Kapittel 12

Interpolasjon, regresjon og markovkjeder

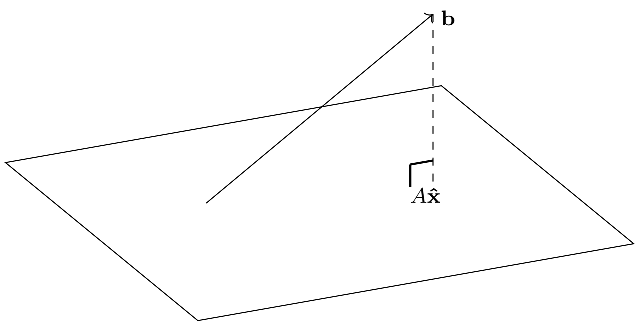
I dette kapittelet skal vi se på noen anvendelser. Interpolasjon er en enkel anvendelse av teknikker fra de første kapitlene, mens minste kvadraters metode og regresjon er anvendelser av projeksjonskapitlet. Tilslutt skal vi se på markovkjeder, som er en anvendelse av kapitlet om egenverdier og egenvektorer.

Minste kvadraters metode

Dette er en teknikk for å finne tilnærmede løsninger til systemer med flere ligninger enn ukjente. La oss si at A er en reell $m \times n$-matrise, og at $\mathbf{x}$ og $\mathbf{b}$ er kolonnevektorer i henholdsvis $\mathbb{R}^n$ og $\mathbb{R}^m$. Vi ønsker å betrakte systemet

$$A\mathbf{x} = \mathbf{b}$$

for $m > n$. Husk fra Vektorromskapitlet, at slike systemer har løsning bare når $\mathbf{b}$ ligger i kolonnerommet til A. Når systemet ikke har løsning, ønsker vi i stedet å finne $\hat{\mathbf{x}}$ som minimerer avstanden fra $A\mathbf{x}$ til $\mathbf{b}$. Vi ønsker altså at $A\hat{\mathbf{x}}$ skal være den ortogonale projeksjonen av $\mathbf{b}$ ned i kolonnerommet til A. Dette oppnår vi ved å kreve at vektoren $A\hat{\mathbf{x}} - \mathbf{b}$ skal stå ortogonalt på kolonnerommet til A.



$A\hat{\mathbf{x}}$ er punktet i kolonnerommet med minst avstand til $\mathbf{b}$

Vi husker fra kapitlet om Projeksjon at nullrommet til A^T er det ortogonale komplementet til kolonnerommet. Altså må vi ha

$$A^T(A\mathbf{x} - \mathbf{b}) = \mathbf{0},$$

som også kan skrives som:

$$A^T A\mathbf{x} = A^T\mathbf{b}.$$

Dette er et $n \times n$-system som kalles *normalligningene*. Løsningen av systemet gir den $\mathbf{x}$-en som minimerer avstanden fra $A\mathbf{x}$ til $\mathbf{b}$.

Vi skriver $\hat{\mathbf{x}}$ for denne løsningen av systemet. Punktet $A\hat{\mathbf{x}}$ er dermed punktet i kolonnerommet til A som ligger nærmest til vektoren $\mathbf{b}$.

Definisjon. Vi kaller $\hat{\mathbf{x}}$ den *minste kvadraters løsning* for $A\mathbf{x} = \mathbf{b}$. △

Vi oppsummerer diskusjonen vår i et teorem:

Teorem 12.1. *La A være en $m \times n$-matrise og $\mathbf{b}$ en kolonnevektor i $\mathbb{R}^m$. Mengden av minste kvadraters løsninger for systemet $A\mathbf{x} = \mathbf{b}$ er lik løsningsmengden for systemet*

$$A^T(A\mathbf{x} - \mathbf{b}) = \mathbf{0}.$$

Hvis $n \times n$-matrisen $A^T A$ er inverterbar, finnes det for hver $\mathbf{b}$, en unik *minste kvadraters løsning $\hat{\mathbf{x}}$ for systemet $A\mathbf{x} = \mathbf{b}$*

Merk. Hvis A er en *kompleks* $m \times n$-matrise og $\mathbf{b}$ er en kolonnevektor i $\mathbb{C}^m$, kan vi spørre om minste kvadraters løsninger $\hat{\mathbf{x}}$ til systemet $A\mathbf{x} = \mathbf{b}$ i $\mathbb{C}^n$. Denne mengden er da lik løsningsmengden for systemet

$$A^*(A\mathbf{x} - \mathbf{b}) = \mathbf{0}$$

der A^* er den adjungerte matrisen $A^* = \overline{A^T}$, som vi definerte på slutten av kapitlet om diagonalisering. △

I dette kapitlet skal vi fokusere på reelle matriser. Før vi ser på noen eksempler, bemerker vi:

Merk. Vi legger også merke til at $A^T A$ alltid er en symmetrisk matrise fordi

$$(A^T A)^T = A^T (A^T)^T = A^T A.$$

I det komplekse tilfellet har vi tilsvarende at $A^* A$ er en hermitsk matrise. △

Merk. Grunnen til at det kalles minste kvadraters metode er at avstanden $\|\mathbf{v} - \mathbf{w}\|$ mellom to punkter $\mathbf{v}$ og $\mathbf{w}$ i $\mathbb{R}^n$ måles ved å ta kvadratroten til summen av kvadratene, det vil si

$$\|\mathbf{v} - \mathbf{w}\|^2 = (v_1 - w_1)^2 + (v_2 - w_2)^2 + \cdots + (v_n - w_n)^2,$$

der v_i og w_i er de $i$te koordinatene til henholdsvis $\mathbf{v}$ og $\mathbf{w}$. Å minimere avstanden fra $\mathbf{b}$ til vektorene $A\mathbf{x}$ betyr derfor at vi minimerer en sum av kvadrater. △

Eksempel 12.2. Vi ønsker å bruke minste kvadraters metode på systemet $A\mathbf{x} = \mathbf{b}$ med

$$A = \begin{bmatrix} -1 & 2 \\ 2 & -3 \\ -1 & 3 \end{bmatrix} \text{ og } \mathbf{b} = \begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix}.$$

Vi ganger matrisen A på venstre side med sin transponerte

$$A^T = \begin{bmatrix} -1 & 2 & -1 \\ 2 & -3 & 3 \end{bmatrix}$$

og får

$$A^T A = \begin{bmatrix} -1 & 2 & -1 \\ 2 & -3 & 3 \end{bmatrix} \begin{bmatrix} -1 & 2 \\ 2 & -3 \\ -1 & 3 \end{bmatrix} = \begin{bmatrix} 6 & -11 \\ -11 & 22 \end{bmatrix}.$$

Vi ganger $\mathbf{b}$ med A^T og får

$$A^T\mathbf{b} = \begin{bmatrix} -1 & 2 & -1 \\ 2 & -3 & 3 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -4 \\ 11 \end{bmatrix}.$$

Vi vil altså finne løsningen av systemet med totalmatrise

$$\left[\begin{array}{cc|c} 6 & -11 & -4 \\ -11 & 22 & 11 \end{array}\right].$$

Gausseliminasjon gir

$$\left[\begin{array}{cc|c} 6 & -11 & -4 \\ -11 & 22 & 11 \end{array}\right] \sim \left[\begin{array}{cc|c} 6 & -11 & -4 \\ 0 & 11 & 22 \end{array}\right]$$
$$\sim \left[\begin{array}{cc|c} 6 & 0 & 18 \\ 0 & 11 & 22 \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & 0 & 3 \\ 0 & 1 & 2 \end{array}\right].$$

Løsningen er

$$\hat{\mathbf{x}} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}.$$

Dette betyr at vektoren

$$A\hat{\mathbf{x}} = \begin{bmatrix} -1 & 2 \\ 2 & -3 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}$$

er det punktet i kolonnerommet til matrisen A som minimerer avstanden til punktet $\mathbf{b}$. △

Eksempel 12.3. Vi vil finne minste kvadraters løsninger for systemet $A\mathbf{x} = \mathbf{b}$ med

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \text{ og } \mathbf{b} = \begin{bmatrix} 1 \\ 3 \\ 8 \\ 2 \end{bmatrix}.$$

Vi ganger matrisen A med sin transponerte på venstre side:

$$A^T A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix}.$$

Videre må vi også beregne $A^T\mathbf{b}$:

$$A^T\mathbf{b} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 8 \\ 2 \end{bmatrix} = \begin{bmatrix} 14 \\ 4 \\ 10 \end{bmatrix}.$$

Nå løser vi systemet $A^T A\mathbf{x} = A^T\mathbf{b}$:

$$\left[\begin{array}{ccc|c} 4 & 2 & 2 & 14 \\ 2 & 2 & 0 & 4 \\ 2 & 0 & 2 & 10 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 1 & 0 & 2 \\ 1 & 0 & 1 & 5 \\ 2 & 1 & 1 & 7 \end{array}\right]$$
$$\sim \left[\begin{array}{ccc|c} 1 & 1 & 0 & 2 \\ 0 & -1 & 1 & 3 \\ 0 & -1 & 1 & 3 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 1 & 0 & 2 \\ 0 & -1 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

Vi ser at systemet faktisk har uendelig mange løsninger som vi kan parametrisere ved hjelp av den frie variablen $t = x_3$. En generell løsning av systemet blir da

$$\begin{bmatrix} 5-t \\ -3+t \\ t \end{bmatrix} = \begin{bmatrix} 5 \\ -3 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}.$$

Med andre ord, er vektoren

$$\hat{\mathbf{x}}_0 = \begin{bmatrix} 5 \\ -3 \\ 0 \end{bmatrix}$$

en minste kvadraters løsning til systemet $A\mathbf{x} = \mathbf{b}$. Men så kan vi legge til alle vektorer som løser systemet

$$\left[\begin{array}{ccc|c} 1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

altså alle vektorer $\mathbf{y}$, slik at $A\mathbf{y} = \mathbf{0}$, som er nøyaktig alle vektorer $t \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$.

Vi konkluderer da at alle vektorer på formen

$$\hat{\mathbf{x}} = \begin{bmatrix} 5 \\ -3 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$$

for alle reelle tall t er minste kvadraters løsninger til systemet $A\mathbf{x} = \mathbf{b}$. Vi bemerker at vektoren $\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$ utspenner nullrommet til A. Det betyr at at vi har

$$A\hat{\mathbf{x}} = A\hat{\mathbf{x}}_0.$$

Vi kan derfor konkludere at punktet

$$A\hat{\mathbf{x}}_0 = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 5 \\ -3 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 5 \\ 5 \end{bmatrix}$$

er det unike punktet i kolonnerommet til A som har minst avstand fra $\mathbf{b}$. △

Interpolasjon og regresjon

Som en anvendelse av minste kvadraters metode ser vi på hvordan vi kan finne grafer som passer best til observerte data. Først ser vi på interpolasjon, og så ser vi på situasjoner der vi trenger minste kvadraters metode.

Dersom vi har $m+1$ punkter (x_i, y_i) i $\mathbb{R}^2$, der x_i-ene er forskjellig for alle punkter, vil det generelt være mulig å finne et (unikt!) reelt polynom

$$p(x) = a_m x^m + a_{m-1}x^{m-1} + ... + a_1 x + a_0,$$

hvor grafen går gjennom alle disse punktene, altså at

$$p(x_i) = y_i$$

for alle $1 \leq i \leq m+1$. Dette kalles *interpolasjon.*

Setter vi inn for (x_i, y_i) i formelen for $p(x)$, får vi et $(m+1)\times(m+1)$-ligningssystem, koeffisientene a_i er de ukjente, med totalmatrise

$$\left[\begin{array}{ccccc|c} x_1^m & x_1^{m-1} & \dots & x_1 & 1 & y_1 \\ x_2^m & x_2^{m-1} & \dots & x_2 & 1 & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ x_{m+1}^m & x_{m+1}^{m-1} & \dots & x_{m+1} & 1 & y_{m+1} \end{array}\right]$$

Det kan vises at dette ligningssystemet alltid har unik løsning så lenge $x_j \neq x_k$ for $j \neq k$, men det skal vi ikke gjøre. Det følger at du alltid kan interpolere $m+1$ punkter med et polynom av orden m på en unik måte.

Eksempel 12.4. Vi prøver å finne et annengradspolyom som går gjennom punktene $(0,1), (1,0)$ og $(2,1)$.

Et annengradspolynom skrives $p(x) = ax^2 + bx + c$, og vi skal altså finne a, b, c, som passer til disse punktene. Når vi setter inn i $p(x)$ vil disse tre punktene gi oss tre ligninger:

$$\begin{cases} \qquad\qquad c = 1 \\ a + \ \ b + c = 0 \\ 4a + 2b + c = 1 \end{cases}$$

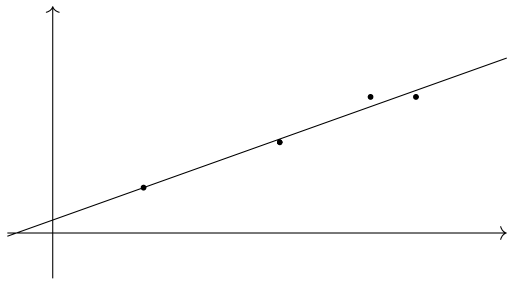
Løsningen er $a = 1$, $b = -2$ og $c = 1$, slik at polynomet blir $p(x) = x^2 - 2x + 1 = (x-1)^2$. Det er nå lett å sjekke at polynomet tar de rette verdiene i $x = 0$, $x = 1$ og $x = 2$. △

Dersom man prøver å utføre den samme prosessen med et polynom som har orden $n < m$, vil man få det overbestemte $(m+1)\times(n+1)$-systemet

$$\left[\begin{array}{ccccc|c} x_1^n & x_1^{n-1} & \dots & x_1 & 1 & y_1 \\ x_2^n & x_2^{n-1} & \dots & x_2 & 1 & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ x_{m+1}^n & x_{m+1}^{n-1} & \dots & x_{m+1} & 1 & y_{m+1} \end{array}\right].$$

Dette systemet har ikke nødvendigvis en løsning. Generelt kan vi derfor bare håpe på å finne et n-tegradspolynom som minimerer avstanden til punktene. Minste kvadraters metode er da akkurat teknikken vi trenger for å finne polynomet som passer til punktene. Prosessen å finne polynomet som passer best til punktene kalles *regresjon.*

Det enkleste tilfellet er at vi har m punkter i $\mathbb{R}^2$ og vil finne den rette linjen som passer best til disse punktene, dvs har minst avstand til alle punktene. Dette problemet oppstår for eksempel når vi observerer data med to koordinater og vi vil finne en lineær sammenheng som passer best til dataene.



Dersom $(x_1, y_1), \dots, (x_m, y_m)$ er datapunkter i $\mathbb{R}^2$, kan vi prøve å finne linjen

$$y = ax + b$$

som passer best til disse punktene. Om punktene ligger på en linje, kan vi enkelt finne a og b slik at

$$y_1 = ax_1 + b,\ y_2 = ax_2 + b, \dots \text{ og } y_m = ax_m + b.$$

Disse ligningene kan vi samle i en matriseligning

$$A\begin{bmatrix} a \\ b \end{bmatrix} = \mathbf{y} \text{ med } A = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_m & 1 \end{bmatrix} \text{ og } \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}. \tag{12.1}$$

Dette betyr at hvis vi kan løse ligningssystemet 12.1, så får vi linjen punktene ligger på. Hvis punktene ikke ligger på en linje, kan vi bruke minste kvadraters metode på systemet 12.1 for å finne a og b som gir oss linjen som approksimerer punktene best.

Merk. Man kan vise at minste kvadraters løsning gir oss linja som approkismerer punktene best, i den forstand at summen av kvadratene av de vertikale avstandene mellom punktene og linja er minst mulig. △

Eksempel 12.5. Vi har observert datapunktene $(0,4)$, $(1,-1)$, $(2,1)$, $(3,-3)$ og $(4,-1)$. Vi vil finne linjen $y = ax + b$ som passer best til disse punktene.

Vi vil altså finne minste kvadraters løsning til systemet $A \cdot \begin{bmatrix} a \\ b \end{bmatrix} = \mathbf{y}$ med

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix} \text{ og } \mathbf{y} = \begin{bmatrix} 4 \\ -1 \\ 1 \\ -3 \\ -1 \end{bmatrix}.$$

Vi må altså løse ligningssystemet $A^T A \begin{bmatrix} a \\ b \end{bmatrix} = A^T \mathbf{y}$ som er

$$\begin{bmatrix} 30 & 10 \\ 10 & 5 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} -12 \\ 0 \end{bmatrix}.$$

Gausseliminasjon gir

$$\left[\begin{array}{cc|c} 30 & 10 & -12 \\ 10 & 5 & 0 \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & -2 & -6 \\ 2 & 1 & 0 \end{array}\right]$$

$$\sim \left[\begin{array}{cc|c} 1 & -2 & -6 \\ 0 & 5 & 12 \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & 0 & -6/5 \\ 0 & 1 & 12/5 \end{array}\right].$$

Løsningen er $a = -\frac{6}{5}$ and $b = \frac{12}{5}$. Linjen som passer best til datapunktene er altså gitt ved

$$y = a + bx = -\frac{6}{5}x + \frac{12}{5}.$$

△

Vi kan også spørre om annengrads- eller n-tegradspolynomer som approksimerer datapunktene best. La oss derfor se på et eksempel til:

Eksempel 12.6. Vi prøver å finne et annengradspolyom $p(x) = ax^2 + bx + c$ som går gjennom punktene $(0, 1), (1, 0), (2, 1)$ og $(3, 2)$.

Vi setter inn de fire punktene i $p(x)$ og får ligningssystemet:

$$\begin{cases} \qquad\qquad c = 1 \\ a + \ \ b + c = 0 \\ 4a + 2b + c = 1 \\ 9a + 3b + c = 2 \end{cases}$$

Dette systemet har ingen løsning, men vi kan bruke minste kvadraters metode. Koeffisientmatrisen er:

$$A = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \end{bmatrix},$$

mens høyresiden $\mathbf{b}$ er:

$$\mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix}.$$

Den transponerte A^T er:

$$A^T = \begin{bmatrix} 0 & 1 & 4 & 9 \\ 0 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

Vi ganger A^T med A og $\mathbf{b}$, og får

$$A^T A = \begin{bmatrix} 0 & 1 & 4 & 9 \\ 0 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 98 & 36 & 14 \\ 36 & 14 & 6 \\ 14 & 6 & 4 \end{bmatrix}$$

og

$$A^T \mathbf{b} = \begin{bmatrix} 0 & 1 & 4 & 9 \\ 0 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 22 \\ 8 \\ 4 \end{bmatrix}$$

Vi må løse systemet $A^T A = A^T \mathbf{b}$, altså systemet med totalmatrise

$$\left[\begin{array}{ccc|c} 98 & 36 & 14 & 22 \\ 36 & 14 & 6 & 8 \\ 14 & 6 & 4 & 4 \end{array}\right].$$

Løsningen er

$$\begin{bmatrix} \frac{1}{2} \\ -\frac{11}{10} \\ \frac{9}{10} \end{bmatrix}$$

slik at polynomet blir

$$p(x) = \frac{1}{2}x^2 - \frac{11}{10}x + \frac{9}{10}.$$

△

Markovkjeder

Vi begynner med et eksempel.

Eksempel 12.7. Barna i en barnehage heier på fotballklubbene Arsenal, Liverpool og Manchester United i Premier League. Barna har enda ikke bestemt seg helt for hvilken klubb de liker best, og noen skifter klubb fra sesong til sesong. Det viser seg at

- når et barn heier på Manchester United, er det 50 % sannsynlighet for at barnet fortsatt heier på Manchester United neste sesong, 30 % sannsynlighet for at barnet heier på Liverpool neste sesong, og 20 % sannsynlighet for at barnet heier på Arsenal neste sesong;

- når et barn heier på Liverpool, er det 20 % sannsynlighet for at barnet heier på Manchester United neste sesong, 80 % sannsynlighet for at barnet fortsatt heier på Liverpool neste sesong, og 0 % sannsynlighet for at barnet heier på Arsenal neste sesong;

- når et barn heier på Arsenal, er det 30 % sannsynlighet for at barnet heier på Manchester United neste sesong, 30 % sannsynlighet for at barnet heier på Liverpool neste sesong, og 40 % sannsynlighet for at barnet fremdeles heier på Arsenal neste sesong.

Vi oppsummerer denne informasjonen i en tabell, der vi uttrykker sannsynlighetene ved desimaltall:

bytter fra	*MU*	*L*	*A*	**til**
	0.5	0.2	0.3	*MU*
	0.3	0.8	0.3	*L*
	0.2	0	0.4	*A*

Dette ser allerede ut som en matrise

$$M = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.8 & 0.3 \\ 0.2 & 0 & 0.4 \end{bmatrix}.$$

Faktisk hjelper matrise- og vektorregning oss å finne ut hvordan klubbsupporten skifter fra år til år. Dersom fordelingen av barna på klubbene i et år er slik at 50% heier på Man United, 30% på Liverpool og 20% på Arsenal, kan vi skrive det som en vektor

$$\mathbf{x}_0 = \begin{bmatrix} 0.5 \\ 0.3 \\ 0.2 \end{bmatrix}.$$

Etter en sesong blir fordelingen, dersom vi følger reglene for total sannsynlighet, da

$$\mathbf{x}_1 = M\mathbf{x}_0 = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.8 & 0.3 \\ 0.2 & 0 & 0.4 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0.3 \\ 0.2 \end{bmatrix} = \begin{bmatrix} 0.37 \\ 0.45 \\ 0.18 \end{bmatrix}.$$

Etter to sesonger blir fordelingen

$$\mathbf{x}_2 = M\mathbf{x}_1 = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.8 & 0.3 \\ 0.2 & 0 & 0.4 \end{bmatrix} \begin{bmatrix} 0.37 \\ 0.45 \\ 0.18 \end{bmatrix} = \begin{bmatrix} 0.329 \\ 0.525 \\ 0.146 \end{bmatrix}.$$

Vi kan fortsette denne prosessen (under antagelsen at barna forblir mange år i barnehagen), og får fordelingene $\mathbf{x}_3 = M\mathbf{x}_2$, $\mathbf{x}_4 = M\mathbf{x}_3$ osv.

Vi ser da at fordelingen endrer seg mindre og mindre etterhvert som tiden går, og konvergerer mot vektoren

$$\mathbf{q} = \begin{bmatrix} 0.3 \\ 0.6 \\ 0.1 \end{bmatrix},$$

som oppfyller ligningen

$$M \cdot \mathbf{q} = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.8 & 0.3 \\ 0.2 & 0 & 0.4 \end{bmatrix} \cdot \begin{bmatrix} 0.3 \\ 0.6 \\ 0.1 \end{bmatrix} = \begin{bmatrix} 0.3 \\ 0.6 \\ 0.1 \end{bmatrix} = \mathbf{q}. \tag{12.2}$$

△

Dette er et eksempel på en Markovkjede, som er en en stokastisk sannsynlighetsmodell hvor man beveger seg fra et stadium til det neste i en tilfeldig prosess uten hukommelse om hva som har skjedd tidligere. Altså er neste trinn kun avhengig av det nåværende og ikke av sekvensen foran.

Vi skal se på Markovkjeder der overgangen fra et trinn til neste kan beskrives ved en matrise.

Definisjon. En *sannsynlighetsvektor* er en vektor $\mathbf{v}$ i $\mathbb{R}^n$ der alle koordinatene er større eller lik 0 og summen av koordinatene er lik 1.

En $n \times n$ matrise M kalles en *stokastisk matrise* hvis kolonnene i M er sannsynlighetsvektorer, det vil si at alle elementene er ikke-negative og kolonnesummene er lik 1. △

Gitt en sannsynlighetsvektor $\mathbf{x}_0$ og en stokastiskmatrise M, kan vi sjekke at $\mathbf{x}_1 = M \cdot \mathbf{x}_0$ også er en sannsynlighetsvektor. Derfor er også $\mathbf{x}_2 = M \cdot \mathbf{x}_1$ og $\mathbf{x}_{n+1} = M \cdot \mathbf{x}_n$ en sannsynlighetsvektor. Vi oppfatter $\mathbf{x}_n$ som tilstanden på tidspunkt n som oppstår fra utgangstilstanden $\mathbf{x}_0$.

Definisjon. La M være en stokastisk matrise og $\mathbf{x}_0$ en sannsynlighetsvektor. Vi kaller følgen av vektorene

$$\{\mathbf{x}_n\} \text{ for } n = 0, 1, 2, \ldots$$

en *Markovkjede.* △

I eksempel 12.7 studerte vi en Markovkjede, og observerte at den konvergerte til en stabil tilstand. Vi stiller da det følgende spørsmålet:

> *Konvergerer alle Markovkjeder til en stabil tilstand?*

For å finne svaret på dette spørsmålet må vi undersøke stokastiske matriser. Ligning 12.2 uttrykker at $\lambda = 1$ er en egenverdi for M med egenvektor $\mathbf{q}$. Vi skal nå se at dette alltid gjelder for stokastiske matriser.

Teorem 12.8. *En stokastisk matrise M har alltid $\lambda = 1$ som en egenverdi.*

Bevis. La m_{ij} være elementet i M i rad i og kolonne j. Fordi M er en stokastisk matrise er alle kolonnesummene lik 1, dvs $\sum_i m_{ij} = 1$ for alle $j = 1, \ldots, n$. I den transponerte matrisen M^T er derfor radsummer lik 1. Vektoren $\mathbf{u} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$ er derfor en egenvektor for M^T til egenverdi $\lambda = 1$:

$$M^T \cdot \mathbf{u} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} \sum_i m_{i1} \\ \sum_i m_{i2} \\ \vdots \\ \sum_i m_{in} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$

Nå observerer vi at egenverdiene til M^T er de samme som egenverdiene til M (men egenvektorene kan være forskjellige): La A være en $n \times n$-matrise. Egenverdiene til A^T og A er like fordi

$$\det(A^T - \lambda I_n) = \det\left((A - \lambda I_n)^T\right) = \det(A - \lambda I_n).$$

Dette viser at λ er egenverdi for A hvis og bare hvis λ er en egenverdi for A^T. Da følger det at fordi $\lambda = 1$ er en egenverdi til M^T, så er $\lambda = 1$ er en egenverdi til M. □

La M være en stokastisk matrise. Fordi $\lambda = 1$ er egenverdi for M, vet vi at det finnes egenvektorer som hører til egenverdi $\lambda = 1$. Vi vet at vi finner egenvektorene som hører til egenverdi $\lambda = 1$ ved å finne ikke-trivielle løsninger for ligningssystemet

$$(M - I_n) \cdot \mathbf{x} = \mathbf{0}.$$

En løsning $\mathbf{q}$ for dette ligningssystemet som også er en sannsynlighetsvektor, dvs. som har koordinatsum lik 1 og alle koordinatene er ikke-negative, har en spesiel betydning for Markovkjeder. Derfor gir vi den et navn:

Definisjon. La M være en stokastisk matrise. En egenvektor for M som hører til egenverdi 1 og er en sannsynlighetsvektor kalles en *likevektsvektor.* △

Eksempel 12.9. I eksempel 12.7 vil vi nå finne likevektsvektoren. Metoden for dette er akkurat som i kapitlet om egenverdier, der vi nå kun er interessert i egenverdien $\lambda = 1$. Vi vil altså løse systemet $(M - I_3)\mathbf{x} = \mathbf{0}$, for å finne tilhørende egenvektorer.

$$\begin{aligned}
(M - I_3) &= \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.8 & 0.3 \\ 0.2 & 0 & 0.4 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
&\sim \begin{bmatrix} -0.5 & 0.2 & 0.3 \\ 0.3 & -0.2 & 0.3 \\ 0.2 & 0 & -0.6 \end{bmatrix} \sim \begin{bmatrix} -5 & 2 & 3 \\ 3 & -2 & 3 \\ 2 & 0 & -6 \end{bmatrix} \\
&\sim \begin{bmatrix} -5 & 2 & 3 \\ 3 & -2 & 3 \\ 1 & 0 & -3 \end{bmatrix} \sim \begin{bmatrix} 0 & 2 & -12 \\ 0 & -2 & 12 \\ 1 & 0 & -3 \end{bmatrix} \\
&\sim \begin{bmatrix} 0 & 1 & -6 \\ 0 & -1 & 6 \\ 1 & 0 & -3 \end{bmatrix} \sim \begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 6 \\ 1 & 0 & -3 \end{bmatrix} \\
&\sim \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & -6 \\ 0 & 0 & 0 \end{bmatrix}.
\end{aligned}$$

Dette viser at løsningene er alle vektorer på formen

$$t \cdot \begin{bmatrix} 3 \\ 6 \\ 1 \end{bmatrix}.$$

Fordi vi leter etter en løsning som også er en sannsynlighetsvektor, er likevektsvektoren for M vektoren

$$\mathbf{q} = \begin{bmatrix} 0.3 \\ 0.6 \\ 0.1 \end{bmatrix}.$$

△

Vi kan spørre om det finnes kun én eller flere forskjellige likevektsvektorer for en stokastisk matrise. Faktisk er likevektsvektoren unik når vi krever at M oppfyller en liten ekstra betingelse.

Definisjon. En stokastisk matrise M kalles *regulær* hvis det finnes en $k \geq 1$ slik at alle elementene i M^k er større enn 0. △

I eksempel 12.7 er M en regulær stokastisk matrise fordi

$$M^2 = \begin{bmatrix} 0.37 & 0.26 & 0.33 \\ 0.45 & 0.7 & 0.45 \\ 0.18 & 0.04 & 0.22 \end{bmatrix}.$$

Oppgave

Er matrisen $\begin{bmatrix} 1 & 0.7 & 0.4 \\ 0 & 0.3 & 0.4 \\ 0 & 0 & 0.2 \end{bmatrix}$ en regulær stokastisk matrise?

Vi så i eksemplet at Markovkjeden konvergerte mot likevektsvektoren. Dette vil alltid være tilfelle for regulære stokastiske matriser.

Teorem 12.10. *La M være en regulær stokastisk matrise. Da har M en unik likevektsvektor $\mathbf{q}$. For enhver utgangssannsynlighetsvektor $\mathbf{x}_0$ konvergerer Markovkjeden $\{\mathbf{x}_n\}$ til $\mathbf{q}$ når $n \to \infty$.*

Bevis. Vi bare skisserer idéen for beviset:
Vektoren Markovkjeden konvergerer til, må være en likevektsvektor fordi

$$\begin{aligned} M(\lim_{n\to\infty} x_n) &= M(\lim_{n\to\infty} M^n x_0) = \lim_{n\to\infty} M^{n+1} x_0 \\ &= \lim_{n\to\infty} x_{n+1} = \lim_{n\to\infty} x_n. \end{aligned}$$

Vi vet allerede at 1 er en egenverdi, og at det derfor alltid finnes en vektor $\mathbf{q} \neq \mathbf{0}$ med $M\mathbf{q} = \mathbf{q}$. Dermed kan vi ved skalering vise at det finnes en egenvektor tilhørende egenverdien 1, som også er en sannsynlighetsvektor. Det gjenstår da å vise at dette er den eneste egenvektoren med denne egenskapen, men det skal vi ikke gjøre i dette emnet (de som er interessert, kan lese om Perron-Frobenius teoremet, for eksempel på Wikipedia). □

Merk. Vi observerer at vektoren Markovkjeden konvergerer til *ikke* avhenger av utgangstilstanden $\mathbf{x}_0$. Det er et ganske overraskende resultat. Det betyr at når vi venter lenge nok, spiller startpunktet ingen rolle. △

Oppgave

Finn likevektsvektorene for de stokastiske matrisene:

- $A = \begin{bmatrix} 0.6 & 0.3 \\ 0.4 & 0.7 \end{bmatrix}$
- $B = \begin{bmatrix} 0.7 & 0.1 & 0.3 \\ 0.2 & 0.8 & 0.3 \\ 0.1 & 0.1 & 0.4 \end{bmatrix}$

Eksempel 12.11. En biolog studerer to kolonier med maur. Hun legger merke til at, når maurene er ute for å hente mat, går noen av maurene seg bort og bytter koloni. Etter lang tid med observasjon finner hun at hver time

- tar 5% av maurene fra koloni A og skifter til koloni B,
- mens 3% av maurene fra koloni B skifter til koloni A.

På et tidspunkt teller hun at 55% av alle maurene hører til koloni A, og 45% hører til koloni B.

Oppgave

Finn en stokastisk matrise som beskriver migrasjonen. Kan du finne ut hva forholdet av maurene i koloniene blir etter mange timer?

△

Eksempel 12.12. Forskere vil løse klimaproblemet, og tester forskjellige metoder for å transformere CO_2 til andre mindre skadelige gasser, vi kaller dem gass A og gass B. I eksperimentet forander det lukkede systemet med gassene seg hver dag etter det følgende mønsteret:

- 15% av gass A går over til gass B og 5% går over til CO_2, (dvs 80% av gass A forblir gass A),
- 15% av gass B går over til gass A og 5% går over til CO_2,
- mens 10% av CO_2 går over til gass A og 10% av CO_2 går over til gass B.

Forskerne begynner med en fordeling der det er like mye av alle tre gassene i systemet.

Oppgave

Kan forskerne være optimistiske basert på dette forsøket?

△