Kompendium til

UNF Machine Learning Camp 2024

Christian Noble Truelsen
David Rasmussen Lolck
Mikkel Riis Rasmussen
Mohamad Dalal
Oskar Ude Jørgensen
Rasmus Frigaard Lemvig
Rebekka Wätzold Høgh Madsen

Navn:

Ungdommens Naturvidenskabelige Forening

Med tidligere materiale og hjælp fra:
Andreas Mosbæk Jensen
Andreas Raaskov Madsen
Emil Fulei Lykke Aagreen
Lotte Maria Bruun
Janne Brøgger
Jonas Ryssel
Jesper Samsø Birch
Mads Rogild
Mathias Weirsøe Klitgaard
Karl Hvarregaard Mose
Nikolaj Simling
Nicklas Vedsted

Kompendium til UNF Machine Learning Camp 2024

Kompendiet er skrevet af Andreas Raaskov Madsen, Christian Noble Truelsen, David Rasmussen Lolck, Lauge Hermansen, Mikkel Riis Rasmussen, Mohamad Dalal, Oskar Ude Jørgensen, Rasmus Frigaard Lemvig og Rebekka Wätzold Høgh Madsen. Kompendiet er trykt i juli 2024, og teksten er

Korrektur: Ria Dole, Ane Sophie Mahler Sonne, Karl Meisner-Jensen, og fagligt team. Design: Sofie Kjærgaard.

Indhold

Kapitel 1

Introduktion til matematik i ML

Rasmus Frigaard Lemvig

1 Kort om forløbet

I dette forløb introducerer vi den matematik, der er nødvendig for at forstå, hvordan machine learning grundlæggende virker. Vi starter med lineær algebra, hvor vi har fokus på at regne med matricer, som man kan tænke på som skemaer med tal. Derefter vil vi introducere differentialregning, som flere af jer nok allerede er bekendte med. Vi kommer ikke til at gå i dybden med det, og fokus kommer primært til at være på konceptuel forståelse. Kort sagt beskriver differentialregning, hvordan funktioner ændres, som inputtet ændres, og det er hjørnestenen i at bestemme minimum og maksimum for funktioner, hvilket indgår mange steder i machine learning. Vi skal herefter arbejde med sandsynlighedsteori, der danner grundlaget for de modeller, man bruger i statistik og machine learning. Til slut skal vi kort diskutere regression, hvor vi udelukkende fokuserer på lineær regression, som er den simpleste model til at beskrive tendenser i data.

2 Indledende matematik

Funktioner

En funktion plejer at være en ligning der tager en variabel og giver en ny værdi, alt efter hvad funktionen er beregnet til at gøre.
Et eksempel er

$$f(x) = x^2,$$

hvor funktionen hedder f, variablet som funktionen tager hedder x og giver x^2. Hvis man sætter variablet til at være lig 2, så at $x = 2$, vil man få at $f(x = 2) = 2^2 = 4$.
Funktioner kan også have flere variabler. Et eksempel for en funktion med flere variabler er funktionen,

$$f(x, y) = x + 2 \cdot y,$$

som tager både x og y som variabler. Hvis man eksempelvis sætter $x = 3$ og $y = 5$, får man $f(x = 3, y = 5) = 3 + 2 \cdot 5 = 3 + 10 = 13$.
I andre dele af undervisningen vil funktioner bliver brugt, hvor dette er ment som at give en kort basal beskrivelse af funktioner der vil kunne arbejdes videre med senere.

Polynomier

Polynomier er et funktion der kan bestå af konstanter og variabler. En variabel er et symbol som kan repræsentere forskellige værdier der kan varierer ud fra det man nu arbejder med, som giver dem navnet variabel.
Her er et eksempel på et førstegrads-, andengrads-, og trejdegradspolynomium:

$$kx^1, kx^2, kx^3,$$

hvor de forskellige grad siger noget om potensen på variablet x, og k er en tilfældig konstant der ganges på.
En n grad polynomier kan skrives som følge

$$p(x) = k_n\, x^n + k_{n-1}\, x^{n-1} + \cdots + k_2\, x^2 + k_1\, x + k_0$$

Potensregneregler

Når man arbejder med tal, eller variabler, som skal ganges eller divideres med hinanden, er der regler for hvordan det gøres for ikke at ende ud i svar med forkerte resultater. Der vil her gå igennem regler man typisk vil støde ind i.
Når man har en variabel x som man ganger med sig selv n gange, vil man skrive det: x^n, og udtales x opløftet i n'te, x opløftet med n, eller bare x i n'te. Det skrives på den måde i stedet for at skrive x, n gange. Det betyder så, at

$$x^n = x \cdot x \cdot x \cdot \cdots \cdot x.$$

Det man opløfter med hedder potensen, som også kaldes exponenten, hvor her n vil være potensen der opløftes.

Til de følgende regler vil vi bruge variablerne x og y, og konstanterne a og b:

$$\begin{aligned}
x^a \cdot x^b &= x^{a+b} \\
\frac{x^a}{x^b} &= x^{a-b} \\
(x^a)^b &= x^{a \cdot b} \\
(x \cdot y)^a &= x^a \cdot y^a \\
\left(\frac{x}{y}\right)^a &= \frac{x^a}{y^a}
\end{aligned}$$

Ud fra de her regler, kan man få disse definitioner:

$$\begin{aligned}
x^0 &= 1 \\
x^{-a} &= \frac{1}{x^a} \\
x^{1/a} &= \sqrt[a]{x} \\
x^{1/2} &= \sqrt{x}
\end{aligned}$$

Opgaver

Opgave 2.1. Tekst eksempel.

Den naturlige logaritme og Eulers tal

Når man har et tal, som vi betegner med variablet x, kan man tage logaritmen af det tal, hvilket skrives som $\log(x)$. Disse logaritmer vil altid have en base, som skrives som et lille tal efter log, f.eks. $\log_{10}(x)$, hvor basen her vil være 10.

Logaritmen kan udregnes ved at spørge sig selv, "hvor mange gange skal man gange basen med sig selv, for at få x?". Et simpelt eksempel er at udregne $\log_{10}(100)$. Hvor mange gange skal man gange 10 med sig selv, for at få 100?, $10^? = 100$. Vi ved at $10 \cdot 10 = 100 = 10^2$, så 2 gange, altså er $\log_{10}(100) = 2$. Vi kunne også prøve at finde $\log_{10}(0.001)$, hvilket vil give $\log_{10}(0.001) = -3$, da $10^{-3} = 0.001$

Basen kan være et hvilket som helst tal. Vi kunne f.eks. prøve at finde $\log_3(9)$, altså, hvor mange gange skal man gange 3 med sig selv, for at få 9? Siden $3 \cdot 3 = 9$, er $\log_3(9) = 2$.

Når man arbejder med tal der ikke går lige så nemt op med hinanden som disse eksempler, anbefaler vi at bruge en lommeregner.

Alt efter hvilken gren af naturvidenskab eller af samfundet man kommer fra, så er der forskellige definitioner af hvad det betyder når man skriver log uden en base. Nogen mener at når man skriver log uden en base betyder det $\log_{10}$, hvor nogen mener $\log_e$, som er den naturlige logaritme, som også skrives ln. Her, vil vi fortsætte med at når man skrive log uden en base, betyder det at vi gerne vil finde den naturlige logaritme, ln.

Bogstavet e i den naturlige logaritme er Eulers tal, og er en konstant der er defineret til at være en konstant med uendelige decimaler, hvor nogle af de første er 2.7182818... . Eulers tals er basen af en potens ligesom andre potenser som beskrevet tidligere. Det her tal dukker op flere steder, og har derfor også fået navnet *exponential funktionen*, og skrives ofte som $\exp(x)$, eller som e^x.

Det er er vigtigt at pointere at $\ln(x)$ og e^x er definerede til at være hinandens inverse. Det vil ikke sige, at $\log_e(x) = \ln(x) = x$, men at $\log_e(\mathrm{e}^x) = \ln(\mathrm{e}^x) = x$. Det gælder også den modsatte vej, hvor $\mathrm{e}^{\ln(x)} = x$. Et eksempel er at se $\ln(\mathrm{e}^9) = 9$, hvor $\ln(9) = 2.1972$.

Når man arbejder med logaritmer, er der også regneregler inden for dette, som vi vil forklare lidt om her.

$$\begin{aligned}
\log(x \cdot y) &= \log(x) + \log(y), \\
\log\left(\frac{x}{y}\right) &= \log(x) - \log(y), \\
\log\left(x^y\right) &= y \cdot \log(x), \\
\log\left(\mathrm{e}^x\right) &= x.
\end{aligned}$$

Opgaver

Opgave 2.2. Tekst eksempel.

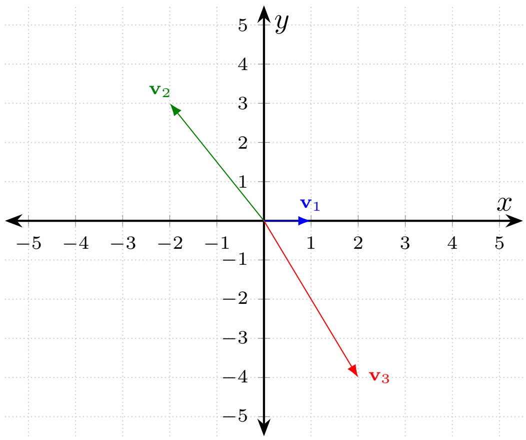
3 Lineær algebra

Regneoperationer af vektorer og matricer

Vi start med eksempler på vektorer i 2D-plane

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} -2 \\ 3 \end{pmatrix} \quad \text{og} \quad \mathbf{v}_3 = \begin{pmatrix} 2 \\ -4 \end{pmatrix}.$$

Disse vektorer kan nemt illustreres. Den øverste koordinat indikerer, hvor langt ud af x-aksen, man skal gå, mens andenkoordinaten indikerer, hvor langt ud af y-aksen, man skal gå. De tre vektorer ser sådan ud i planet:



Princippet med vektorer i 3D-rum er det samme. I er velkomne til at tænke på en vektor

$$\mathbf{v} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

i 3D som et punkt med koordinaterne (x, y, z). Og til en mere general tænkegang af vektor i n-dimension vi giver den følger definition.

Definition 3.1. En n-dimensional (reel) *vektor* er en ordnet liste

$$\mathbf{v} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

hvor $x_1, x_2, ..., x_n \in \mathbb{R}$ kaldes vektorens *indgange*. Vi betegner mængden af n-dimensionale vektorer med $\mathbb{R}^n$. Ønsker vi at angive dimensionen af en vektor, skriver vi $\dim \mathbf{v}$ for dimensionen af $\mathbf{v}$.

Selvom en vektor er defineret som en søjle, vil vi ofte bedrive misbrug af notation og skrive $\mathbf{v} = (x_1, x_2, ..., x_n)$ for en vektor. Vi indfører nu to fundamentale regneoperatorer for vektorer, nemlig addition og skalarmultiplikation. hvor vi giver ved eksemple

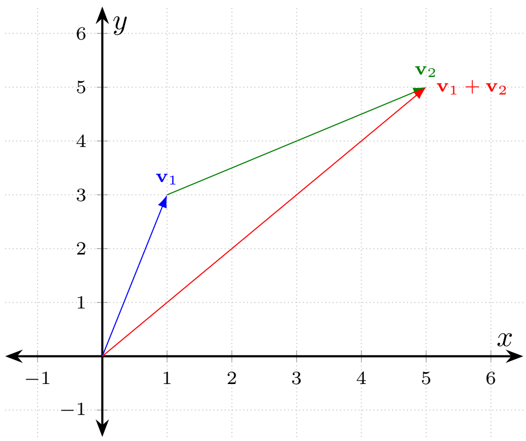
Både addition af vektorer og skalarmultiplikation har en geometrisk fortolkning. Summen af to vektorer giver vektoren, som fås ved at lægge de to vektorer i forlængelse af hinanden. Lad os se på vektorerne

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 3 \end{pmatrix} \quad \text{og} \quad \mathbf{v}_2 = \begin{pmatrix} 4 \\ 2 \end{pmatrix}.$$

Deres sum er

$$\begin{pmatrix} 1 \\ 3 \end{pmatrix} + \begin{pmatrix} 4 \\ 2 \end{pmatrix} = \begin{pmatrix} 5 \\ 5 \end{pmatrix}$$

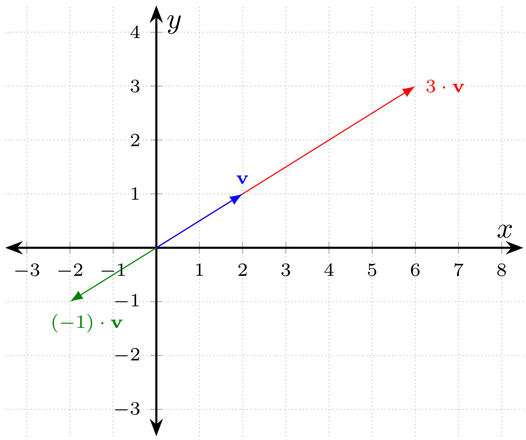
og i et koordinatsystem ser det hele således ud:



Skalarmultiplikation skal fortolkes som en skalering. Betragt vektoren

$$\mathbf{v} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

Skalerer vi vektoren med hhv. -1 og 3 fås



Skalarmultiplikation med -1 er et vigtig specialtilfælde, da det svarer til at spejle vektoren. Den skalerede vektor har samme længde, men peger i den modsatte retning.

Ved den addition og skalarmultiplikation eksempler vi giver den formale definition til de to koncepter

Definition 3.2. Lad $\mathbf{v}_1 = (x_1, x_2, ..., x_n) \in \mathbb{R}^n$, $\mathbf{v}_2 = (y_1, y_2, ..., y_n) \in \mathbb{R}^n$ og $a \in \mathbb{R}$. Summen af $\mathbf{v}_1$ og $\mathbf{v}_2$ er defineret som

$$\mathbf{v}_1 + \mathbf{v}_2 = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix}$$

og skalarmultiplikationen af a på $\mathbf{v}_1$ er defineret som

$$a \cdot \mathbf{v}_1 = a \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a \cdot x_1 \\ a \cdot x_2 \\ \vdots \\ a \cdot x_n \end{pmatrix}$$

hvor $\cdot$ inde i vektoren er almindelig multiplikation af reelle tal.

Definitionen ovenover forudsætter, at de to vektorer $\mathbf{v}_1$ og $\mathbf{v}_2$ har samme dimension. Vi har ingen definition af addition for vektorer med forskelligt antal indgange. F.eks. giver det ikke mening at lægge vektorerne

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ -2 \\ 8 \end{pmatrix} \quad \text{og} \quad \mathbf{v}_2 = \begin{pmatrix} 4 \\ 6 \end{pmatrix}$$

sammen, idet $\dim \mathbf{v}_1 = 3 \neq 2 = \dim \mathbf{v}_2$.

Definition 3.3. En $m \times n$-*matrix* $\mathbf{A}$ er et skema af reelle tal i m *rækker* og n *søjler*

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix},$$

hvor $a_{ij} \in \mathbb{R}$ betegner det reelle tal i række i og søjle j. Disse kaldes matricens *indgange*. Mængden af $m \times n$-matricer betegnes $\mathbb{R}^{m \times n}$.

Konvention: Vi benytter fed skrift til at betegne vektorer og matricer. Matricer betegnes med store bogstaver og indgangene med små bogstaver.

Eksempler på matricer i $\mathbb{R}^{2\times 3}$ er

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 0 & 1 & -4 \\ 11 & -3 & -1 \end{pmatrix},$$

mens matricer i $\mathbb{R}^{3\times 3}$ kunne være

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{og} \quad \mathbf{D} = \begin{pmatrix} -3 & 7 & 10 \\ 1/2 & 0 & 0 \\ 6 & -2 & -2 \end{pmatrix}.$$

I mange henseender er $n \times n$-matricer (altså matricer med samme antal rækker og søjler) særligt interessante. De kaldes *kvadratiske* matricer. Det skal vi se senere, når vi arbejder med inverse matricer. Matricen $\mathbf{C}$ ovenover er i øvrigt ikke helt tilfældig, men spiller derimod en meget central rolle. Faktisk er den så vigtig, at den har en særlig betegnelse. I definitionen nedenunder definerer vi en række centrale matricer og vektorer.

Definition 3.4. $\mathbf{0} \in \mathbb{R}^n$ defineres til at være vektoren, hvori alle indgange er 0:

$$\mathbf{0} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Denne vektor betegnes *nulvektoren.* $\mathbf{O} \in \mathbb{R}^{n \times n}$ defineres tilsvarende til at være den kvadratiske matrix med n rækker og søjler, hvori alle indgange er 0:

$$\mathbf{O} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}.$$

$\mathbf{I}_n \in \mathbb{R}^{n \times n}$ defineres til at være den kvadratiske matrix, hvor diagonalen består af 1-taller, og alle andre indgange er 0:

$$\mathbf{I}_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

$\mathbf{I}_n$ kaldes *identitetsmatricen* af dimension n.

Matricer er nyttige, fordi de indeholder information på en form, vi kan regne med. Vi skal dog først have defineret, hvordan vi kan regne med matricer. Tilfældet med addition minder meget om addition af vektorer, men når vi indfører multiplikation af matricer, bliver det straks mere regnetungt. Lad os starte med at få addition og skalarmultiplikation af matricer på plads.

Addition af to matricer er kun defineret, hvis de har samme dimensioner, dvs. samme antal rækker og samme antal søjler. Det giver altså kun mening at lægge matricerne $\mathbf{A}$ og $\mathbf{B}$ sammen, hvis de begge ligger i $\mathbb{R}^{m\times n}$.

Som eksempel kunne vi tage matricerne

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \quad \text{og} \quad \mathbf{B} = \begin{pmatrix} 0 & 1 & -4 \\ 11 & -3 & -1 \end{pmatrix}$$

fra tidligere. Vi har

$$\mathbf{A}+\mathbf{B} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} + \begin{pmatrix} 0 & 1 & -4 \\ 11 & -3 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 3 & -1 \\ 15 & 2 & 5 \end{pmatrix}.$$

Eksample af skalarmultiplikation af matricer er givet den matrice $\mathbf{A}$ fra eksample af addition af to matricer og en reele tal $a = 2$

$$a\mathbf{A} = 2\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} = \begin{pmatrix} 2\cdot 1 & 2\cdot 2 & 2\cdot 3 \\ 2\cdot 4 & 2\cdot 5 & 2\cdot 6 \end{pmatrix} = \begin{pmatrix} 2 & 4 & 6 \\ 8 & 10 & 12 \end{pmatrix}.$$

vi giver nu den formale definition til addition af to matricer og skalarmultiplikation af matricer

Definition 3.5. Lad $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times n}$. Skriv

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \quad \text{og} \quad \mathbf{B} = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix},$$

da er summen $\mathbf{A}+\mathbf{B}$ defineret ved

$$\mathbf{A}+\mathbf{B} = \begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \cdots & a_{2n}+b_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \cdots & a_{mn}+b_{mn} \end{pmatrix}.$$

Altså skal man blot lægge indgangene sammen. Hvis $a \in \mathbb{R}$ er skalarmultiplikationen $a\mathbf{A}$ defineret som

$$a\mathbf{A} = \begin{pmatrix} a \cdot a_{11} & a \cdot a_{12} & \cdots & a \cdot a_{1n} \\ a \cdot a_{21} & a \cdot a_{22} & \cdots & a \cdot a_{2n} \\ \vdots & \vdots & & \vdots \\ a \cdot a_{m1} & a \cdot a_{m2} & \cdots & a \cdot a_{mn} \end{pmatrix}$$

altså skal man blot gange a på alle indgangene i $\mathbf{A}$.

Vi bevæger os nu raskt videre til multiplikation med matricer. Vi starter med en eksample af specialtilfælde, nemlig multiplikation af en matrix og en vektor.

Vi farvelægger indgangene i vektoren, så man kan se, hvad der foregår i multiplikationen:

$$\begin{pmatrix} 1 & 2 \\ -1 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \cdot 1 + 2 \cdot 3 \\ -1 \cdot 1 + 4 \cdot 3 \end{pmatrix} = \begin{pmatrix} 7 \\ 11 \end{pmatrix}$$

Lad os tage et eksempel med en større matrix og vektor:

$$\begin{aligned} \begin{pmatrix} 1 & 0 & 0 & -1 \\ 3 & -2 & 4 & 1 \\ 10 & 0 & -1 & 4 \\ 8 & 7 & -1 & 0 \\ 7 & -9 & -2 & 11 \end{pmatrix} \begin{pmatrix} 2 \\ 0 \\ 5 \\ 6 \end{pmatrix} &= \begin{pmatrix} 1 \cdot 2 + 0 \cdot 0 + 0 \cdot 5 - 1 \cdot 6 \\ 3 \cdot 2 - 2 \cdot 0 + 4 \cdot 5 + 1 \cdot 6 \\ 10 \cdot 2 + 0 \cdot 0 - 1 \cdot 5 + 4 \cdot 6 \\ 8 \cdot 2 + 7 \cdot 0 - 1 \cdot 5 + 0 \cdot 6 \\ 7 \cdot 2 - 9 \cdot 0 - 2 \cdot 5 + 11 \cdot 6 \end{pmatrix} \\ &= \begin{pmatrix} -4 \\ 32 \\ 39 \\ 11 \\ 70 \end{pmatrix} \end{aligned}$$

følge med den formal definition af multiplikation af en matrix og en vektor.

Definition 3.6. Lad $\mathbf{A} \in \mathbb{R}^{m\times n}$ og $\mathbf{v} \in \mathbb{R}^n$ med indgange a_{ij} og x_j, hvor $i = 1, ..., m$ og $j = 1, ..., n$. Da definerer vi produktet af $\mathbf{A}$ og $\mathbf{v}$ til

$$\mathbf{A}\mathbf{v} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n \end{pmatrix} \in \mathbb{R}^m.$$

Matrix-vektorproduktet $\mathbf{A}\mathbf{v}$ er kun defineret, hvis $\mathbf{A}$ har samme antal søjler som indgange i $\mathbf{v}$. Resultatet af at gange en $m \times n$-matrix på en n-dimensional vektor er en m-dimensional vektor.

Kender man til skalarproduktet/prikproduktet af to vektorer, findes en god måde at huske, hvordan man beregner matrix-vektorprodukter. Definitionen af prikproduktet genkalder vi herunder.

Definition 3.7. Lad $\mathbf{v}_1 = (x_1, x_2, ..., x_n)$ og $\mathbf{v}_2 = (y_1, y_2, ..., y_n)$ være vektorer i $\mathbb{R}^n$. Da er *prikproduktet/skalarproduktet* af $\mathbf{v}_1$ og $\mathbf{v}_2$ defineret til

$$\mathbf{v}_1 \cdot \mathbf{v}_2 = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = x_1y_1 + x_2y_2 + ... + x_ny_n \in \mathbb{R}.$$

Vi bemærker, at prikproduktet af to vektorer kun er defineret, hvis de to vektorer har samme antal indgange, og at prikproduktet giver et reelt tal og ikke en vektor. Nu kan vi forklare sammenhængen mellem prikproduktet og matrix-vektormultiplikation. Antag, at vi har en $m \times n$-matrix $\mathbf{A}$ og en vektor $\mathbf{v}$ i $\mathbb{R}^n$. Lad $\mathbf{A}_1, \mathbf{A}_2, ..., \mathbf{A}_m$ betegne rækkerne i $\mathbf{A}$. Da er $\mathbf{A}_1, \mathbf{A}_2, ..., \mathbf{A}_m$ vektorer i $\mathbb{R}^n$, og vi har

$$\mathbf{A}\mathbf{v} = \begin{pmatrix} \mathbf{A}_1 \cdot \mathbf{v} \\ \mathbf{A}_2 \cdot \mathbf{v} \\ \vdots \\ \mathbf{A}_m \cdot \mathbf{v} \end{pmatrix}.$$

Lad os nu definere matrix-matrixmultiplikation.

Definition 3.8. Lad $\mathbf{A} \in \mathbb{R}^{m\times n}$ og $\mathbf{B} \in \mathbb{R}^{n\times p}$. Lad $\mathbf{B}_1, \mathbf{B}_2, ..., \mathbf{B}_p$ betegne søjlerne i $\mathbf{B}$. Da definerer vi produktet $\mathbf{AB}$ til at være $m \times p$-matricen givet ved

$$\mathbf{AB} = \begin{pmatrix} \mathbf{AB}_1 & \mathbf{AB}_2 & \cdots & \mathbf{AB}_p \end{pmatrix}.$$

I ord: $\mathbf{AB}$ er matricen, hvis søjler er produktet af $\mathbf{A}$ med søjlerne i $\mathbf{B}$.

Bemærk, at matrixproduktet $\mathbf{AB}$ er defineret hvis og kun hvis $\mathbf{A}$ har samme antal søjler, som $\mathbf{B}$ har rækker. Lad os tage en række eksempler. Betragt de to matricer

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \quad \text{og} \quad \mathbf{B} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \\ 3 & -1 \end{pmatrix}.$$

Da $\mathbf{A}$ har tre søjler, og $\mathbf{B}$ har tre rækker, er $\mathbf{AB}$ veldefineret. Vi beregner

$$\begin{aligned} \mathbf{AB} &= \begin{pmatrix} \mathbf{A}\begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix} & \mathbf{A}\begin{pmatrix} 0 \\ 2 \\ -1 \end{pmatrix} \end{pmatrix} \\ &= \begin{pmatrix} 1\cdot 1 + 2\cdot 0 + 3\cdot 3 & 1\cdot 0 + 2\cdot 2 + 3\cdot(-1) \\ 4\cdot 1 + 5\cdot 0 + 6\cdot 3 & 4\cdot 0 + 5\cdot 2 + 6\cdot(-1) \end{pmatrix} \\ &= \begin{pmatrix} 10 & 1 \\ 22 & 4 \end{pmatrix}. \end{aligned}$$

Vi kan også beregne produktet $\mathbf{BA}$ fordi antal søjler i $\mathbf{B}$ er lig antal rækker i $\mathbf{A}$. Da fås

$$\mathbf{BA} = \begin{pmatrix} 1 & 2 & 3 \\ 8 & 10 & 12 \\ -1 & 1 & 3 \end{pmatrix}$$

I dette eksempel kunne man tage matrixproduktet fra begge sider, men dette er ikke altid muligt. Vi kan f.eks. se på matricen

$$\mathbf{C} = \begin{pmatrix} 1 & 0 \\ 4 & 2 \end{pmatrix}.$$

Da eksisterer matrixproduktet $\mathbf{CA}$, men $\mathbf{AC}$ giver ikke mening. Et simpelt og vigtigt eksempel på matrix-produkter er mellem kvadratiske matricer. Her kan man altid tage produktet fra begge sider, men det er ikke nødvendigvis det samme resultat, man får. Definér

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ -1 & 7 \end{pmatrix} \quad \text{og} \quad \mathbf{B} = \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix}.$$

Da udregner vi

$$\mathbf{AB} = \begin{pmatrix} 4 & 2 \\ 5 & 7 \end{pmatrix} \quad \text{men} \quad \mathbf{BA} = \begin{pmatrix} 2 & 4 \\ 0 & 9 \end{pmatrix}$$

så $\mathbf{AB} \neq \mathbf{BA}$. Dette står i kontrast til, hvad vi kender fra tallene normalt. Hvis $a, b \in \mathbb{R}$ gælder $ab = ba$.

Invertering og transponering af matricer

Vi vil nu se på invertering af matricer. Genkald identitetsmatricen

$$\mathbf{I}_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

Denne matrix har den interessante egenskab, at $\mathbf{I}_n\mathbf{v} = \mathbf{v}$ for alle $\mathbf{v} \in \mathbb{R}^n$. Det følger dermed, at $\mathbf{AI}_n = \mathbf{A} = \mathbf{I}_n\mathbf{A}$ for alle $n \times n$ matricer $\mathbf{A}$. Dermed fungerer $\mathbf{I}_n$ på samme måde som tallet 1 i de reelle tal. Der har vi nemlig $1 \cdot a = a = a \cdot 1$ for alle $a \in \mathbb{R}$. Hvis $a \neq 0$, kan vi bestemme en invers til a, som vi kan skrive som a^{-1} eller $1/a$. Vi skal nu gøre noget lignende for matricer. Det viser sig dog, at ikke alle kvadratiske matricer forskellig fra nulmatricen er invertible. Heldigvis findes der metoder til at tjekke det, i hvert fald for små matricer.

Definition 3.9. En matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ (bemærk, at $\mathbf{A}$ er kvadratisk) er *invertibel*, hvis der findes en matrix $\mathbf{A}^{-1} \in \mathbb{R}^{n \times n}$ som opfylder

$$\mathbf{AA}^{-1} = \mathbf{I}_n = \mathbf{A}^{-1}\mathbf{A}.$$

Lad os se på et eksempel. Betragt matricen

$$\mathbf{A} = \begin{pmatrix} 2 & 3 \\ 1 & -4 \end{pmatrix}.$$

Vi hævder, at $\mathbf{A}$ er invertibel med invers

$$\mathbf{A}^{-1} = \frac{1}{11}\begin{pmatrix} 4 & 3 \\ 1 & -2 \end{pmatrix}.$$

Lad os tjekke, at dette er tilfældet. Vi udregner

$$\begin{aligned}
\mathbf{A}\mathbf{A}^{-1} &= \begin{pmatrix} 2 & 3 \\ 1 & -4 \end{pmatrix}\frac{1}{11}\begin{pmatrix} 4 & 3 \\ 1 & -2 \end{pmatrix} = \frac{1}{11}\begin{pmatrix} 2 & 3 \\ 1 & -4 \end{pmatrix}\begin{pmatrix} 4 & 3 \\ 1 & -2 \end{pmatrix} \\
&= \frac{1}{11}\begin{pmatrix} 4 \cdot 2 + 3 \cdot 1 & 3 \cdot 2 + 3 \cdot (-2) \\ 4 \cdot 1 + 1 \cdot (-4) & 3 \cdot 1 + (-4) \cdot (-2) \end{pmatrix} \\
&= \frac{1}{11}\begin{pmatrix} 11 & 0 \\ 0 & 11 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.
\end{aligned}$$

Tilsvarende kan vi vise ved udregning, at

$$\mathbf{A}^{-1}\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Altså er $\mathbf{A}^{-1}$ den inverse til $\mathbf{A}$. Dette eksempel virker sikkert som snyd. Hvor kommer $\mathbf{A}^{-1}$ fra? Generelt er det svært at invertere matricer, men i det simple tilfælde med en 2×2 matrix er der en løsningsformel.

Sætning 3.10. Betragt en matrix $\mathbf{A} \in \mathbb{R}^{2\times 2}$ givet ved

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Da gælder, at hvis $ad - bc \neq 0$, så er $\mathbf{A}$ invertibel med invers

$$\mathbf{A}^{-1} = \frac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Bevis. Beviset overlades til jer. Se opgaverne. □

Prøv som øvelse at bruge sætning 3.10 på matricen $\mathbf{A}$ fra før. Da vil I også få $\mathbf{A}^{-1}$, som vi før blot hev op af hatten. Størrelsen $ad-bc$ er så vigtig, at den har sit eget navn.

Definition 3.11. Lad $\mathbf{A} \in \mathbb{R}^{2\times 2}$ være givet ved

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Da defineres *determinanten* af $\mathbf{A}$ ved

$$\det \mathbf{A} = ad - bc.$$

Vi ønsker at bemærke, at man sagtens kan definere determinanten for kvadratiske matricer af enhver dimension. Dog bliver problemet med at regne determinanten markant sværere, og vi ønsker derfor at holde os til 2×2-tilfældet. Determinanten har et væld af interessante og brugbare egenskaber. Vi har allerede set en af dem, nemlig, at hvis $\det \mathbf{A} \neq 0$, da er $\mathbf{A}$ invertibel. En anden interessant egenskab er såkaldt *multiplikativitet.*

Definition 3.12. Lad $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{2\times 2}$. Da gælder, at

$$\det(\mathbf{AB}) = \det \mathbf{A} \cdot \det \mathbf{B} = \det \mathbf{B} \cdot \det \mathbf{A}.$$

de Som det sidste i lineær algebra-delen vil vi se på transponering af matricer. Her ser vi ikke kun på kvadratiske matricer.

Definition 3.13. Lad $\mathbf{A} \in \mathbb{R}^{m\times n}$. Den transponerede matrix $\mathbf{A}^T \in \mathbb{R}^{n\times m}$ er matricen, hvis i'te søjle er lig den i'te række i $\mathbf{A}$.

Lidt løst kan man sige, at $\mathbf{A}^T$ fås ved at bytte rundt på rækker og søjler i $\mathbf{A}$. Lad os tage et eksempel. Se på

$$\mathbf{A} = \begin{pmatrix} 1 & -3 & 2 \\ 0 & 1 & -4 \end{pmatrix}.$$

Da vil

$$\mathbf{A}^T = \begin{pmatrix} 1 & 0 \\ -3 & 1 \\ 2 & -4 \end{pmatrix}.$$

Bemærk, at matrixprodukterne $\mathbf{A}\mathbf{A}^T$ og $\mathbf{A}^T\mathbf{A}$ altid findes. $\mathbf{A}\mathbf{A}^T \in \mathbb{R}^{m\times m}$, og $\mathbf{A}^T\mathbf{A} \in \mathbb{R}^{n\times n}$, så $\mathbf{A}\mathbf{A}^T$ og $\mathbf{A}^T\mathbf{A}$ er ovenikøbet kvadratiske. Dog har de ikke samme dimension, medmindre $\mathbf{A}$ er kvadratisk (dvs. $m = n$). Vi har faktisk allerede indirekte set transponering. Bemærk, at prikproduktet af to vektorer $\mathbf{v}_1$ og $\mathbf{v}_2$ i $\mathbb{R}^n$ kan skrives som

$$\mathbf{v}_1 \cdot \mathbf{v}_2 = \mathbf{v}_1^T\mathbf{v}_2,$$

hvor højresiden er et matrixprodukt. Vi har nogle regneregler for transponering, der er nyttige at have i baghovedet.

Sætning 3.14. Vi har følgende egenskaber/regneregler for transponering

1. For enhver matrix $\mathbf{A}$ gælder $(\mathbf{A}^T)^T = \mathbf{A}$.
2. For matricer $\mathbf{A}$ og $\mathbf{B}$ af samme dimensioner gælder $(\mathbf{A}+\mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$.
3. Lad $\mathbf{A} \in \mathbb{R}^{m\times n}$ og $\mathbf{B} \in \mathbb{R}^{n\times p}$. Da er $(\mathbf{A}\mathbf{B})^T = \mathbf{B}^T\mathbf{A}^T$ (bemærk, at rækkefølgen vender).
4. $\det \mathbf{A}^T = \det \mathbf{A}$ for enhver kvadratisk matrix $\mathbf{A}$.
5. $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$.

For en matrix $\mathbf{A}$, se igen på den kvadratiske matrix $\mathbf{A}\mathbf{A}^T$. Ved at bruge regnereglerne ovenover har vi

$$(\mathbf{A}\mathbf{A}^T)^T = (\mathbf{A}^T)^T\mathbf{A}^T = \mathbf{A}\mathbf{A}^T.$$

Altså sker der ingenting, når vi transponerer $\mathbf{A}\mathbf{A}^T$. Det samme sker for $\mathbf{A}^T\mathbf{A}$, som læseren kan verificere. En matrix $\mathbf{B}$ med egenskaben $\mathbf{B}^T = \mathbf{B}$ fortjener sit eget navn. Bemærk, at dette kun kan lade sig gøre for kvadratiske $\mathbf{B}$.

Opgaver

Opgave 3.1. Lad

$$\mathbf{v}_1 = \begin{pmatrix} -3 \\ 5 \end{pmatrix} \quad \text{og} \quad \mathbf{v}_2 = \begin{pmatrix} 2 \\ 3 \end{pmatrix}.$$

1. Beregn $\mathbf{v}_1 + \mathbf{v}_2$ og $\mathbf{v}_1 - \mathbf{v}_2$.
2. Tegn $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_1 + \mathbf{v}_2$ og $\mathbf{v}_1 - \mathbf{v}_2$ ind i et plan.

Opgave 3.2. Lad

$$\mathbf{A} = \begin{pmatrix} 5 & 1 \\ 2 & 0 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 0 & 3 \\ -1 & -5 \end{pmatrix}, \quad \mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{og} \quad \mathbf{v}_2 = \begin{pmatrix} 0 \\ -7 \end{pmatrix}.$$

Beregn $\mathbf{A}\mathbf{v}_1, \mathbf{A}\mathbf{v}_2, \mathbf{B}\mathbf{v}_1$ og $\mathbf{B}\mathbf{v}_2$.

Opgave 3.3. Lad

$$\mathbf{A} = \begin{pmatrix} 1 & 7 & -2 \\ 0 & 0 & -3 \\ -1 & 4 & 6 \end{pmatrix} \quad \text{og} \quad \mathbf{v} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

Beregn $\mathbf{A}\mathbf{v}$.

Opgave 3.4. Lad $\mathbf{A} \in \mathbb{R}^{n \times n}$ være givet ved

$$\mathbf{A} = \begin{pmatrix} a & 0 & \cdots & 0 \\ 0 & a & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & a \end{pmatrix}.$$

Sådan en matrix kaldes en *skaleringsmatrix*. Vi ser på tilfældet $n = 2$ og definerer

$$\mathbf{A} = \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix} \quad \text{og} \quad \mathbf{B} = \begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix}.$$

Lad $\mathbf{v} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

1. Beregn $\mathbf{Av}$ og $\mathbf{Bv}$.
2. Tegn $\mathbf{v}, \mathbf{Av}$ og $\mathbf{Bv}$ i et koordinatsystem. Forklar, hvad matricerne $\mathbf{A}$ og $\mathbf{B}$ gør.

Opgave 3.5. Lad $\mathbf{A} \in \mathbb{R}^{2\times 2}$, $\mathbf{B} \in \mathbb{R}^{2\times 5}$ og $\mathbf{C} \in \mathbb{R}^{5\times 2}$. Hvilke af følgende udtryk er veldefinerede?

- $\mathbf{AB}$
- $\mathbf{BA}$
- $\mathbf{BC}$
- $\mathbf{CB}$
- $\mathbf{BC} + \mathbf{A}$
- $\mathbf{CB} + \mathbf{A}$
- $\mathbf{AB} + \mathbf{C}$
- $\mathbf{ABC}$
- $\mathbf{CAB}$

Opgave 3.6. Beregn $\mathbf{AB}$ for:

1. $$\mathbf{A} = \begin{pmatrix} 2 & -1 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 4 & 6 \\ 3 & 11 \end{pmatrix}$$
2. $$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 2 & 0 \\ 3 & 5 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 7 & 1 & 0 \\ 0 & -2 & -8 \end{pmatrix}$$

3.

$$\mathbf{A} = \begin{pmatrix} 1 \\ 4 \\ 0 \\ 2 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} -2 & 3 & -5 & -1 \end{pmatrix}$$

4.

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}$$

Opgave 3.7. Bestem determinanterne af følgende matricer:

1.

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

2.

$$\mathbf{B} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

3.

$$\mathbf{C} = \begin{pmatrix} 4 & 7 \\ 0 & 17 \end{pmatrix}$$

4.

$$\mathbf{D} = \begin{pmatrix} 2 & 4 \\ 6 & 12 \end{pmatrix}$$

Opgave 3.8. Bestem den inverse af følgende matricer:

1.

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

2.
$$\mathbf{B} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

3.
$$\mathbf{C} = \begin{pmatrix} 4 & 3 \\ 0 & 1 \end{pmatrix}$$

4.
$$\mathbf{D} = \begin{pmatrix} -2 & 8 \\ 2 & 4 \end{pmatrix}$$

Opgave 3.9. Bestem den transponerede af følgende matricer

1.
$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 0 & 0 \end{pmatrix}$$

2.
$$\mathbf{B} = \begin{pmatrix} 7 & -1 & -3 \\ 9 & 0 & 2 \end{pmatrix}$$

3.
$$\mathbf{C} = \begin{pmatrix} 10 \\ 4 \\ 8 \end{pmatrix}$$

Opgave 3.10. Betragt matricen

$$\mathbf{Q} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}.$$

Bestem $\mathbf{Q}^T$. Vis herefter ved udregning, at $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}_3$ og $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}_3$. En matrix, hvis inverse er lig den transponerede, kaldes *ortogonal*.

Opgave 3.11. Givet en matrix $\mathbf{A}$ kan man under visse betingelser lave en såkaldt LU-faktorisering af $\mathbf{A}$. En LU-faktorisering af $\mathbf{A}$ er en opskrivning $\mathbf{A} = \mathbf{LU}$, hvor $\mathbf{L}$ er nedre triangulær, og $\mathbf{U}$ er øvre triangulær. Hvis alle diagonalindgangene i $\mathbf{L}$ vælges til at være 1, kaldes faktoriseringen Doolittle's faktorisering. Vælges alle diagonalindgangene i $\mathbf{U}$ til at være 1, kaldes faktoriseringen Crout's faktorisering.

Find Doolittle-faktoriseringen og Crout-faktoriseringen for følgende matricer.

1.
$$\mathbf{A} = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}.$$

2.
$$\mathbf{B} = \begin{pmatrix} 4 & 1 \\ 1 & 4 \end{pmatrix}.$$

4 Differentialregning

Det skal bemærkes at vi herunder vil beskrive det nødvendige materiale for differentialregning, til at kunne arbejde videre med det i forhold til machine learning. Emnet *differentialregning* er en del bredere end hvad der vil blive nævnt her og der lægges op til at arbejde videre med det i fremtiden.

Differentialregning er et middel i matematik der bruges til at sige noget om hvordan en funktion ændre sig ved forskellige punkter. Dette kan være sværere at forstå konceptuelt end funktioner eller polynomier. Derfor vises der eksempler med grafer for hvordan funktioner ændre sig ved bestemte punkter:

Der er flere måder at betegne at man differentierer et udtryk, eller en funktion. Hvis man f.eks. vil differentiere en funktion $f(x)$, i forhold til x, kan man bl.a. skrive det på følgende måder:

$$f'(x), \frac{\mathrm{d}f(x)}{\mathrm{d}x}, \dot{f}(x), \ldots$$

De mest almindelige måder, og de måder der vil blive brugt, er $f'(x)$, og $\frac{\mathrm{d}f}{\mathrm{d}x}$.
Der findes regneregler for hvordan man kan differentiere forskellige opbygninger af funktioner og ligninger, hvor de mere relevante for undervisningen kan findes i Tabel 1.1.

$f(x)$	$f'(x)$
k	0
kx	k
x^n	nx^{n-1}
e^x	e^x
e^{kx}	ke^{kx}
$\ln(x)$	$1/x$

Tabel 1.1: Tabel over nogle vigtige funktioner og deres afledte.

Hvis man bliver nødt til at differentiere flere funktioner på samme tid, findes der differentationsregler for det, som kan ses i Tabel 1.2 med funktionerne $f(x)$ og $g(x)$.

$f(x), g(x)$	$f'(x), g'(x)$
$(f \pm g)'(x)$	$f'(x) \pm g'(x)$
$(f \cdot g)'(x)$	$f'(x) \cdot g(x) + f(x) \cdot g'(x)$
$\left(\frac{f}{g}\right)'(x)$	$\frac{f'(x) \cdot g(x) + f(x) \cdot g'(x)}{g(x)^2}$

Tabel 1.2: Tabel over nogle vigtige funktioner og deres afledte.

Man kan selvfølgelig støde ind i, at skulle differentiere en funktion med flere variabler i forhold til hver af de variabler. Dette hedder at finde den *partielt afledte*, eller at *partielt differentiere* en funktion og skrives med ∂ i stedet for d. Hvis man vil partiel differentiere en funktion f som har variablerne x og y, så $f(x,y)$, i forhold til x, skriver man det:

$$\frac{\partial f(x,y)}{\partial x}.$$

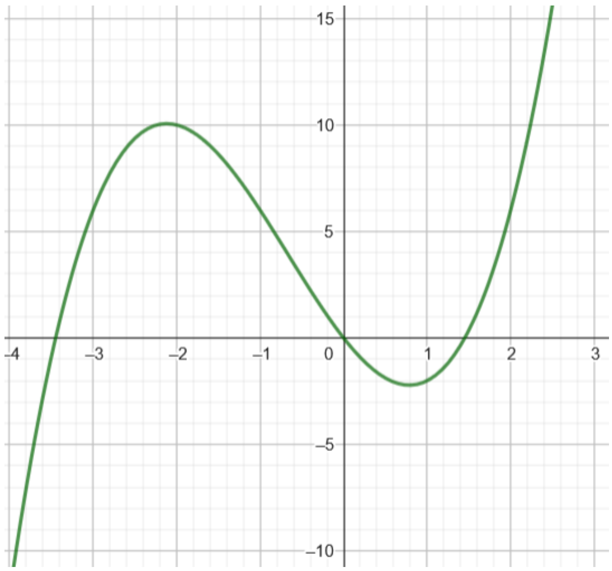
Det er vigtigt her at huske, at når man differentiere en funktion med flere variabler, i forhold til et bestem variabel, så betragtes de andre variabler som konstanter.
Eksempelvis man kan have funktionen $f(x,y) = x \cdot y^2$, og vil finde den partielt afledte i forhold til x, så $\frac{\partial f(x,y)}{\partial x}$, vil man behandle y som en konstant der ganges den differentierede på x og ikke differentiere y. Ligeledes hvis man vil finde den partielt afledte i forhold til y, $\frac{\partial f(x,y)}{\partial y}$, vil man behandle x som en konstant der ganges på den differentierede y og ikke differentiere x. Det samme koncept vil selvfølgelig gælde for funktioner med flere, men der er her kun brugt to variabler for nemhedens skyld.
Hvis man så vil finde den partielt afledte i forhold til alle de variabler funktionen har, kaldes det at finde *gradienten* af funktionen og skrives med tegnet ∇, som hedder nabla, foran funktionen med flere variabler, hvor svaret vil give en vektor med hver indgang der er den partiel afledte i forhold til hver variable. Et eksempel med hvordan det vi se ud er:

$$\nabla f(x,y) = \left(\frac{\partial f(x,y)}{\partial x}, \frac{\partial f(x,y)}{\partial y}\right) = \begin{bmatrix} \frac{\partial f(x,y)}{\partial x} \\ \\ \frac{\partial f(x,y)}{\partial y} \end{bmatrix},$$

alt efter hvordan man skriver en vektor op. Her vil man finde hvordan funktionen ændre sig ved forskellige punkter, i forhold til de forskellige variabler.

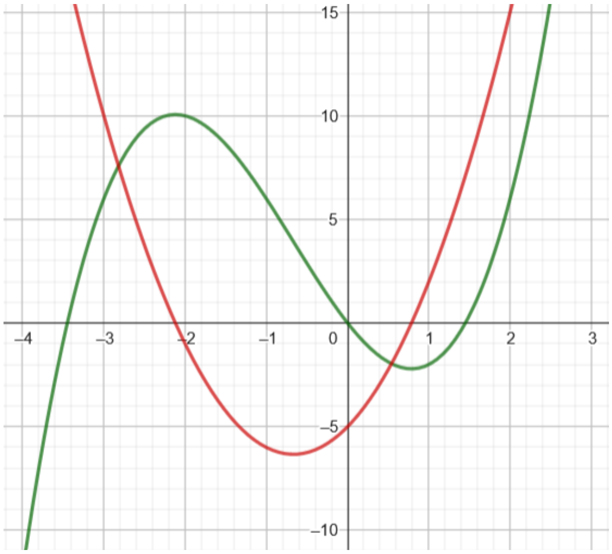
Man kan bruge differentialregning til at finde minimums- og maksimumspunkter ved at differentiere udtrykket man vil finde disse værdier for, og løse for variablet når man sætter det lig 0.
Grunden at man kan finde minimums- og maksimumspunkter ved at differentiere er at man finder hvordan funktioner ændre sig. Når funktioner ændre sig ved at gå fra opadgående til nedadgående, eller nedadgående til

opadgående, vil funktionerne begynde med at hælde negativt og positivt, respektivt. Man siger så, at man har en negativ eller positiv hældning. Når en funktion har en positiv hældning er den differentierede funktion positiv, og når funktion har en negativ hældning er den differentierede funktion negativ.
Et eksempel er at finde minimums- og maksimumspunkter for funktionen $f(x) = x^3 + 2x^2 - 5x$, som er illustreret i Figur 1.1.



Figur 1.1: Illustration af funktionen $f(x) = x^3 + 2x^2 - 5x$ (grøn).

Som man kan se på illustrationen, så har funktionen $f(x)$ et lokalt maksimum og et lokalt minimum. Vi kan finde hvad der her punkter er på x-aksen når vi finder $f'(x)$, ved at bruge de regneregler man kan se i Tabel 1.1.
Vi får så, at $f'(x) = 3x^2 + 4x - 5$, som man kan se en illustration af i Figur 1.2.



Figur 1.2: Illustration af funktionen $f(x) = x^3 + 2x^2 - 5x$ (grøn) og $f'(x) = 3x^2 + 4x - 5$ (rød).

Her ser vi, at $f'(x)$ skærer x-aksen der hvor $f(x)$ har maksimums- og minimumspunkter. Vi kan så finde disse værdier ved at sætte $f'(x)$ lig 0 og løse for x, hvilket vi kan gøre med andengradsligningen $x = \frac{-b \pm \sqrt{b^2-4ac}}{2a}$, hvor $a = 3, b = 4, c = -5$:

$$
\begin{aligned}
f'(x) &= 3x^2 + 4x - 5 = 0, \\
x &= \frac{-4 \pm \sqrt{4^2 - 4 \cdot 3 \cdot (-5)}}{2 \cdot 3}, \\
x &= \frac{-4 \pm \sqrt{16 + 60}}{2 \cdot 3}, \\
x &= \frac{-4 \pm \sqrt{76}}{2 \cdot 3}, \\
x & \frac{-2 \pm \sqrt{19}}{3} \approx \begin{cases} 0.79 & ,+ \\ -2.12 & ,- \end{cases}
\end{aligned}
$$

Altså, får vi at $f(x) = 0$ når $x \approx 0.79$ og -2.12.
Fra $f'(x)$ ser vi også hvordan $f(x)$ ændrer sig. Til venstre for maksimumspunktet i $f(x)$, ved $x \approx -2.12$ ser vi at funktionen har en positiv hældning hvor den stiger langsomt indtil den når punktet. Ved maksimumspunktet er $f'(x)$ lig nul, hvor den fortsætter negativt så længe at $f(x)$ har en negativ hældning. Der hvor $f(x)$ er stejlest, lidt før $x = -1$, ændre funktionen sig ikke, og derfor har $f'(x)$ et nulpunkt ved det punkt. Derefter vil $f'(x)$ have en positiv hældning og gå fra at være negativ til at være positiv ved minimumspunktet for $f(x)$.

Opgaver

Opgave 4.1. Differentiér funktionen $f(x) = x^2 + 7$.

Opgave 4.2. Differentiér funktionen $f(x) = x^2 + 2x + 9$.

Opgave 4.3. Differentiér funktionen $f(x) = x^2 + x - 2x + 5 - 1 + \ln(x)$.

Opgave 4.4. Differentiér funktionen $f(x) = \mathrm{e}^x + x^2 + 2x + 7$.

Opgave 4.5.

1. Tegn $f(x) = (x + 3)^2$.

2. Tegn den afledte $f'(x)$

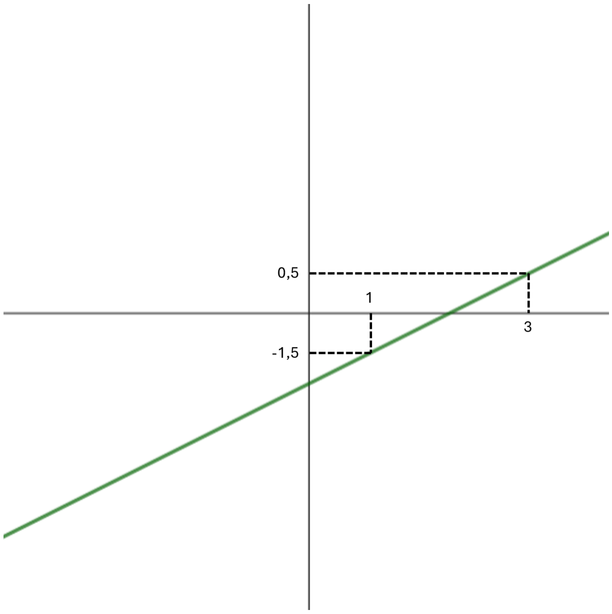
Opgave 4.6. Se på Figur 1.3.

1. Tegn den differentierede af grafen på figuren.

2. Hvor skærer den differentierede graf på y-aksen?

Opgave 4.7. Denne opgave regnes uden tal og er mere en intuitions opgave. Til det opremser vi følgende brugbare punkter:

- Afledte maksimums- og minimumspunkter (også kaldet ekstremums punkter, eller stationære punkter) er lig 0.

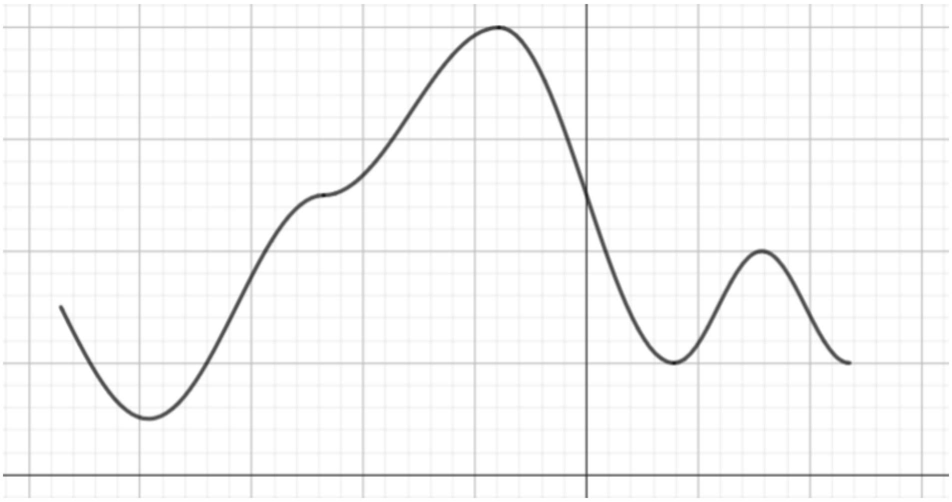


Figur 1.3: Figur for Opgave 4.6

- Der hvor funktionen er aftagende, er den afledte negativ.
- Der hvor funktionen er voksende, er den afledte positiv.
- Afledte bøjningspunkter (det stejleste punkt) bliver til nye maksimums- eller minimumspunkter.

Se på grafen i Figur 1.4.

1. Tegn den afledte funktion $f'(x)$
2. Snak med en sidemakker om hvorfor $f'(x)$ ser ud som den gør.



Figur 1.4: Figur for Opgave 4.7

5 Sandsynlighedsteori

Udfaldsrum

Når vi arbejder med sandsynligheder, støder man ind på begreber såsom *udfaldsrum*, som beskriver de mulige *udfald* der kan komme fra et *eksperiment*. Her er et eksperiment det tilfælde man nu vil finde sandsynlighederne for.
Vi skriver et udfaldsrum på formen $\{a_1, a_2, ...\}$, hvor a_1, a_2, osv., betegner de forskellige udfald. Vi betegner som regel et udfaldsrum med bogstavet Ω, kaldet *omega*.

Lad os se på et simpelt eksempel, nemlig et møntkast. Der er to udfald, nemlig K (krone) og P (plat). Dermed er udfaldsrummet lig

$$\Omega = \{K, P\}.$$

Hvad hvis vi har to møntkast? Da har vi fire muligheder, og vi kan skrive dem som KK, KP, PK, PP. F.eks. svarer PK til, at første kast er plat og andet kast krone. Dermed er udfaldsrummet for dette eksperiment lig

$$\Omega = \{KK, KP, PK, PP\}.$$

Et andet typisk eksempel er med terningekast. Lad os sige, at vi slår med en sekssidet terning. Da er udfaldsrummet lig

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

Til tider er man interesseret i at betragte bestemte samlinger af udfald. Disse kaldes *hændelser*.

Når vi har et eksperiment, som har et udfaldsrum Ω, kan vi have en *hændelse*, eller en *begivenhed*, A som er en samling af udfald i Ω. Dette kan til tider skrives som $A \subseteq \Omega$, når A er en hændelse. Hændelsen uden nogle begivenheder kaldes den *tomme begivenhed* og betegnes $\emptyset$.

Lad os se på eksemplet med at slå en mønt to gange, altså $\Omega = \{KK, KP, PK, PP\}$. Vi kunne være interesseret i hændelsen at slå mindst én plat. Denne hændelse beskrives matematisk med

$$A = \{KP, PK, PP\}.$$

Med terningekastet fra før, $\Omega = \{1, 2, 3, 4, 5, 6\}$, ville

$$A = \{2, 4, 6\}$$

være hændelsen at slå et lige tal. Inden vi kan tale om sandsynligheder, er vi nødt til at tale om, hvordan man kan relatere hændelser til hinanden.

Hvis vi har hændelserne A og B i udfaldsrummet Ω, kan vi have de følgende fire muligheder:

1. Hændelsen $A \cup B$ består af alle de udfald, der ligger enten i A eller i B (eller i begge to). $A \cup B$ kaldes *foreningen* af A og B.

2. Hændelsen $A \cap B$ består af alle de udfald, der ligger i både A og B. $A \cap B$ kaldes *snittet* af A og B.

3. Hændelsen $A \setminus B$ består af alle de udfald, der ligger i A, men ikke i B. $A \setminus B$ kaldes *differensen* af A med B.

4. Hændelsen $\Omega \setminus A$ kaldes *komplementet* til A. Det er det samme som differensen, men forskellen er at det hedder komplementet når det er med udfaldsrum.

Ser vi igen på $\Omega = \{1, 2, 3, 4, 5, 6\}$, kunne vi f.eks. betragte $A = \{2, 4, 6\}$ og $B = \{1, 2, 3\}$. Da har vi

$$\begin{aligned} A \cup B &= \{1, 2, 3, 4, 6\} \\ A \cap B &= \{2\} \\ A \setminus B &= \{4, 6\} \\ B \setminus A &= \{1, 3\} \\ \Omega \setminus A &= \{1, 3, 5\} \end{aligned}$$

Vi er nu klar til at definere sandsynligheder.

Sandsynligheder

I matematik skriver man sandsynligheder med P, som kommer fra ordet for sandsynlighed på engelsk *probability*. Sandsynligheden P for alle hændelser A kan findes i udfaldsrummet Ω, $A \subseteq \Omega$, vil være et tal mellem, og inkluderende, 0 til 1. Derudover skal P opfylde følgende egenskaber:

1. $P(\Omega) = 1$. Den samlede sandsynlighed af alle udfald er lig 1.

2. Hvis A og B er hændelser med $A \cap B = \emptyset$, altså at hændelserne A og B er uafhængige, da vil $P(A \cup B) = P(A) + P(B)$.

Det er vigtigt at huske følgende egenskab, så den gentages for at vi husker det.
Sandsynligheden for hele udfaldsrummet Ω er lig 1.

Sandsynligheden for hele udfaldsrummet Ω er lig 1.
Sandsynligheden for hele udfaldsrummet Ω er lig 1.

Man skal tænke på $P(A) = 0$ som, at A har 0% sandsynlighed for at indtræffe, mens $P(A) = 1$ betyder, at A har 100% sandsynlighed for at ske. Kravet $P(\Omega) = 1$ siger blot, at der sker noget med 100% sandsynlighed. Ω ville jo også være et dårligt udfaldsrum, hvis alle mulige udfald ikke var medtaget. Lad os se på et konkret eksempel, nemlig terningekastet fra før. En fair terning svarer til

$$P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}) = \frac{1}{6}.$$

Hvad er sandsynligheden for at slå 1 eller 3? Den er

$$P(\{1,3\}) = P(\{1\}) + P(\{3\}) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3},$$

hvor vi bruger regel to, idet $\{1\} \cup \{3\} = \{1,3\}$, og de to begivenheder overlapper ikke. Regel to fortæller os, at det er tilstrækkeligt at specificere, hvad sandsynligheden er på hvert udfald.

I sandsynlighedsteori er man ofte interesseret i at kende en sandsynlighed, hvor man ved, at en bestemt hændelse er indtruffet. Hvad er f.eks. sandsynligheden for at slå en sekser på en sekssidet terning, hvis man ved, at man har slået et lige tal? Svaret burde være $1/3$, da halvdelen af mulighederne, nemlig 1, 3 og 5, er udelukket. Vi formaliserer en betinget sandsynlighed som følger.

Når vi arbejder med et udfaldsrum Ω, og hændelserne A og B, hvor $P(B) \neq 0$, så vil den betingede sandsynlighed for A, givet B, være defineret ved

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

Altså, sagt igen kun med ord. Sandsynligheden at et tal er i hændelsen A, givet at tallet er i hændelsen B, er lig sandsynligheden for at hændelsen A

er i hændelsen B ($A \cap B$) divideret med sandsynligheden for at tallet er i hændelsen B.

Lad os tage den forrige overvejelse som eksempel. Hvad er sandsynligheden for at slå 6, hvis man ved, at man har slået et lige tal? Lad $A = \{6\}$ og $B = \{2, 4, 6\}$. Da er den ønskede sandsynlighed lig $P(A \mid B)$. Vi udregner

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{6\})}{P(\{2,4,6\})} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{2}{6} = \frac{1}{3}$$

som forventet. Lad os tage et andet eksempel. Betragt $\Omega = \{KK, KP, PK, PP\}$, altså at vi slår en mønt for at få plat eller krone to gange. Hvad er sandsynligheden for at slå krone to gange, hvis vi ved, at vi slår mindst én krone? Lad $A = \{KK\}$ være hændelsen, at vi slår krone to gange og $B = \{KK, KP, PK\}$ hændelsen, at vi slår mindst én krone. Da får vi sandsynligheden

$$P(A \mid B) = \frac{P(\{KK\})}{P(\{KK, KP, PK\})} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3},$$

som vi også ville forvente.

Lad os tage endnu et eksempel. En familie har to børn. Udfaldsrummet for kønnene på børnene er

$$\Omega = \{DD, DP, PD, PP\},$$

hvor D står for dreng og P for pige. Rækkefølgen indikerer, hvilket barn der er født først. Vi antager, at alle udfald er lige sandsynlige. Dermed er sandsynligheden for, at begge børn er piger, lig

$$P(\{PP\}) = \frac{1}{4}.$$

Hvad hvis vi ved, at én af børnene er en pige? Da får vi

$$P(\{PP\} \mid \{DP, PD, PP\}) = \frac{P(\{PP\})}{P(\{DP, PD, PP\})} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}.$$

Hvad hvis vi ved, at den ældste af børnene er en pige? Dette svarer til at betinge på begivenheden $\{PP, DP\}$. Dermed har vi

$$P(\{PP\} \mid \{PP, DP\}) = \frac{P(\{PP\})}{P(\{PP, DP\})} = \frac{\frac{1}{4}}{\frac{2}{4}} = \frac{1}{2},$$

så nu er sandsynligheden større end før. Hvorfor er sandsynligheden større blot fordi, vi ved, at den ældste er en pige? Det er fordi, vi udelukker endnu et udfald, nemlig PD. Dermed må sandsynligheden være større for, at de begge er piger.

Er der en relation mellem de to størrelser $P(A \mid B)$ og $P(B \mid A)$? De er i hvert fald ikke det samme. I skal finde et eksempel i opgaverne. Generelt har vi følgende sammenhæng, som hedder **Bayes sætning**.

Bayes sætning siger, at for hændelser A og B gælder

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}.$$

Bevis. Per definition har vi

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \quad \text{og} \quad P(B \mid A) = \frac{P(A \cap B)}{P(A)}.$$

Vi har altså $P(A \cap B) = P(B \mid A)P(A)$. Indsætter vi dette udtryk på $P(A \cap B)$'s plads i definitionen for $P(A \mid B)$, får vi det ønskede resultat. □

Sagt med ord, så når vi prøver at finde den betingede sandsynlighed for hændelsen A, givet at hændelsen B er sket ($P(A \mid B)$), er lig sandsynligheden for hændelsen B, givet at hændelsen A er sket ($P(B \mid A)$), ganget med sandsynligheden for at hændelsen A sker ($P(A)$), divideret med at hændelsen B sker ($P(B)$). Altså, at vores tro om noget opdateres baseret på nye beviser, men ikke at de nye bevise bestemmer vores tro.

Et eksempel vil være, at prøve at finde sandsynligheden for at en person har en sygdom (A), givet at en test er kommet tilbage positiv (B).
Her er

- $P(A)$: Den generelle sandsynlighed for at have sygdommen (A), før at man har taget en test.
- $P(B \mid A)$: Sandsynligheden for, at testen er positiv (B), hvis personen har sygdommen (A).
- $P(B)$: Den samlede sandsynlighed for, at testen er positiv, uanset om personen har sygdommen eller ej.

For at sætte tal på, kan vi forestille os, at vi har at gøre med en sygdom der rammer 1 % af befolkningen, $P(A) = 0.01$.
Vi har en test der kan opdage sydommen i 90 % af tilfælden, $P(B \mid A) = 0.9$.
Dog har vi problemet at testen giver også falske positive i 5 % af tilfældene, $P(B \mid \Omega \setminus A) = 0.05$.
Den totale sandsynligheden for at en test vil være positiv er:
$P(B) = P(B \mid A) \cdot P(A) + P(B \mid \Omega \setminus A)P(\Omega \setminus A)$.
Det her er fordi vi skal kigge på alle de muligheder hvor testen kan være positiv, som inkludere de muligheder hvor det passer at personen er syg, og inkludere de muligheder hvor det ikke passer at personen er syg.
$P(B) = (0.9 \cdot 0.01) + (0.05 \cdot 0.99) = 0.0585$.

Hvis vi så bruger Bayes sætning for at finde sandsynligheden for, at personen virkelig er har sygdommen:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} = \frac{0.9 \cdot 0.01}{0.0585} \approx 0.154.$$

Selvom testen kom tilbage positiv, er der kun 15,4 % sandsynlighed for, at personen faktisk har sygdommen. Dette skyldes, at sygdommen er sjælden, og der er en relativt høj falsk positiv rate.
Vi har her opdateret vores tro om at man har en sygdom, baseret på nye

beviser. Før at man havde sygdommen, kunne der være 1 % sandsynlighed for man har sygdommen, men efter at have test er der 15,4 % sandsynlighed for, at personen faktisk har sygdommen. Den her sandsynlighed kan stige, hvis man f.eks. har symptomer af sygdommen, som vil være beviser der opdaterer vores tro om at man har sygdommen.

Et andet eksempel, er at overveje om en kan lide dig, baseret på at de smilte til dig. Skrevet ind i Bayes sætning, vil dette være:

$$P(\text{personen kan lide dig} \mid \text{personen smilte til dig}) \\ = \frac{P(\text{personen smilte til dig} \mid \text{personen kan lide dig})P(\text{personen kan lide dig})}{P(\text{personen smiler bare generelt})},$$

hvor $P(\text{personen smiler bare generelt})$ inkludere at personen smiler til dig fordi at de kan lide dig:
$P(\text{personen smilte til dig} \mid \text{personen kan lide dig})P(\text{personen kan lide dig})$,
og at personen smilte til dig selvom de ikke kan lide dig:
$P(\text{personen smilte til dig} \mid \text{personen kan ikke lide dig})P(\text{personen kan ikke lide dig})$,
måske er det bare en venlig person.

Lad os sætte nogle tal på. Lad os sige:
- Hvis personen kan lide dig, smiler de til dig i 80 % af tilfældene.
- Hvis personen ikke kan lide dig, smiler de stadig nogle gange, f.eks. 30 % af tilfældene. - Generelt tror du, at der er 40 % sandsynlighed for, at personen kan lide dig.
Altså er:
$P(\text{personen smilte til dig} \mid \text{personen kan lide dig}) = 0.8$,
$P(\text{personen smilte til dig} \mid \text{personen kan ikke lide dig}) = 0.3$
$P(\text{personen kan lide dig}) = 0.4$
$P(\text{personen kan ikke lide dig}) = 1 - 0.4 = 0.6$.
Hvis vi så finder:
$P(\text{personen smiler bare generelt}) = 0.8 \cdot 0.4 + 0.3 \cdot 0.6 = 0.5$.

Vi kan så finde sandsynligheden for at personen kan lide dig, baseret på at de smilte til dig:
$P(\text{personen kan lide dig} \mid \text{personen smilte til dig}) = \frac{0.8 \cdot 0.4}{0.5} = \frac{0.32}{0.5} = 0.64.$
Hvis du så møder den her person som smiler til dig, er der så en 64 % sandsynlighed for, at personen faktisk kan lide dig.

Diskrete stokastiske variable

I praksis er det ikke så smart at skulle angive et udfaldsrum eksplicit. Vi vil hellere tildele hvert udfald et tal. Dette gøres gennem stokastiske variable.

Definition 5.1. En stokastisk variabel X er en funktion, der tildeler alle værdier i et udfaldsrum Ω et reelt tal.

Lad os som eksempel se på plat eller krone, $\Omega = \{K, P\}$. Vi kunne da se på den stokastiske variabel X givet ved

$$X(K) = 1 \quad \text{og} \quad X(P) = 0.$$

Begivenheden $\{K\}$ er da det samme som begivenheden, at $X = 1$. Vi skriver dette som $\{K\} = (X = 1)$. Tilsvarende er $\{P\} = (X = 0)$. I dette eksempel har vi mere eller mindre bare omskrevet de oprindelige udfald, men stokastiske variable er langt mere fleksible.

Lad os se på et mere kompliceret eksempel, nemlig to kast med en sekssidet terning. Der er i alt 36 udfald. Vi har nemlig seks udfald for hvert kast. Vi beskriver et udfald som et par (i, j), hvor i er øjnene i første kast og j er øjnene i andet kast. Dermed bliver Ω alle par af disse, hvor i og j løber fra 1 til 6. Lad nu X være den stokastiske variabel, der angiver summen af de to kast

$$X(i, j) = i + j.$$

Hvad er $P(X = 7)$, altså sandsynligheden for, at X er lig 7? Følgende udfald giver en sum på 7:

$$(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1),$$

som er seks ud af 36 udfald, så

$$P(X = 7) = \frac{6}{36} = \frac{1}{6}.$$

En stokastisk variabel er tilfældig per konstruktion. Vi ved jo ikke på forhånd, hvilken værdi den vil tage. Derfor er det bedste, vi kan håbe på, at beskrive størrelserne $P(X = x)$, altså sandsynligheden for, at X antager værdien x.

Definition 5.2. For en stokastisk variabel X kaldes funktionen

$$f_X(x) = P(X = x)$$

for *tætheden* af X.

Lad os arbejde videre med eksemplet fra før med de to terningekast. Her er tætheden givet ved

$$f_X(x) = \begin{cases} \frac{1}{36}, & x = 2, x = 12 \\ \frac{2}{36}, & x = 3, x = 11 \\ \frac{3}{36}, & x = 4, x = 10 \\ \frac{4}{36}, & x = 5, x = 9 \\ \frac{5}{36}, & x = 6, x = 8 \\ \frac{6}{36}, & x = 7 \\ 0, & \text{ellers} \end{cases}$$

Inden vi kan arbejde videre med stokastiske variable og sandsynligheder, er vi nødt til at få noget notation på plads, nemlig sumtegn. Lad $a_1, ..., a_n$ være nogle tal. Summen af dem skriver vi som

$$\sum_{i=1}^{n} a_i.$$

i skal fortolkes som en parameter, der løber fra 1 til n. Først er $i = 1$, og vi lægger a_1 til. Derefter er $i = 2$, og vi lægger a_2 til osv. Det er samme logik som i en for-løkke. Vi kan tage et konkret eksempel. Lad os udregne

$$\sum_{i=i}^{5} i.$$

Først er $i = 1$, så vi lægger 1 til. Derefter er $i = 2$, så vi lægger 2 til osv. Altså er

$$\sum_{i=1}^{5} i = 1 + 2 + 3 + 4 + 5 = 15.$$

På samme måde er

$$\sum_{i=1}^{4} i^2 = 1^2 + 2^2 + 3^2 + 4^2 = 30.$$

i behøver ikke at starte i 1. F.eks. er

$$\sum_{i=3}^{4} \frac{i}{2} = \frac{3}{2} + \frac{4}{2} = \frac{7}{2}.$$

Nu er vi blevet bekendte med sumtegn og kan derfor tale om en central egenskab for tætheder af en stokastisk variabel.

Sætning 5.3. Lad X være en stokastisk variabel, der kan tage værdierne $x_1, ..., x_n$. Da er f_X en ikke-negativ funktion, og vi har

$$\sum_{i=1}^{n} f_X(x_i) = 1.$$

Bevis. f_X er en ikke-negativ funktion, da sandsynligheder altid er større end eller lig nul. Da $x_1, ..., x_n$ udgør alle de værdier, X kan antage, må der være sandsynlighed 1 for, at X antager en af disse værdier. Dermed har vi

$$1 = P((X = x_1) \cup (X = x_2) \cup \cdots \cup (X = x_n)) = \sum_{i=1}^{n} P(X = x_i),$$

fordi ingen af begivenhederne $(X = x_i)$ overlapper. Dette beviser sætningen. $\square$

Vi slutter dette kapitel af med et vigtigt eksempel indenfor machine learning, nemlig *softmax*-funktionen. Lad os sige, at vi har noget data i form af nogle tal $z_1, ..., z_n$. Softmax-funktionen f på dataen er givet ved

$$f(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}}.$$

Da eksponentialfunktionen altid er strengt positiv, har vi $0 < f(z_i) < 1$ for alle i. Vi ser også, at vi får 1, såfremt vi summerer alle $f(z_i)$:

$$\sum_{i=1}^{n} f(z_i) = \frac{\sum_{i=1}^{n} e^{z_i}}{\sum_{j=1}^{n} e^{z_j}} = 1.$$

Dermed opfylder f de samme egenskaber som en tæthed for en stokastisk variabel. Vi har altså taget vores data og omdannet det til en sandsynlighedsfordeling.

Opgaver

Opgave 5.1. Du har en firesidet terning, som du slår med én gang. Hvad er udfaldsrummet Ω for eksperimentet? Opskriv hændelsen at slå 3 eller 4.

Opgave 5.2. Betragt en ottesidet terning, altså hvor udfaldsrummet er $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8\}$. Lad $A = \{1, 2, 5, 7\}$ og $B = \{2, 4, 5, 6\}$. Bestem følgende hændelser:

1. $A \cup B$
2. $A \cap B$
3. $A \setminus B$
4. $B \setminus A$
5. $\Omega \setminus A$

Opgave 5.3. Betragt udfaldsrummet for at slå plat eller krone to gange, nemlig $\Omega = \{KK, KP, PK, PP\}$. Opskriv hændelsen at slå krone mindst én gang. Bestem sandsynligheden for denne hændelse.

Opgave 5.4. Du har en pose med 5 kugler. To af dem er blå, to af dem er røde, og en er grøn. Alle kugler har samme sandsynlighed for at blive trukket.

1. Du trækker én kugle fra posen. Opskriv udfaldsrummet Ω. Angiv sandsynligheden for hvert udfald.
2. Du trækker nu to kugler med tilbagelægning. Opskriv udfaldsrummet Ω og angiv sandsynligheden for hvert udfald.

Opgave 5.5. Udregn sandsynligheden for at slå 2 med en firesidet terning, såfremt du ved, at du har slået et lige tal. Hvad bliver sandsynligheden, hvis du i stedet betinger på at have slået et ulige tal?

Opgave 5.6. Find et eksempel, der viser, at $P(A \mid B)$ og $P(B \mid A)$ ikke behøver at være det samme.

Opgave 5.7. Du har et almindeligt kortspil med 52 kort. Hvad er sandsynligheden for at trække en hjerter 2, hvis du ved, at du trækker et hjerter?

Opgave 5.8. Vis, at for alle hændelser A har vi $P(A \mid A) = 1$. Kan du give en fortolkning af dette?

Opgave 5.9. Du har en firesidet terning, som du slår to gange. Lad X være den stokastiske variabel, der angiver summen af de to kast. Opskriv tætheden for X.

Opgave 5.10. Udregn følgende summer:

1. $$\sum_{i=1}^{5}(2+i)$$

2.

$$\sum_{i=1}^{3} i^3$$

3.

$$\sum_{i=1}^{7} 3$$

Opgave 5.11. Opskriv summen af de første 100 ulige tal som et sumtegn. Gentag for de første 100 lige tal.

6 Regression

Når man har en masse punkter og gerne vil forstå sammenhængen mellem dem, kan man bruge regression til at finde en model, der beskrive den sammenhæng. Regression handler om at finde den bedste linje der passer til punkterne, så den kan forklare mønsteret i dataen, at du nemmere kan arbejde med dataen, og forudsige nye værdier. Det er lidt ligesom at tegne en den mest passende linje gennem en samling af punkter, så den fanger den tendens de har og forklarer hvilken fordeling de kommer fra.
Måden man finder den model der bedst forklarer den data man arbejder med, er at kende en masse funktioner. Men her vil vi kigge på en lineær funktion for at gøre ting lidt simplere. Men når man så har en funktion og vil undersøge hvordan den bedst kan forklarer de punkter man arbejder med, en måde at gøre det på er at kigge på afstanden mellem punkterne man arbejder med, og den forventede værdi af punkterne fra funktionen, hvor man så prøver at balancere sin funktion så at den overordnede afstand til sin funktion er mindst.
Opskriver vi det her matematisk, arbejder vi med punkter der har x og y værdier, og vi vil forklarer dem med funktionen $f(x, \beta)$, hvor β er de variabler der er en del af funktionen. Hvis vi arbejder med en lineær funktion,

$$f(x, a, b) = a \cdot x + b,$$

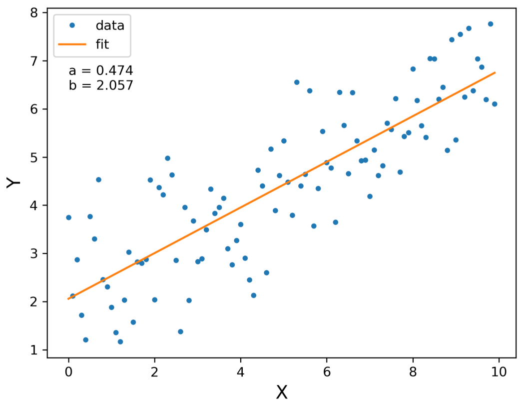
hvor a er hældningen og b er skæringspunktet med y-aksen, vil β være variablerne a og b.
Man kan så finde forskellen mellem de faktiske y-værdier og de værdier funktionen forudsiger, som kaldes fra *residualer*:

$$\varepsilon_i = y_i - f(x_i, a, b) = y_i - (a \cdot x_i + b).$$

Vi vil minimere summen af residualerne i anden, for at få den bedste lineære model:

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - (a \cdot x_i + b))^2 .$$

Grunden til at man minimere de kvadratiske residualer og ikke bare residualer, er for at det forhindre positive og negative residualer går ud med hinanden, og det straffer mere når man har store residualer end når man har små residualer, altså at det straffer mere når man rammer forkert.
Den her metode til at finde de bedste variabler kaldes for *mindste kvadraters metode* (på engelsk: lease squares method).
Man kan så indsætte forskellige værdier af variablerne a og b og se hvilke værdier af dem minimere residualerne. Vi har lavet sådan et eksempel med punkter lavet fra en lineær linje $0.5 \cdot x + 2$, som vi har tilføjet noget spredning på og lavet regression for, hvor man kan se dette i Figur 1.5 med de fundene værdier for den lineære linje.



Figur 1.5: Lineær regression af en spredningen af punkter som er fundet til at være $0.474 \cdot x + 2.057$.

De spredte punkter er fundet til at være beskrevet ved den lineære linje $0.474 \cdot x + 2.057$. Jo flere punkter der er til at beskrive funktionen vi selv prøver at beskrive den, jo bedre kan vi beskrive den.