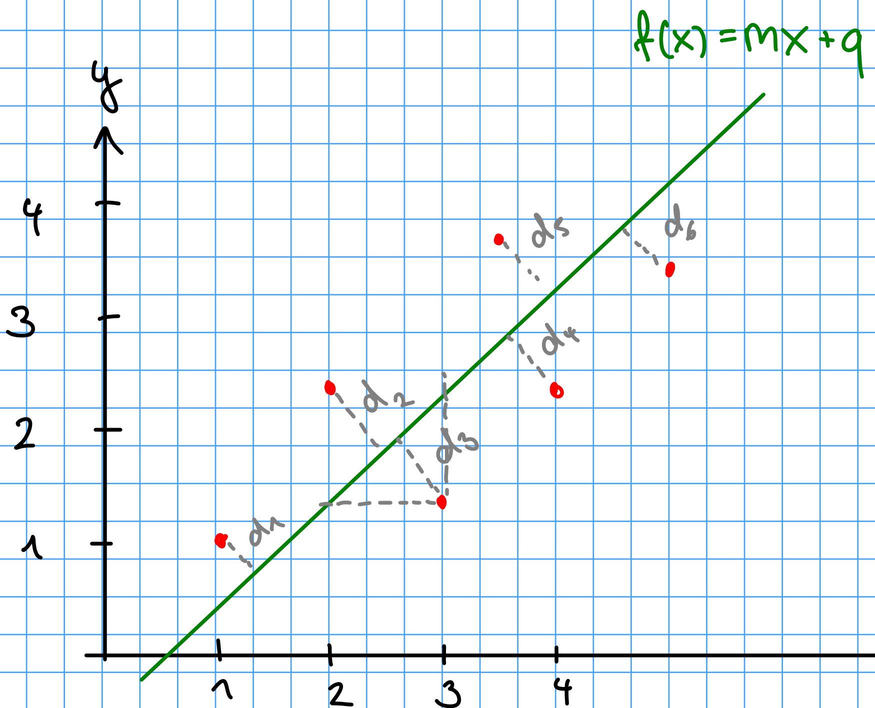
Lineare Regression

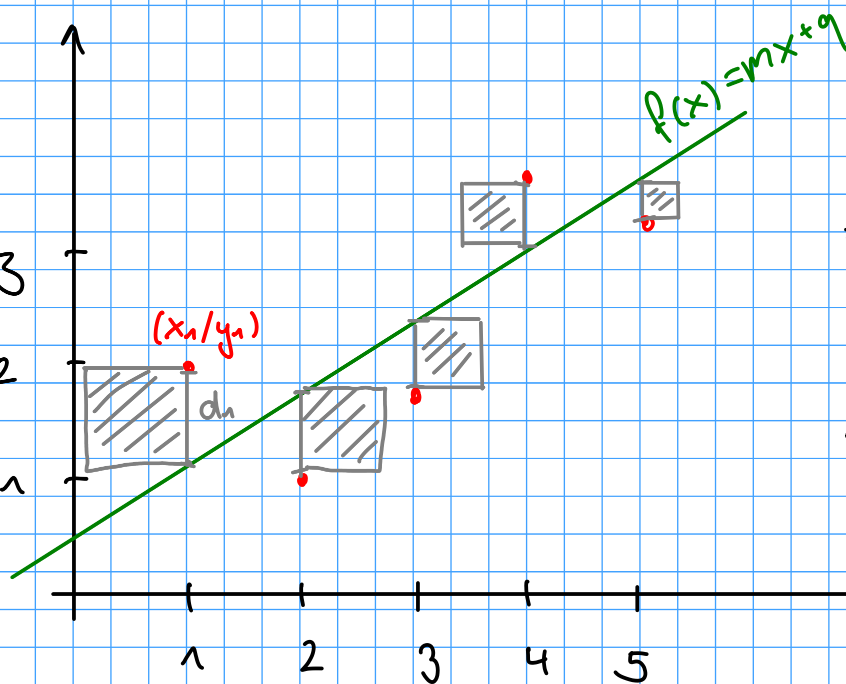
19. September 2016



Idee:

$\sum_{i=1}^{n} d_i$ soll möglichst klein sein.

Ist diese Gerade wirklich eindeutig?



$d_1 = |f(x_1) - y_1|$

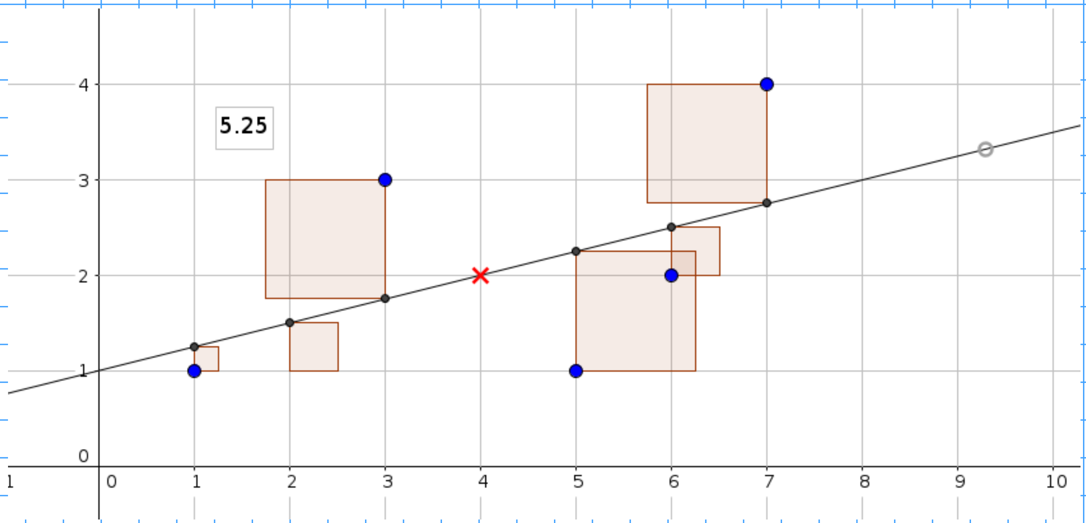
$\Rightarrow$ Betragstriche sind mühsam!

Alternative: $d_1 = (f(x_1) - y_1)^2$

Ziel: Die Summe aller Fehlerquadratflächen //// soll minimal werden.

Wie gross sind m und q?

Aufgabe 1



Berechnen von m und q:

x_i	y_i	$f(x_i)$	$(y_i - f(y_i))^2$
1	1	$m \cdot 1 + q$	$(1 - m + q)^2$
2	1	$m \cdot 2 + q$	$(1 - 2m + q)^2$
3	3	$m \cdot 3 + q$	$(3 - 3m + q)^2$
4	2	⋮	$(2 - 4m + q)^2$
5	1	⋮	$(1 - 5m + q)^2$
6	2	⋮	$(2 - 6m + q)^2$
7	4	$m \cdot 7 + q$	$(4 - 7m + q)^2$

Schwerpunkt (4,2) liegt auf der Geraden f(x)

$2 = m \cdot 4 + q$

$q = 4m - 2$

Einsetzen

+ Alternativer Lösungsweg für Aufgabe 1

x_i	y_i	$\hat{y}(x_i)$	$\hat{y} - y_i$	$(\hat{y} - y_i)^2$
1	1	$m\cdot(-3)+2$	$-3m+1$	$9m^2-6m+1$
2	1	$m\cdot(-2)+2$	$-2m+1$	$4m^2-4m+1$
3	3	$m(-1)+2$	$-m-1$	m^2+2m+1
4	2	2	0	0
5	1	$m+2$	$m+1$	m^2+2m+1
6	2	$2m+2$	$2m$	$4m^2$
7	4	$3m+2$	$3m-2$	$9m^2-12m+4$

FQS: $S(m) = 28m^2 - 18m + 8$

Fehler-Quadrat-Summe

$$\hat{y}(x) = mx + q \underset{\uparrow}{=} m(x - x_0) + y_0 = m(x-4) + 2$$

Wir wissen

$\overline{x} = 4$ (Durchschnitt der x_i); $\overline{y} = 2$ (Durchschnitt der y_i)

$A(4/2)$ liegt auf der Regressionsgerade

An welcher Stelle / für welche Steigung m wird

$$S(m) = 28m^2 - 18m + 8$$

minimal? $\longrightarrow$ Scheitelpunkt

$28 > 0 \Rightarrow$ nach oben geöffnete Parabel

$$x_{SP} = \frac{-b}{2a} = \frac{18}{56} = 0.32 = m$$

$$\hat{y}(x) = 0.32(x-4) + 2$$ Gleichung der Regressionsgeraden.

$\rightarrow$ $\boxed{\text{Aufgabe 2}}$

$xi:=\{4,2,3\}$ $\quad\{4,2,3\}$

$yi:=\{34,45,60\}$ $\quad\{34,45,60\}$

$\mathrm{sum}(yi)$ $\quad 139$

$\mathrm{sum}(xi)$ $\quad 9$

$\frac{9}{3}$ $\quad 3$

$avx:=3$ $\quad 3$

$avy:=\frac{139}{3}$ $\quad \frac{139}{3}$

$$\sum_{i=1}^{3}\left(\left(m\cdot(xi[i]-avx)+avy-yi[i]\right)^2\right)$$

$$2\cdot m^2+22\cdot m+\frac{1022}{3}$$

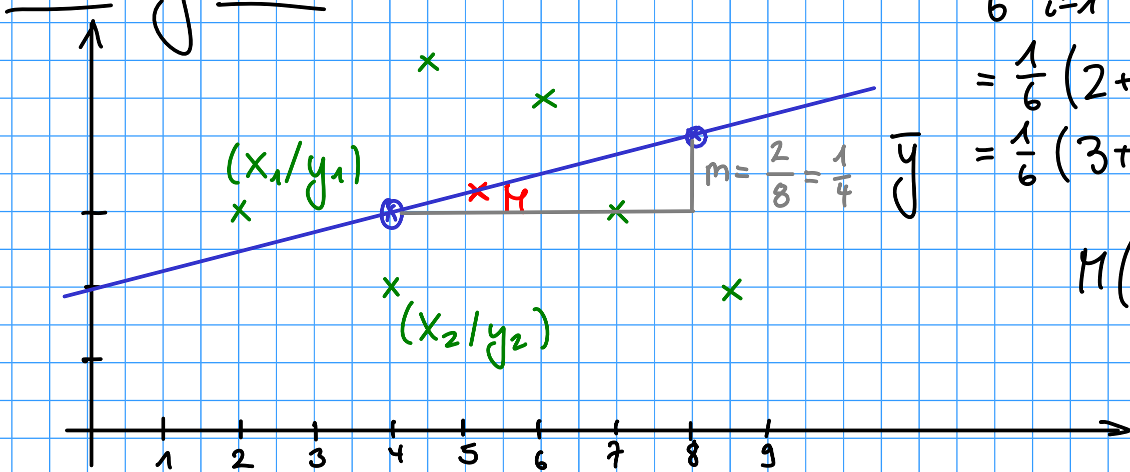
HA: Bis Aufgabe c)

24. Oktober 2016

Ü1: Alles

Ü2: Aufgaben 1, 2 (Rest)

Übung 2



$$\overline{x}=\frac{1}{6}\sum_{i=1}^{6}x_i$$

$$=\frac{1}{6}(2+4+4.5+6+7+8.5)$$

$$\overline{y}=\frac{1}{6}(3+2+5+4.5+3+2)$$

$$M(5.\overline{3}/3.25)$$

Geschätzte Gerade: $y = m(x - 5.\overline{3}) + 3.25$

$$y = \frac{1}{4}(x - 5.\overline{3}) + 3.25$$

x_i	y_i	$\hat{y}(x_i)$	$\hat{y}(x_i) - y_i$	$(\hat{y}(x_i) - y_i)^2$
2	3	$m(2-5.\overline{3}) + 3.25$	$-3.\overline{3}m + 0.25$	
				Σ

$\hat{y}$: Regressionsgerade $m(x - x_0) + y_0 = m(x - 5.\overline{3}) + 3.25$

$y(x):=m\cdot(x-5.33)+3.25$ — Done

$xi:=\{2,4,4.5,6,7,8.5\}$ — $\{2,4,4.5,6,7,8.5\}$

$yi:=\{3,2,5,4.5,3,2\}$ — $\{3,2,5,4.5,3,2\}$

$\text{sum}\left((y(xi)-yi)^2\right)$

$26.8334\cdot m^2 + \boxed{5.}\cdot m + 7.875$

↳ Float 5.0

$$m_s = \frac{-b}{2a} = \frac{-5}{2\cdot 26.8334} = -0.09$$

$$\boxed{y = -0.9(x - 5.\overline{3}) + 3.25}$$