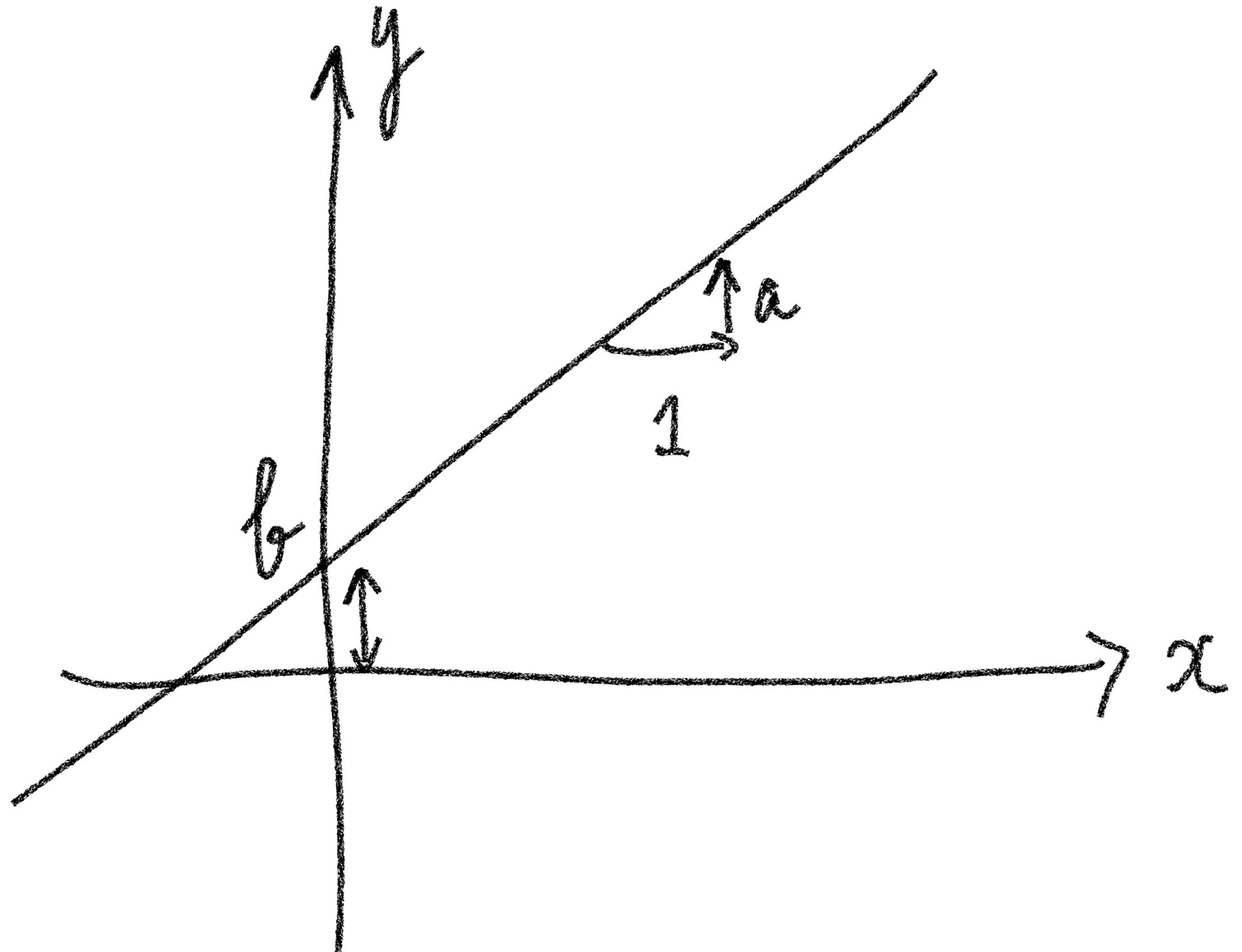


Régression linéaire

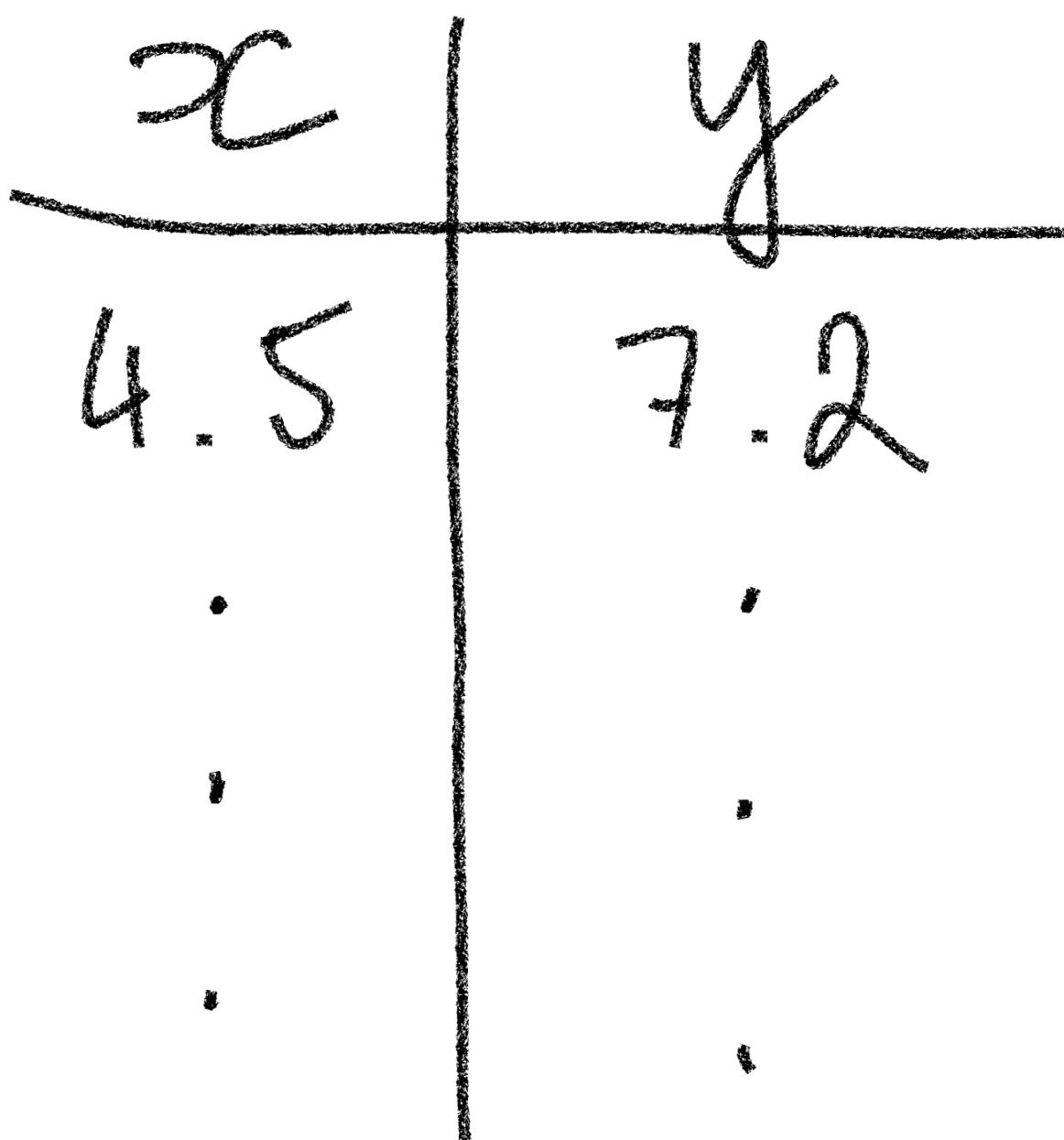
L



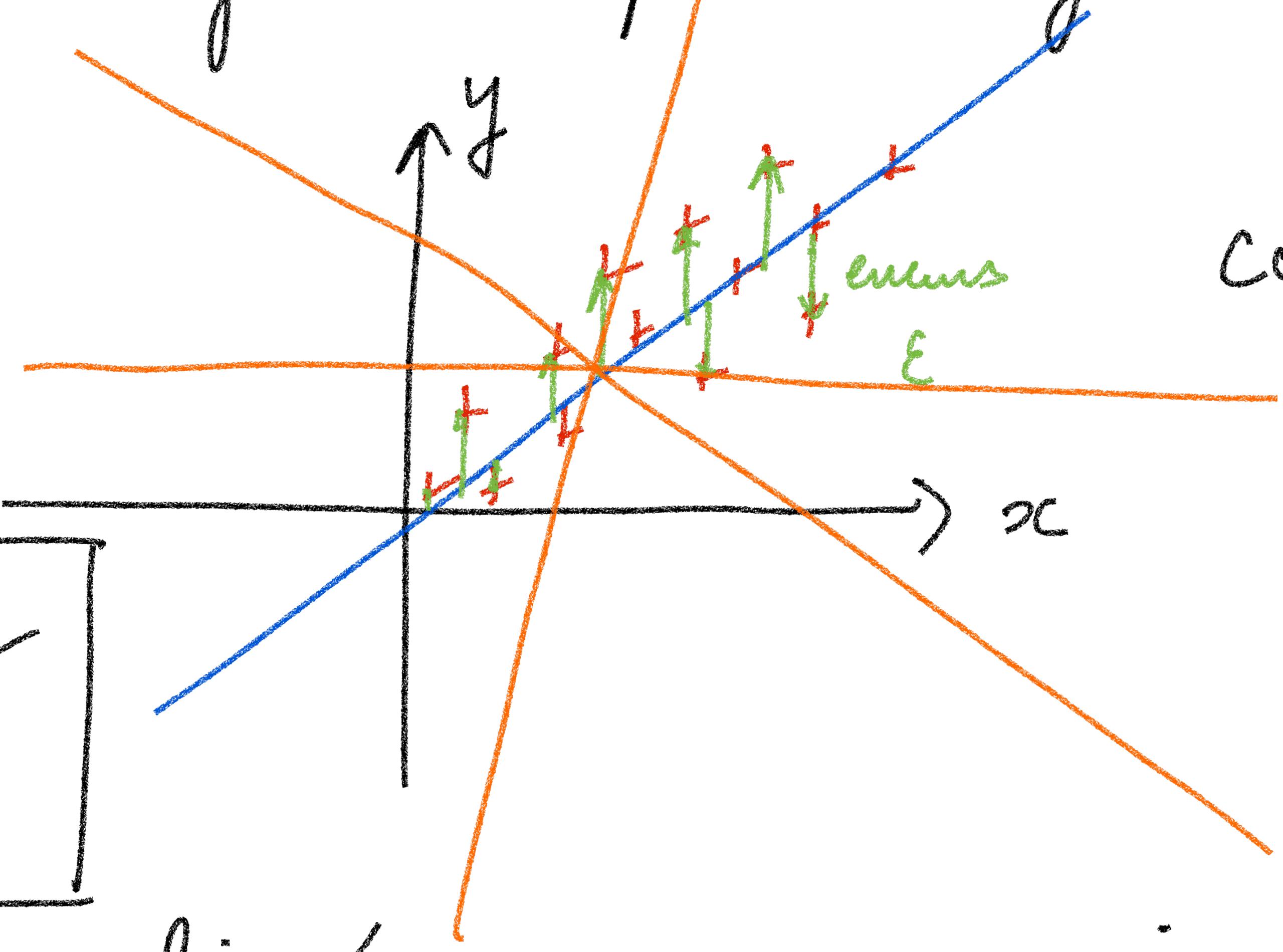
$$y = a x + b$$

"slope"

"intercept"



je veux prédire y à partir de x .



$$\text{cor}(x, y) \geq 1$$

$$y = a(x) + b \\ + \epsilon$$

régression linéaire avec moindres carrés
 $\min \sum_{i=1}^n \epsilon_i^2$

y	x_1	x_2	\dots	x_p
.
.
.
.

$$y = a_1 x_1 + a_2 x_2 + \dots + a_p x_p + b \quad ("intercept") + \epsilon$$

Hypothèses :

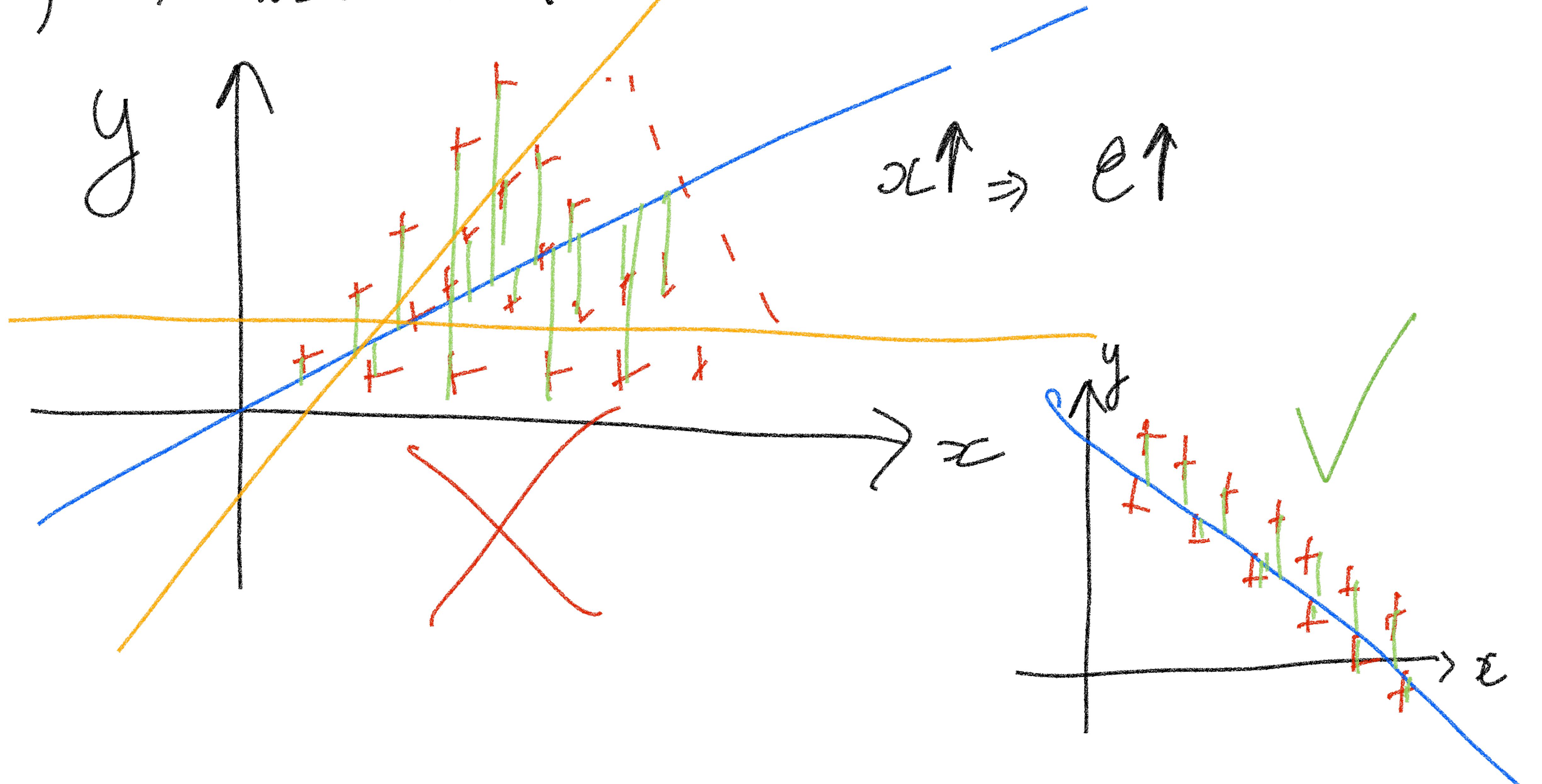
1): Pas de colinearité entre les variables prédictives.

Ex): Si

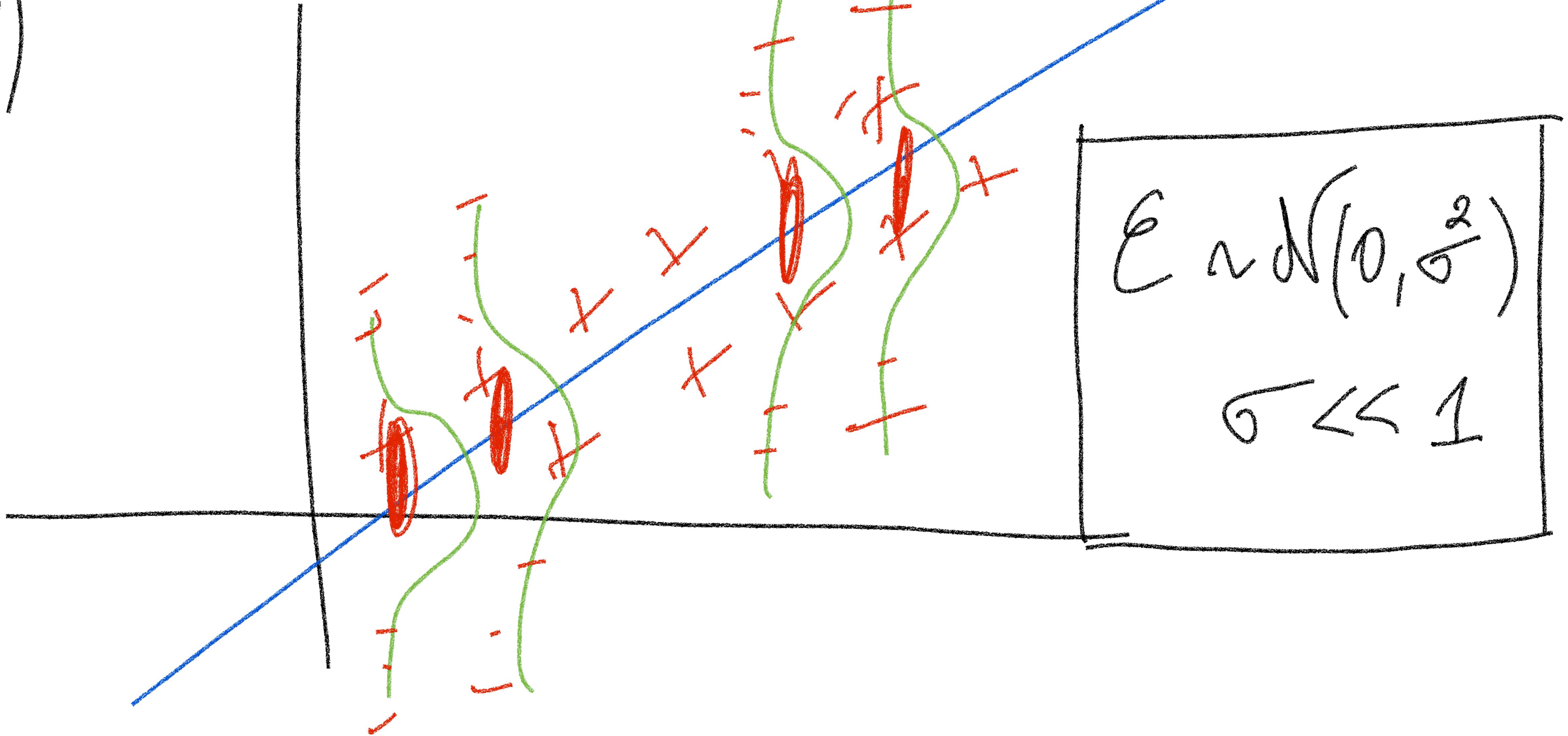
Prix	Prix	il faut
€	\$	/ diviser une
		des 2
		colonnes

$\text{cor}(x_A, x_B) \geq \pm 1$

2) Homoscedashicité



3)



$$y = a_1 x_1 + a_2 x_2 + \dots + a_p x_p + \epsilon$$

a_1, \dots, a_p : poids / pente
de la relation entre

y et x_i

. $R^2 \in [0, 1]$ 1 : parfait
 0 : marche pas du tout

R^2 : "la part de la variance expliquée par le modèle"

$$R^2 = (\text{coefficent de corrélation de Pearson})^2$$

• RMSE = Root Mean Square Error

$$= \sqrt{\frac{1}{N} \sum_{i=1}^N e_i^2}$$

Si y est
en mètres,

RMSE est en
mètres.

$$= \sqrt{\frac{1}{N} \sum_{i=1}^N (\text{prediction} - \text{réalité})^2}$$

$$\text{Anse} = 100 \text{ $}$$

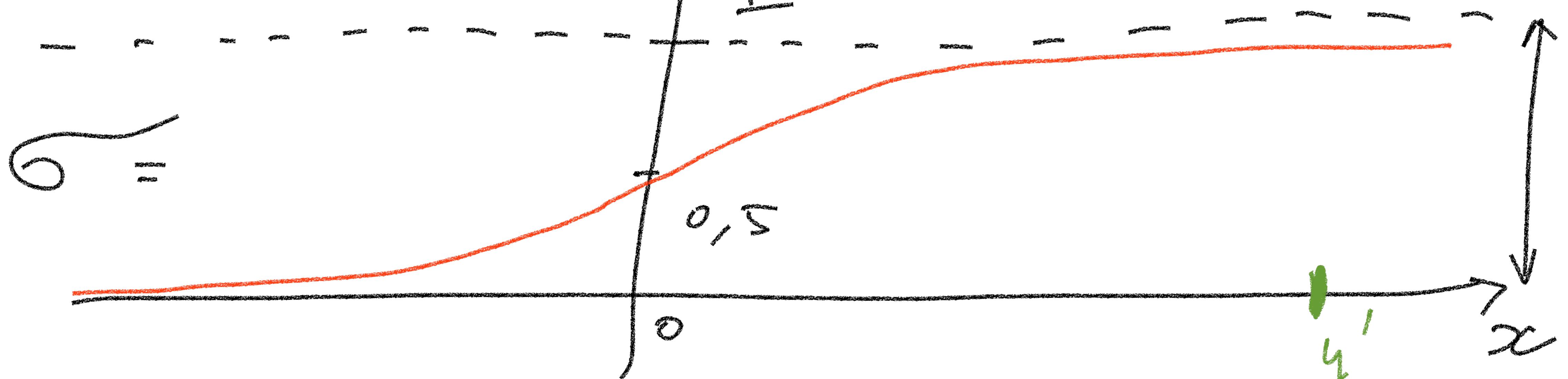
y : Var numérique

(x_i) : Var numériques

y : Var catégorielle (binaire,
0 ou 1)

Régression Logistique

Reg Log = Reg Lin + Sigmoid



$$y' = a_1 x_1 + \dots + a_p x_p + b$$

$$y_{\text{pred}} = \sigma(y') \in [0, 1]$$

Si $y_{\text{pred}} > 0,5$,
 $y_{\text{pred}} = 1$

One VS All Classification

ex: 3 classes : $\{A, B, C\}$

$\{B, C, A\}$

$\{C, A, B\}$