Aflevering i uge 17

Bemærkning. Da man ender med at skal løse nogle „tossede“ polynomiumsligninger, hvis der er tale om at skulle finde en formel, der minimerer summen af kvadraterne af de *vinkelrette* afstande fra nogle givne datapunkter til grafen for den fittede kurve, antages det, at der i begge opgaver er tale om de *lodrette* afstande.

Opgave 6.5.6

Først nogle generelle beregninger for at finde en forskrift for det bedste fit til en ret linje med forskriften $y = ax + b$, hvor $a, b \in \mathbb{R}$ med $a \neq 0$: Definitionen på det *bedste fit* er, at summen af kvadraterne af de lodrette afstande fra nogle givne datapunkter (x_i, y_i) for $i \in \{1, 2, \ldots, n\}$, til grafen for den fittede kurve (denne afstand betegnes ofte med R^2), skal være så lille som mulig. Det vil sige, at vi skal have minimeret følgende størrelse:

$$R^2(a,b) \coloneqq \sum_{i=1}^{n} \big(y_i - (ax_i + b)\big)^2. \tag{1}$$

Som bekendt gøres dette for $\frac{\partial R^2}{\partial a}(a,b) = \frac{\partial R^2}{\partial b}(a,b) = 0$, så af (1) finder vi nu nemt, at

$$\frac{\partial R^2}{\partial a}(a,b) = -2\sum_{i=1}^{n} \big(y_i - (ax_i + b)\big)x_i = 0,$$

$$\frac{\partial R^2}{\partial b}(a,b) = -2\sum_{i=1}^{n} \big(y_i - (ax_i + b)\big) = 0.$$

Ved at forkorte med -2, splitte summerne op (de er oplagt konvergente, så det må man godt) og bytte rundt på leddene, fås følgende:

$$a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} 1 = \sum_{i=1}^{n} y_i,$$

$$a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i.$$

Med fordel kan dette ligningssystem skrives på matrixform, som

$$\begin{pmatrix} \sum_{i=1}^{n} x_i & n \\ \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{pmatrix}. \tag{2}$$

Nu kan man så finde eksplicitte udtryk for a og b, men det er hurtigere først at indsætte følgende data i (2):

$$\{(x_i, y_i)\} = \{(1,1), (2,4), (3,6), (4,9)\}. \tag{3}$$

Af (2) og (3) fås nu med $n = 4$, at

$$\mathsf{A} \coloneqq \begin{pmatrix} \sum_{i=1}^{4} x_i & 4 \\ \sum_{i=1}^{4} x_i^2 & \sum_{i=1}^{4} x_i \end{pmatrix} = \begin{pmatrix} 1+2+3+4 & 4 \\ 1^2+2^2+3^2+4^2 & 1+2+3+4 \end{pmatrix} = \begin{pmatrix} 10 & 4 \\ 30 & 10 \end{pmatrix},$$

$$\mathsf{A}^{-1} = \cdots = \frac{1}{10}\begin{pmatrix} -5 & 2 \\ 15 & -5 \end{pmatrix},$$

$$\mathsf{B} \coloneqq \begin{pmatrix} \sum_{i=1}^{4} y_i \\ \sum_{i=1}^{4} x_i y_i \end{pmatrix} = \begin{pmatrix} 1+4+6+9 \\ 1\cdot 1 + 2\cdot 4 + 3\cdot 6 + 4\cdot 9 \end{pmatrix} = \begin{pmatrix} 20 \\ 63 \end{pmatrix}.$$

Lad nu $\mathbf{s} \coloneqq (a, b)$.[1] Så giver (2), at

$$\mathsf{A}\mathbf{s} = \mathsf{B} \implies \mathbf{s} = \mathsf{A}^{-1}\mathsf{B},$$

$$\begin{pmatrix} a \\ b \end{pmatrix} = \frac{1}{10}\begin{pmatrix} -5 & 2 \\ 15 & -5 \end{pmatrix}\begin{pmatrix} 20 \\ 63 \end{pmatrix} = \frac{1}{10}\begin{pmatrix} (-5)20 + 2\cdot 63 \\ 15\cdot 20 + (-5)63 \end{pmatrix} = \begin{pmatrix} 13/5 \\ -3/2 \end{pmatrix}.$$

Disse værdier minimerer altså (1) mest muligt, så den bedste rette linje gennem punkterne i (3) har forskriften

$$y = \frac{13}{5}x - \frac{3}{2}. \tag{4}$$

[1] Dette er en vektor, og må derfor ikke forveksles med et talpar, som eksempelvis i (3).