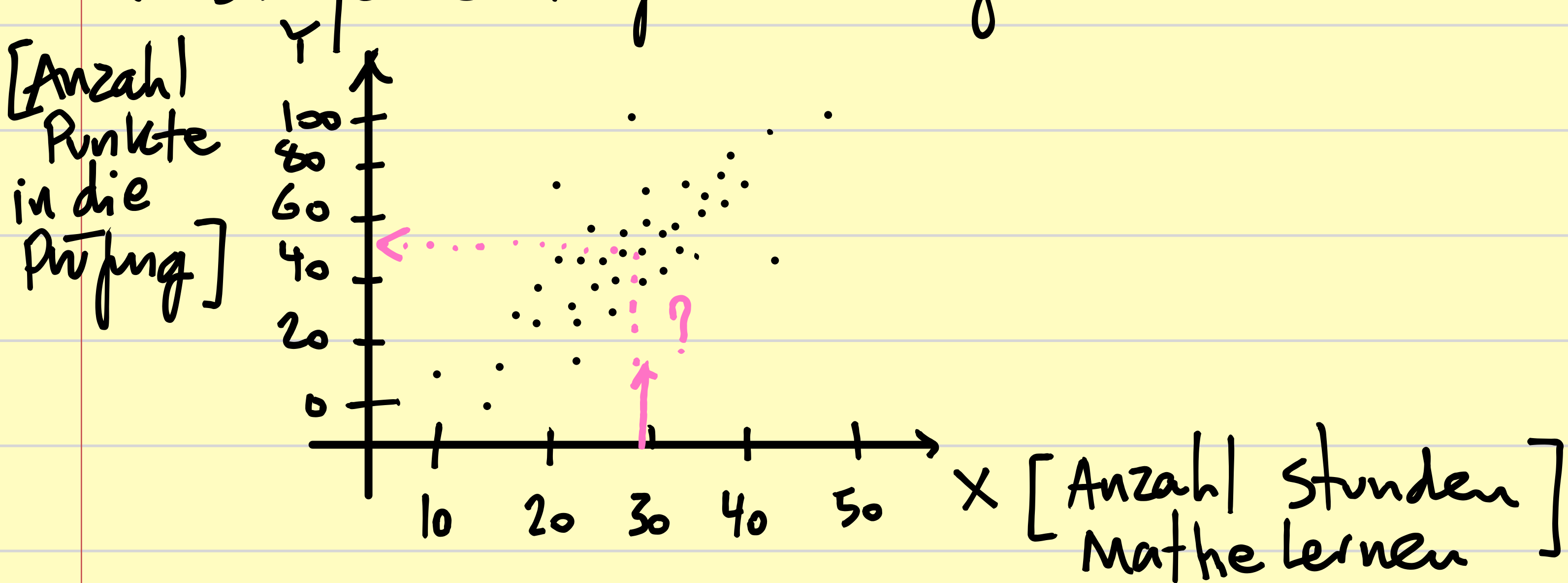
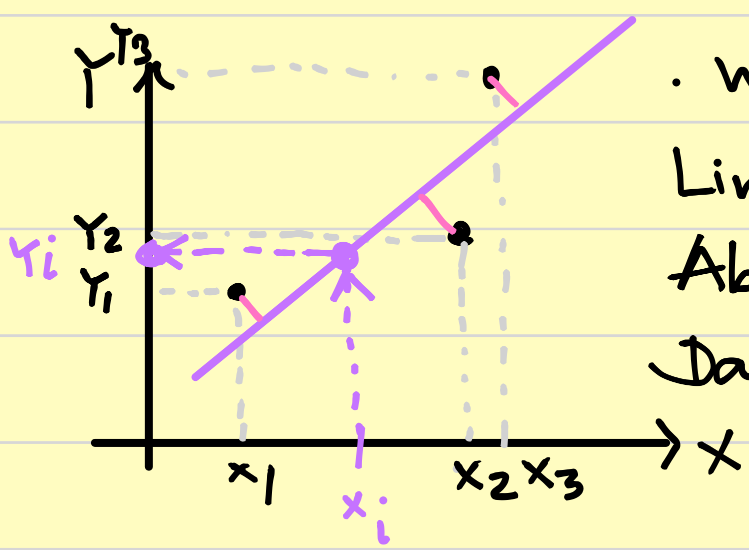
Regressionsalgorythmen. Prognostizieren anhand Daten.

- LINEARE REGRESSION
- POLYNOMISCHE REGRESSION

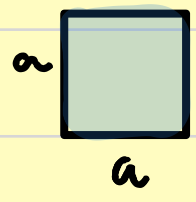
Hypothese. Wir haben Daten aus einem System erhoben, und wir wollen das Verhalten vom System bei anderen Daten mit Hilfe der Regression erfahren.



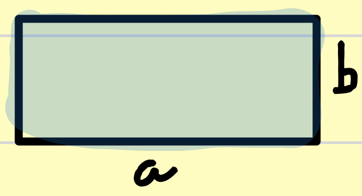
LINEARE PROGNOSE



- Wir streben die Ermittlung einer Linie $y = b_0 + b_1 x$ an, dessen Abstand zu den Punkten des Datasets minimal ist.



Fläche = a^2



Fläche = $a \cdot b$

Schritte der linearen Prognose (Regression)

1. DATENERMITTLUNG.

	X	Y
cw_1	3	6'5
cw_2	4	8'5
cw_3	6	13
cw_4	3	3'5

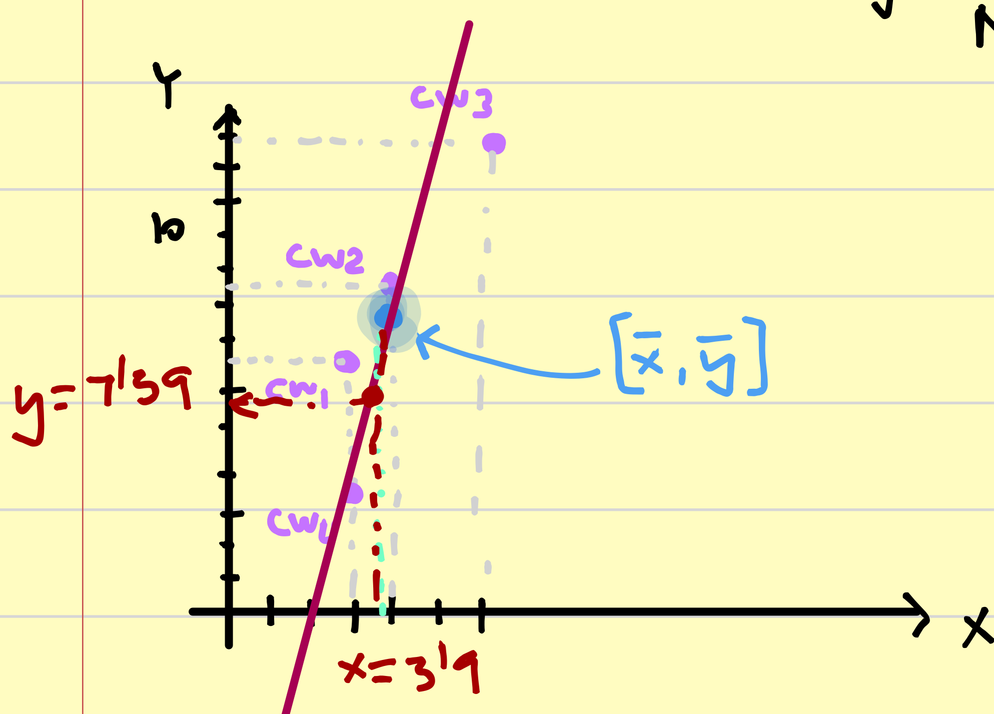
2. MITTELWERT DER VARIABELN (Schwerpunkt des Datasets)

DIE LINEARE REGRESSION GEHT DURCH DEN SCHWERPUNKT

$$[\bar{x}, \bar{y}] \longrightarrow \bar{y} = b_0 + b_1 \bar{x}$$

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{1}{4}\left[3+4+6+3\right] = 4$$

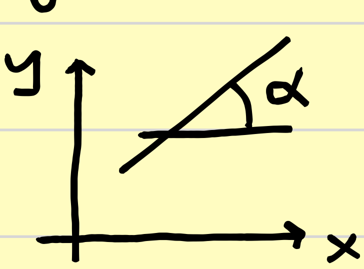
$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i = \frac{1}{4}\left[6'5+8'5+13+3'5\right] = 7'875$$



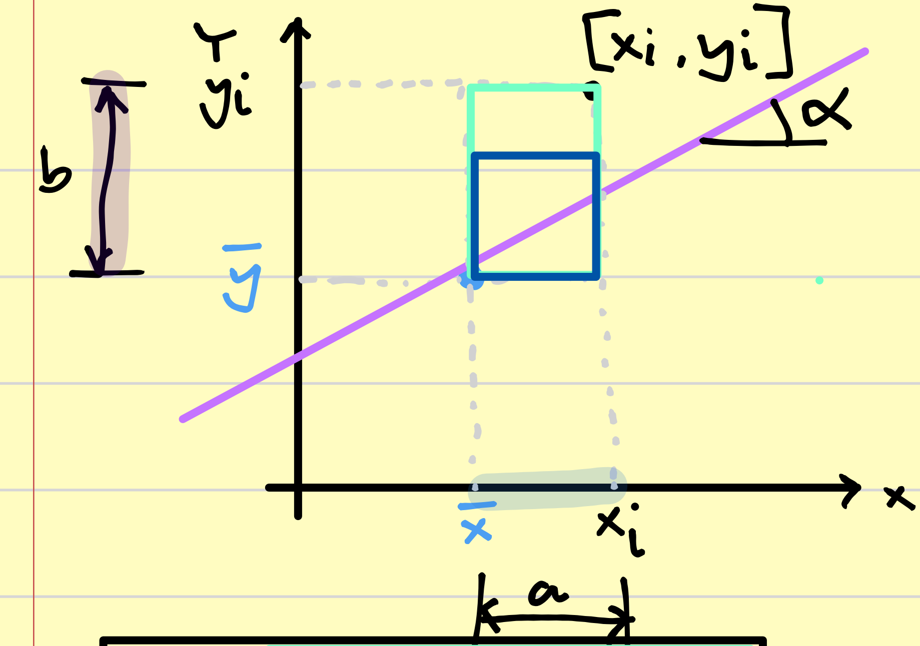
3. NEIGUNG DER LINEAREN PROGNOSE

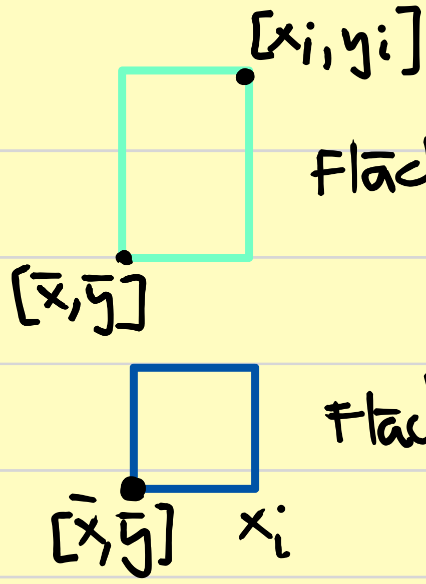
b_0

$$y = b_0 + b_1 \cdot x$$



$$\operatorname{tg}\alpha = b_1 = \frac{dy}{dx}$$





Fläche $= (x_i - \bar{x})(y_i - \bar{y})$

Fläche $= (x_i - \bar{x})^2$

$$\rightarrow \boxed{b_1 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N} (x_i - \bar{x})^2}} = \frac{(3-4)(6'5-7'875)+(4-4)(8'5-7'875)+(6-4)(13-7'875)+(3-4)(3'5-7'875)}{(3-4)^2+(4-4)^2+(6-4)^2+(3-4)^2} =$$

	X	Y
CW_1	3	6'5
CW_2	4	8'5
CW_3	6	13
CW_4	3	3'5

$\bar{x} = 4; \bar{y} = 7'875$

$$= \frac{29}{6} = 4'833$$

$$y = b_0 + 4'833\ x$$

4. SCHRITT . Ermittlung von b_0 .

$$\bar{y} = b_0 + 4'833\ \bar{x} \longrightarrow 7'875 = b_0 + 4'833 \cdot 4 \rightarrow b_0 = -11'458$$

$$\boxed{y = -11'458 + 4'833\ x}$$

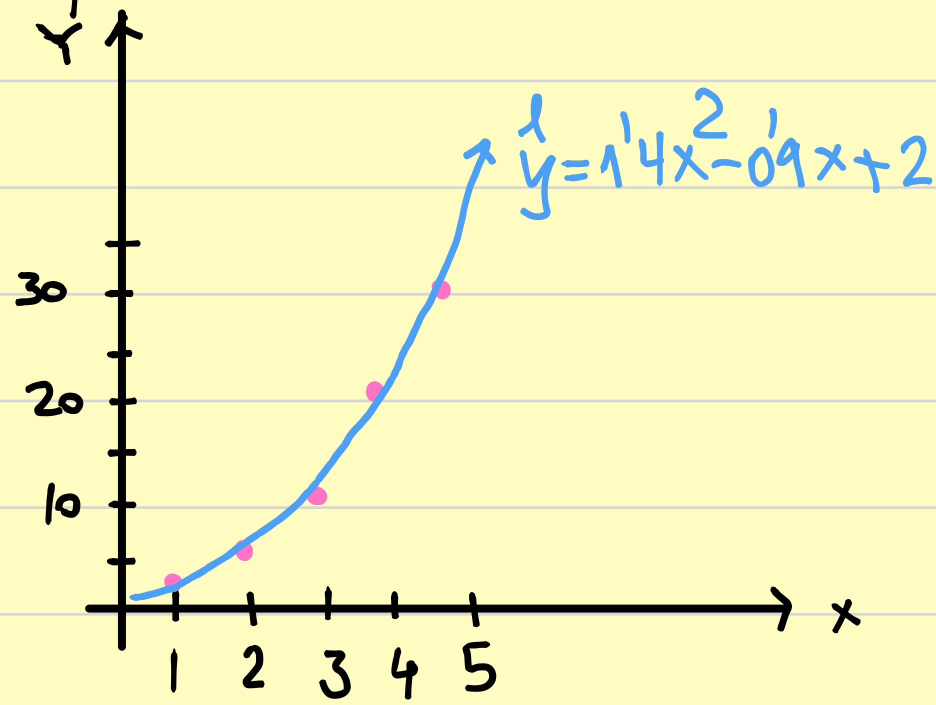
Wenn $x = 3'9$, wie ist die Lineare Prognose für den Y-Wert?

$$y_{[x=3'9]} = -11'458 + 4'833\,(3'9) = 7'3903$$

NICHT-LINEARE REGRESSION

Wir wollen die Beziehung zwischen einer Variable X und einer Variable Y mithilfe einer NICHT LINEAREN REGRESSION analysieren.

	X	Y
cw_1	1	2'5
cw_2	2	5'8
cw_3	3	11'9
cw_4	4	21'4
cw_5	5	31'2



Hypothese. Parabel. $y = ax^2 + bx + c$

1. $[1, 2'5] \longrightarrow 2'5 = a \cdot 1^2 + b \cdot 1 + c = a + b + c$ (1)
2. $[2, 5'8] \longrightarrow 5'8 = a \cdot 2^2 + b \cdot 2 + c = 4a + 2b + c$ (2)
3. $[3, 11'9] \longrightarrow 11'9 = a \cdot 3^2 + b \cdot 3 + c = 9a + 3b + c$ (3)

$(2) - (1) \longrightarrow 5'8 - 2'5 = 4a - a + 2b - b + c - c$

$3'3 = 3a + b$ (4)

$(3) - (2) \longrightarrow 11'9 - 5'8 = 9a - 4a + 3b - 2b + c - c$

$6'1 = 5a + b$ (5)

$(5) - (4) \longrightarrow 6'1 - 3'3 = 5a - 3a + b - b$

$2'8 = 2a \rightarrow a = 1'4 \xrightarrow[(5)]{} b = -0'9 \xrightarrow[(1)]{} c = 2$

$\hat{y} = 1'4x^2 - 0'9x + 2$

Das Modell überprüfen:

	x	Y	Vorhergesagte $\hat{y} = 1{,}4x^2 - 0{,}9x + 2$	ERROR Δ $(\hat{y} - y)$
CW1	1	$2{,}5$	$1{,}4 \cdot (1^2) - 0{,}9 \cdot 1 + 2 = 2{,}5$	0
CW2	2	$5{,}8$	$1{,}4 \cdot (2^2) - 0{,}9 \cdot 2 + 2 = 5{,}7$	$-0{,}1$
CW3	3	$11{,}9$	$1{,}4 \cdot (3^2) - 0{,}9 \cdot 3 + 2 = 11{,}9$	0
CW4	4	$21{,}4$	$1{,}4 \cdot (4^2) - 0{,}9 \cdot 4 + 2 = 21{,}4$	0
CW5	5	$31{,}2$	$1{,}4 \cdot (5^2) - 0{,}9 \cdot 5 + 2 = 31{,}5$	$0{,}3$

Das Modell passt gut zu den Daten, auch wenn es leichte Abweichungen gibt

Welche Y-Werte sind in der CW7 zu erwarten?

$$\hat{y}_{[CW_7]} = 1{,}4 \cdot 7^2 - 0{,}9 \cdot 7 + 2 = 64{,}3$$