Arbejdsseddel: Bevis for lineær regression

KBJ, januar 2025 3u MA

**Udangspunkt**

Vi får givet et datasæt med $n$ datapunkter:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| $$x$$ | $$x\_{1}$$ | $$x\_{2}$$ | $$…$$ | $$x\_{i}$$ | $$…$$ | $$x\_{n}$$ |
| $$y$$ | $$y\_{1}$$ | $$y\_{2}$$ | $$…$$ | $$y\_{i}$$ | $$…$$ | $$y\_{n}$$ |

Vi søger en lineær funktion som model over vores data på formen: $f\left(x\right)=a·x+b$

Vi vil i det følgende bruge sumtegn i betydningen:

$$\sum\_{i=1}^{n}x\_{i}=x\_{1}+x\_{2}+…+x\_{n}=∑x\_{i}$$

Bemærk at vi forkorter sum-tegns-notationen, så vi beslutter at det altid er summen fra 1 til $n$ uden at skrive dette eksplicit. Ellers bliver det følgende ulideligt.

**Opgave 1**

Vi betegner middelværdien af alle $x$-værdier i tabellen $\overbar{x}$ (udtales $x$-streg) og tilsvarende middelværdien af alle y-værdier med $\overbar{y}$.

1. Opskriv udtryk for $\overbar{x}$ og $\overbar{y}$ ved sumtegn, og udtryk derpå summerne $x$ og $y$ ved deres middelværdier.

**Opgave 2**

1. Mind dig selv om hvad begrebet *residual* dækker over.
2. Opstil et udtryk for residual nr. $i$, dvs $r\_{i}$ dels ved $f\left(x\right)$ dels ved $a$ og $b$.
3. Bestem et udtryk for det kvadrerede residual $r\_{i}^{2}$ - ophæv alle parenteser.

**Opgave 3**

Den bedste rette linje er den linje, hvor summen af kvadrerede residualer bliver mindst mulig.

Vi indfører derfor funktionen ”Sum of Squared Errors”, kaldet $SSE$:

$$SSE\left(a,b\right)=∑r\_{i}^{2}$$

1. Hvorfor er $SSE$ en funktion af $a$ og $b$?
2. Opskriv den fulde forskrift for $SSE\left(a,b\right)$.
3. Hvorfor er vi særligt interesserede i *minimum* for $SSE$?

**Opgave 4**

Vi ønsker nu at finde *minimum* for $SSE\left(a, b\right)$:

1. Bestem partialt afledede $SSE\_{a}^{'}\left(a,b\right)$ og $SSE\_{b}^{'}\left(a,b\right)$.
2. Bestem stationære punkter for $SSE$ ved at løse ligningssystemet $SSE\_{a}^{'}\left(a,b\right)=0$ og $SSE\_{b}^{'}\left(a,b\right)=0$. (Hint! Start med anden ligning).
3. Hvilke antagelser må der laves undervejs.
4. Hvor mange stationære punkter har $SSE$?

**Opgave 5**

1. Undersøg at det stationære punkt for $SSE$ faktisk er af arten ”et minimum”.

**Opgave 6**

1. Udfør manuel lineær regression på følgende data, med formlerne udledt i opgave 4.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| $$x$$ | $$1$$ | $$2$$ | $$3$$ | $$6$$ |
| $$y$$ | $$2$$ | $$7$$ | $$8$$ | $$13$$ |

1. Tjek med Nspire at det passer.