_{k2}b_{2j}b_{mj}) \end{aligned}$$

$$\begin{aligned} \frac{\delta d}{\delta a_{km}} &= 0c + I_{kj}\frac{\delta c}{\delta a_{km}} = I_{kj}\frac{\delta c}{\delta a_{km}} \qquad \text{Produktregel} \\ &= I_{kj}2(M_{kj}b_{mj} - a_{k1}b_{1j}b_{mj} + a_{k2}b_{2j}b_{mj}) \end{aligned}$$

dvs. at :

$$\begin{aligned} \frac{\delta E}{\delta a_{km}} &= \sum_{j=1}^{10} \frac{\delta d}{\delta a_{km}} \qquad \text{Sumregel} \\ &= \sum_{j=1}^{10} I_{kj}2(M_{kj}b_{mj} - a_{k1}b_{1j}b_{mj} + a_{k2}b_{2j}b_{mj}) \\ &= 2\sum_{j=1}^{10} I_{kj}(M_{kj}b_{mj} - a_{k1}b_{1j}b_{mj} + a_{k2}b_{2j}b_{mj}) \end{aligned}$$

Vi kan flytte 2 udenfor sumtegnet da den er konstant gennem alle iterationer.

Vi ønsker nu at vise at følgende er korrekt:

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

Vi benytter samme fremgang som før og dele udtrykket op i flere dele.

$$v = a_{i1}b_{1l} + a_{i2}b_{2l}$$
$$g = M_{il} - v$$
$$c = g^2$$
$$d = I_{il} * c$$

$$\frac{\delta v}{\delta b_{bl}} = 1 * a_{im} + 1 * a_{im} = a_{im} \qquad \text{Sumregel}$$

$$\frac{\delta g}{\delta b_{ml}} = 0 - \frac{\delta v}{\delta b_{ml}} = \frac{\delta v}{\delta b_{ml}} \qquad \text{Sumregel}$$

$$\frac{\delta c}{\delta b_{ml}} = 2(g)\frac{\delta g}{\delta b_{ml}} \qquad \text{Kæderegel}$$

$$= 2(M_{il} - a_{i1}b_{1l} + a_{i2}b_{2l})(a_{im})$$
$$= 2(M_{il}a_{im} - a_{i1}b_{1l}a_{im} + a_{i2}b_{2l}a_{im})$$

$$\frac{\delta d}{\delta b_{ml}} = 0c + I_{il}\frac{\delta c}{\delta b_{ml}} = I_{il}\frac{\delta c}{\delta b_{ml}} \qquad \text{Produktregel}$$

$$= I_{il}2(M_{il}a_{im} - a_{i1}b_{1l}a_{im} + a_{i2}b_{2l}a_{im})$$

dvs. at :

$$\frac{\delta E}{\delta b_{ml}} = \sum_{j=1}^{10} \frac{\delta d}{\delta b_{ml}} \qquad \text{Sumregel}$$

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

Vi kan flytte 2 udenfor sumtegnet da den er konstant gennem alle iterationer.

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

Derved har vi bevist det som vi ønskede nemlige at

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

Exercise 4

a)

Vi har blot implementeret vores gradienter ved at hardcode vores partielle afledte inde i nested for-loops hvor vi køre igennem range i, j og m. På den måde fylder vi så A og B i deres respektive elementer. Dette betyder at vi mange udregninger og at køretiden måske ikke er helt ideel:

```
for j in range(10) :
        for i in range (6) :
            for m in range(2) :
                grad_A[i][m]+=(2*(I[i][j]*(-M[i][j]*B[m][j]+A[i][0]*B[0][j]*B[m][j]+A[i][1]*B[1][j]*

for i in range(6) :
    for m in range(2) :
        for j in range(10) :
            grad_B[m][j]+=(2*(I[i][j]*(-M[i][j]*A[i][m]+(A[i][0]*A[i][m]*B[0][j]+A[i][1]*A[i][m]*B[1
```

vores A og B med learningrate på 0.0001

```
A :
 [[ 0.68077562 -3.45126028]
 [ 0.33806753 -3.71483297]
 [-0.3709511  -3.65481161]
 [ 3.11374529 -1.03741474]
 [ 3.49466423 -1.00942082]
 [ 3.23077017 -0.93285998]]
B :
 [[-0.332536   -0.40034714 -0.39084033 -0.28513409 -0.33167868  2.9713013
   2.38105348  2.61667987  1.87402078  2.25258975]
 [-2.42618641 -2.41096423 -2.70876364 -2.30650333 -2.71836364 -0.01100319
  -0.63517278 -0.4773239  -0.47893001 -1.40728801]]
```

Vores M' (Round AB) endte med at se sådan ud med en learningrate på 0.0001:

```
[[ 8.   8.   9.   8.   9.   2.   4.   3.   3.   6.]
 [ 9.   9.  10.   8.  10.   1.   3.   3.   2.   6.]
 [ 9.   9.  10.   9.  10.  -1.   1.   1.   1.   4.]
 [ 1.   1.   2.   2.   2.   9.   8.   9.   6.   8.]
 [ 1.   1.   1.   1.   2.  10.   9.  10.   7.   9.]
 [ 1.   1.   1.   1.   1.  10.   8.   9.   7.   9.]]
```

Som man kan se har den enkelte steder hvor den har den få forskelle fra M udover hvor $M_{ij} = 0$. Derudover har vi et enkelt punkt som er minus. Ser vi på et matrix som viser forskellen på M og M' i alle værdier som ikke er 0 i M. Ser vi at en enkelt værdi har en forskel på 3:

```
[[-1.  0.  0.  0.  0. -1.  0. -1.  0.  3.]
 [ 0.  0.  0.  1.  0.  1.  0.  0.  0. -1.]
 [ 1.  0.  0. -1.  0.  0.  0.  1.  0. -1.]
 [ 0.  0.  0.  0.  0.  0.  0.  0.  0.  0.]
 [ 0.  0.  0.  0.  0.  0.  0.  0.  0.  0.]
 [ 1.  0.  0.  0.  0.  0.  0.  0.  0. -1.]]
```

Her ser vi altså at sidste element i første række har en forskel på 3 og må altså tolkes som at være en outlier.

b)

Generelt kan vi sige om denne implementering at vi den har en ok foreudsigelses rate, men det kræver finjustering af learningrate, tolerance osv. For at skabe bedre resultater kunne man forstille sig at film havde mere metadata end blot titel og rating som kunne bruges til at kvalificere forudsigelsen af personers filmsmag yderligere.