Least Squares

Computational Photography (CSCI 3240U) & Computer Vision (CSCI 4220U)

Faisal Z. Qureshi
http://vclab.science.ontariotechu.ca
OntarioTech UNIVERSITY

Find accompanying notes at

https://csundergrad.science.uoit.ca/courses/cv-notes/notebooks/15-least-squares.html

https://csundergrad.science.uoit.ca/courses/cv-notes/notebooks/16-robust-least-squares.html

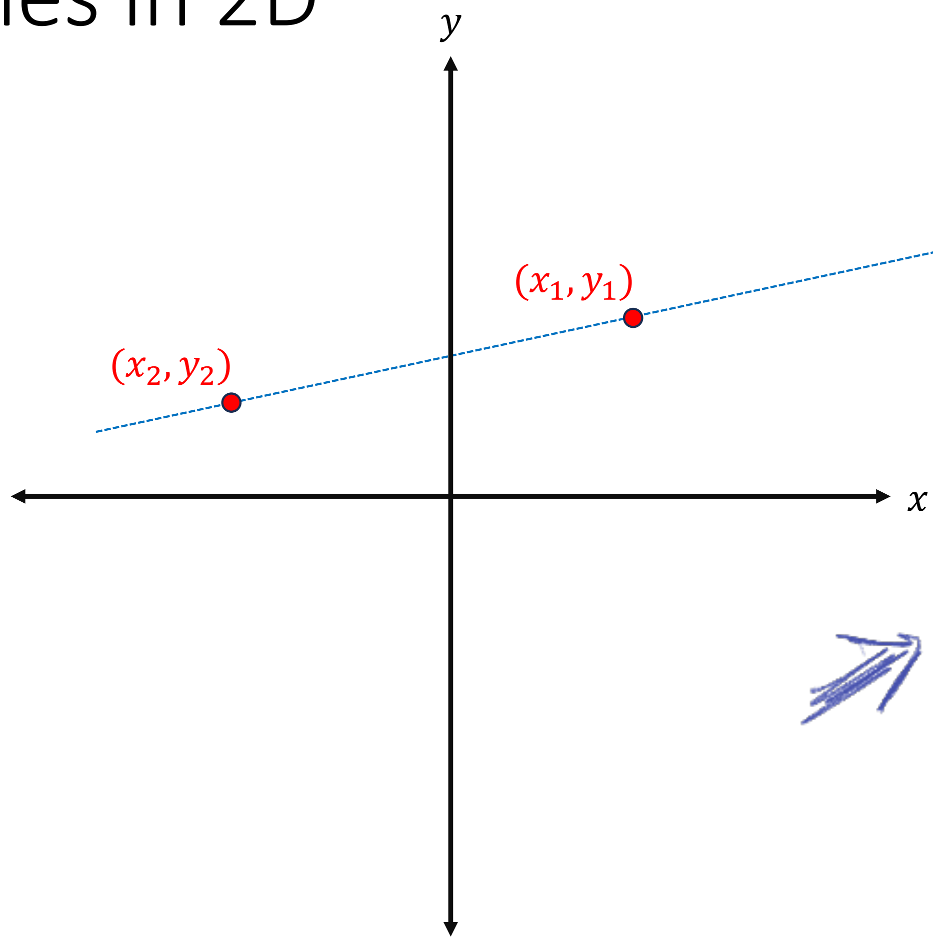
https://csundergrad.science.uoit.ca/courses/cv-notes/notebooks/17-ransac.html

https://csgrad.science.uoit.ca/courses/ist/notebooks/linear-regression.html


401
Broadway
SCOOP
NYC
MO
CA
Pearl River
CRYSTAL
"BEHIND A ROLLING
BALL COMES A
RUNNING CHILD"
Please Drive Safely
2614
HOW IS MY DRIVING?
1-800-327-4677
LOOMIS
XA-932Z
x27
x28

Lines in 2D

Lines in 2D



output

input

Equation of a line in 2D

$$y = mx + c$$

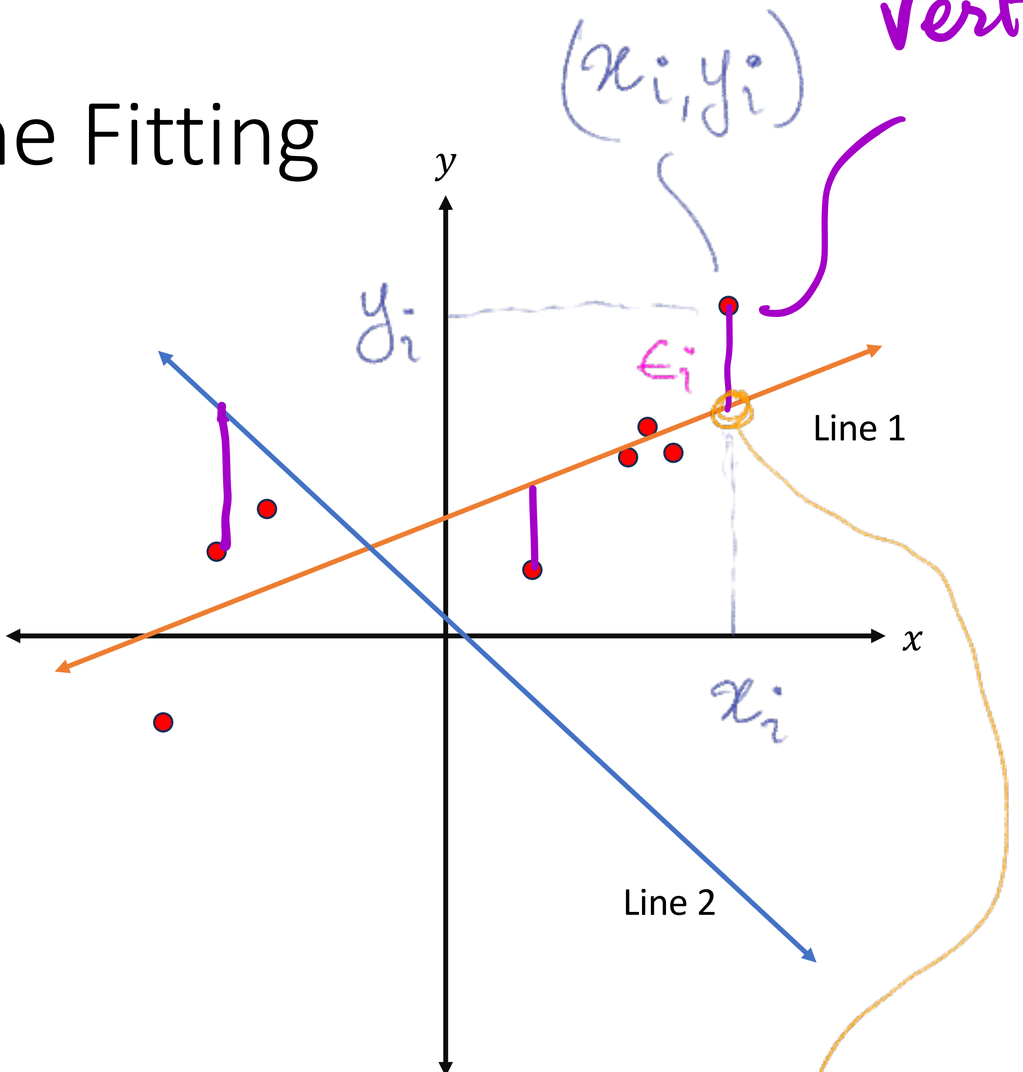
parameters

$$\frac{y - y_1}{y_2 - y_1} = \frac{x - x_1}{x_2 - x_1}$$

$$\begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} \times \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

Line Fitting

vertical distances

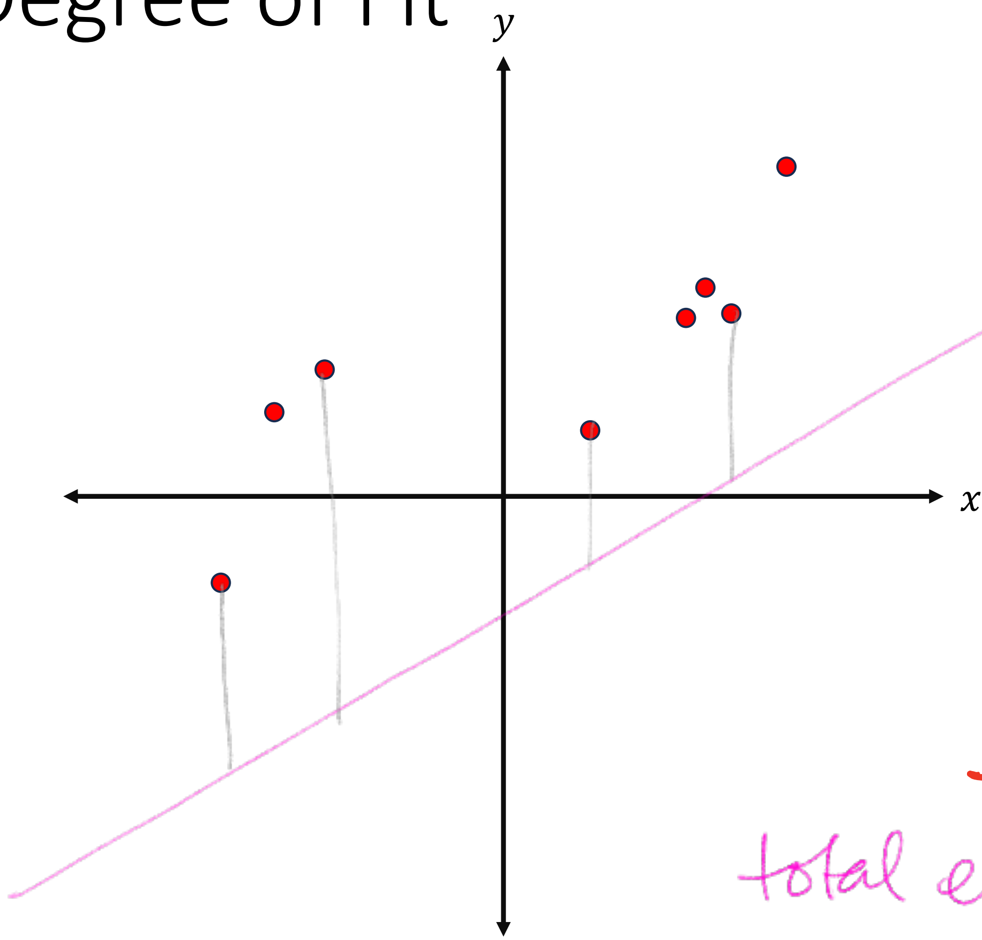


Which of the two lines better represents the data?

Line 1 seems to represent data better than Line 2.

$$\epsilon_i = \left(y_i - (m x_i + b)\right)^2$$

Degree of Fit



Mean Squared Error (MSE)

lower value → better fit

m, b

$$\epsilon_i = (y_i - m x_i - b)^2$$

total error = $\sum_i \epsilon_i$

Least Squares Fitting (2D Lines)

Given

$$(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$$

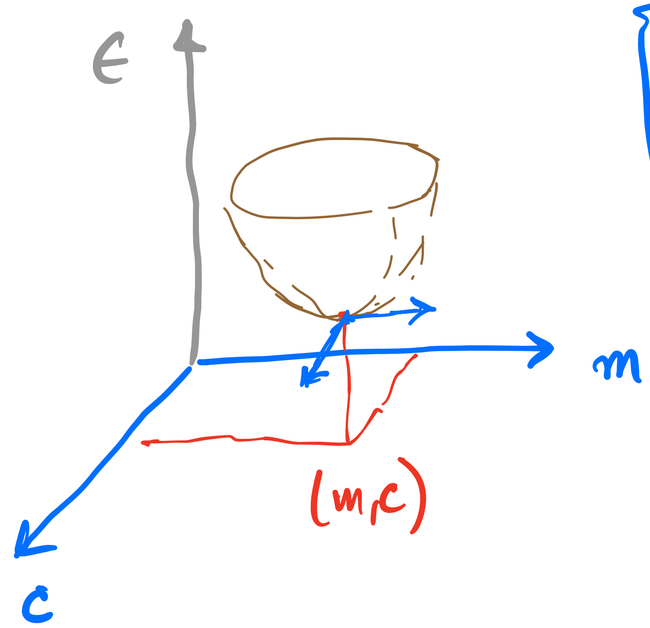
Estimate m and b to setup the 2D line

$$y = mx + c$$

By minimizing MSE

$$\epsilon = \frac{1}{n}\sum_{i=1}^{n}\Big(\underbrace{y_i}_{\text{Ground truth}} - \underbrace{(m\, x_i + c_i)}_{\text{Model prediction}}\Big)^2$$

$$m^*, c^* = \arg\min_{m,c} \epsilon$$



$$\begin{bmatrix} \frac{\partial \epsilon}{\partial m} \\ \frac{\partial \epsilon}{\partial c} \end{bmatrix} = \nabla \epsilon$$

At location (m,c) where ϵ is minimum,

$$\frac{\partial \epsilon}{\partial c} = 0 \quad \text{and} \quad \frac{\partial \epsilon}{\partial m} = 0$$

$$\epsilon = \frac{1}{n} \sum_{i=1}^{n} (y_i - m x_i - c)^2$$

$$\frac{\partial \epsilon}{\partial m} = \frac{1}{n} \sum_{i=1}^{n} -2x_i (y_i - m x_i - c)$$

$$\frac{\partial \epsilon}{\partial c} = \frac{1}{n} \sum_{i=1}^{n} -2 (y_i - m x_i - c)$$

$$-\frac{2}{n}\sum_{i=1}^{n} x_i(y_i - m x_i - c) = 0 \qquad — \; 2$$

$$-\frac{2}{n}\sum_{i=1}^{n} (y_i - m x_i - c) = 0 \qquad — \; 1$$

1

$$\frac{1}{n}\sum_{i=1}^{n} (y_i - m x_i - c) = 0$$

$$\Rightarrow \underline{\frac{1}{n}\sum_{i=1}^{n} y_i} - \underline{\frac{m}{n}\sum_{i=1}^{n} x_i} - \underline{\frac{1}{n}\sum_{i=1}^{n} c} = 0$$

$$\Rightarrow \underline{\langle y\rangle - m\langle x\rangle} - c = 0$$

2

$$\frac{1}{n}\sum_{i=1}^{n} x_i(y_i - m x_i - c) = 0$$

$$\Rightarrow \frac{1}{n}\sum_{i=1}^{n} x_i(y_i - m x_i \underline{- \langle y\rangle + m\langle x\rangle}) = 0$$

$$\Rightarrow \frac{1}{n}\sum_{i=1}^{n} x_i\left(y_i - \langle y\rangle - m(x_i - \langle x\rangle)\right) = 0$$

$$\Rightarrow \frac{1}{n}\sum_{i=1}^{n} x_i(y_i - \langle y\rangle) - \frac{1}{n}\sum_{i=1}^{n} m x_i(x_i - \langle x\rangle) = 0$$

$$\Rightarrow m = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i(y_i - \langle y\rangle)}{\frac{1}{n}\sum_{i=1}^{n} x_i(x_i - \langle x\rangle)}$$

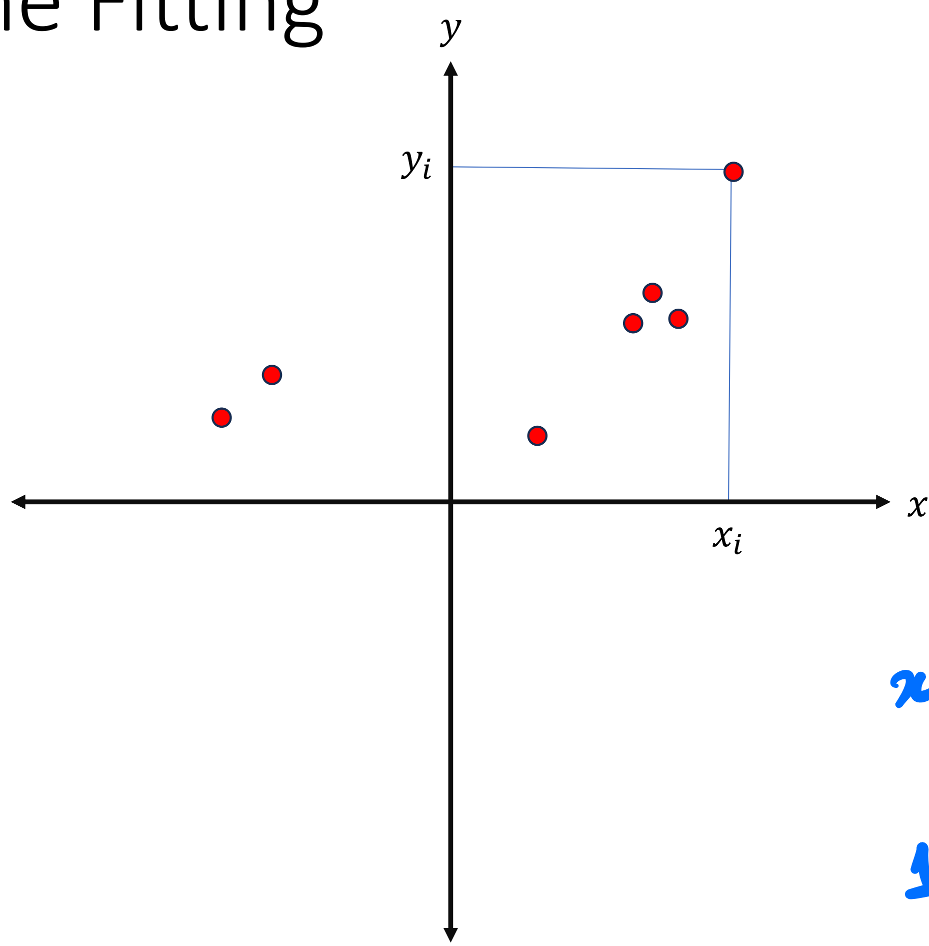
$$\Rightarrow \quad m = \frac{\sum_{i=1}^{n} x_i (y_i - \langle y \rangle)}{\sum_{i=1}^{n} x_i (x_i - \langle x \rangle)}$$

Compute c.

Linear Regression

Going beyond 2D

Line Fitting



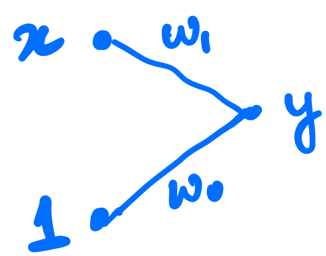
1D features: x

Ground truth: y

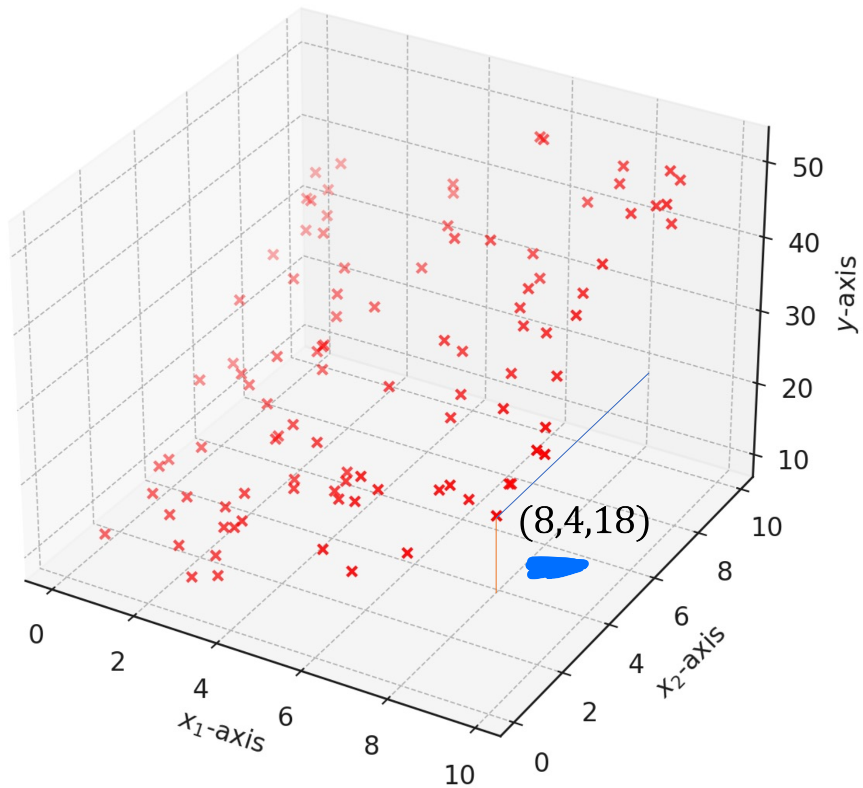
Model

$$y = mx + c \qquad \text{(Line)}$$

$$y = w_1 x + w_0 \qquad \text{(Re-write)}$$



2D Features



2D features: (x_1, x_2)

Ground truth: y

Model

parametere

$$y = w_2 x_2 + w_1 x_1 + w_0$$

feature 1

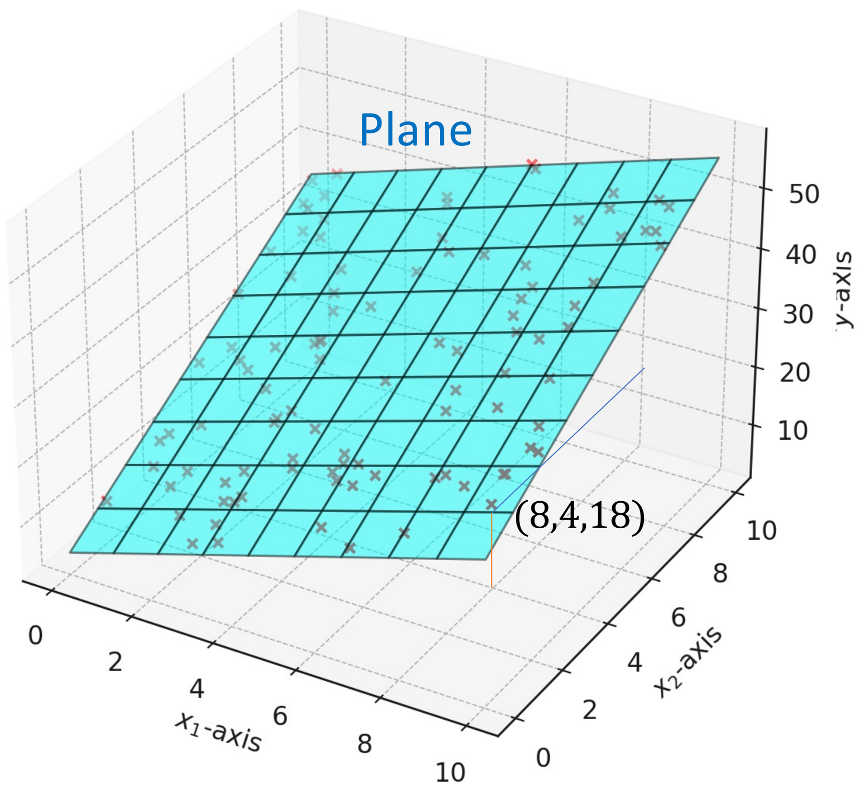
feature 2

$$\epsilon_i = \left(y_i - \left(w_2 x_{2i} + w_1 x_{1i} + w_0\right)\right)^2$$

? vector form

$$\epsilon = \frac{1}{n}\sum_i \epsilon_i$$

2D Features



2D features: (x_1, x_2)

Ground truth: y

Model

$$y = w_2 x_2 + w_1 x_1 + w_0$$

Features.

Linear Models in Higher Dimensions

$$y = w_d x_d + \cdots + w_2 x_2 + w_1 x_1 + w_0$$

parameters.

$$= \begin{bmatrix} w_d & \cdots & w_2 & w_1 & w_0 \end{bmatrix} \begin{bmatrix} x_d \\ \vdots \\ x_2 \\ x_1 \\ 1 \end{bmatrix}$$

not going to help us.

$$y_1 = \begin{bmatrix} x_{d_1} & \cdots & x_{21} & x_{12} & 1 \end{bmatrix}$$

$$y_2 = \begin{bmatrix} x_{d_2} & \cdots & x_{22} & x_{12} & 1 \end{bmatrix} \begin{bmatrix} w_d \\ \vdots \\ w_2 \\ w_1 \\ w_0 \end{bmatrix}$$

ground truth

prediction

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{d_1} & \cdots & x_{21} & x_{11} & 1 \\ x_{d_2} & \cdots & x_{22} & x_{21} & 1 \\ \vdots & & \vdots & \vdots & \vdots \\ x_{d_n} & \cdots & x_{2n} & x_{1n} & 1 \end{bmatrix} \begin{bmatrix} w_d \\ \vdots \\ w_2 \\ w_1 \\ w_0 \end{bmatrix}$$

$$\in \mathbb{R}^{n\times 1} \qquad \in \mathbb{R}^{n\times(d+1)} \qquad \in \mathbb{R}^{(d+1)\times 1}$$

$$Y \qquad X \qquad \Theta$$

Model:

ground truth — predictions

$$Y = \underline{X\Theta}$$

Estimate

Error:

$$\epsilon = (X\Theta - Y)^T (X\Theta - Y)$$

$$X\Theta = w_d x_d + \cdots + w_2 x_2 + w_1 x_1 + w_0$$

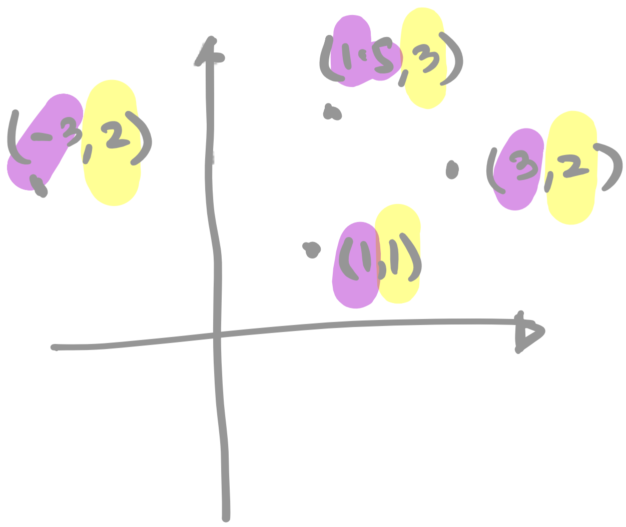
$$Y = y$$

$$(X\Theta - Y) = (w_d x_d + \cdots + w_2 x_2 + w_1 x_1 + w_0) -$$

$$(X\theta - Y) = \begin{bmatrix}\epsilon_1\\ \epsilon_2 \\ \vdots \\ \epsilon_n\end{bmatrix} \in \mathbb{R}^{n\times 1}$$

$$(X\theta - Y)^T (X\theta - Y) = \epsilon_1^2 + \epsilon_2^2 + \cdots + \epsilon_n^2$$

$$= \sum_{i=1}^{n} \epsilon_i^2$$



$$Y = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 2 \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & 1 \\ 3 & 1 \\ 1.5 & 1 \\ -3 & 1 \end{bmatrix}$$

$$\theta = \begin{bmatrix} m \\ b \end{bmatrix}$$

$$X\theta = \begin{bmatrix} (1)(m) + (1)(b) \\ (3)(m) + (1)(b) \\ (1.5)(m) + (1)(b) \\ (-3)(m) + (1)(b) \end{bmatrix}$$

$$\epsilon = (X\theta - Y)^T (X\theta - Y)$$

$$\boxed{\frac{\partial \epsilon}{\partial \theta} = \emptyset}\ ; \text{ Solve for } \theta$$

?

y

Linear Models in Higher Dimensions

$$y = w_n x_n + \cdots + w_2 x_2 + w_1 x_1 + w_0$$

2D line fitting is a special case

$$y = \underbrace{\cancel{w_n x_n + \cdots + w_2 x_2}}_{0} + w_1 x_1 + w_0$$

$$y = w_1 x_1 + w_0$$

$$y = mx + c$$

Linear Regression

Given

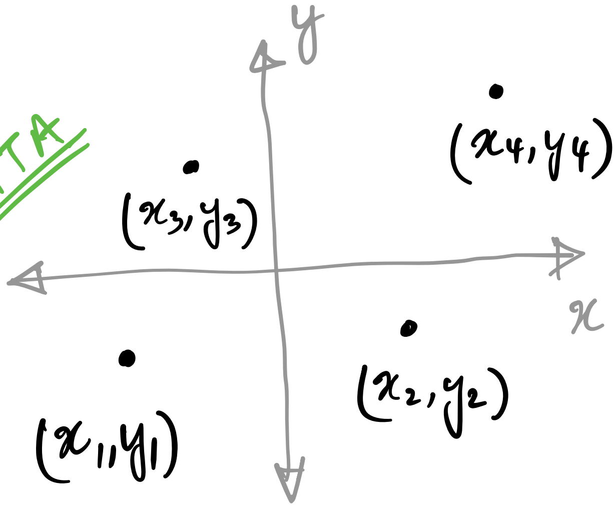
$$(\vec{x}_1\,, y), (\vec{x}_2\,, y), \cdots, (\vec{x}_n\,, y) \quad \text{where} \quad \vec{x}_i \in \mathbb{R}^d$$

Fit

$$y = w_d x_d + \cdots + w_2 x_2 + w_1 x_1 + w_0$$

Feb 14, 2025

DATA



Model

$f : \mathbb{R} \to \mathbb{R}$

$: x \to y$

$$y = mx + c$$
$$= \theta_1 x + \theta_0$$
$$= X\theta$$

design matrix

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_0 \end{bmatrix}$$

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \hat{y}_4 \end{bmatrix} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \\ x_4 & 1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_0 \end{bmatrix}$$

$X \qquad \theta$

predictions $(\hat{Y})$

Q. What is required to fit a model?

A. Minimize the $\underbrace{\text{fit error}}_{\text{loss}}$.

Example: MSE

$$\epsilon = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

ground truth

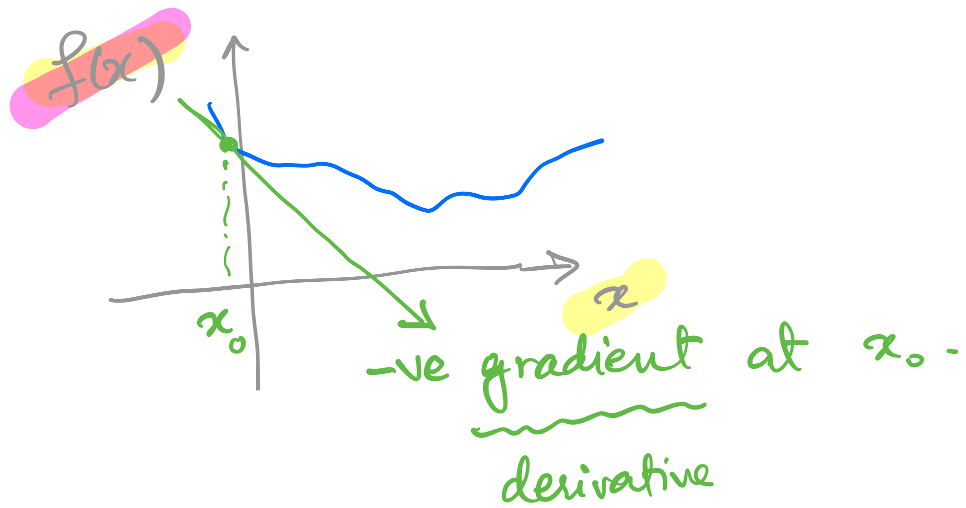
Using $\hat{Y} = X\theta$, $\longrightarrow$ $\underline{\underline{A}}\,\underline{\underline{\vec{x}}} = \underline{\underline{\vec{b}}}$

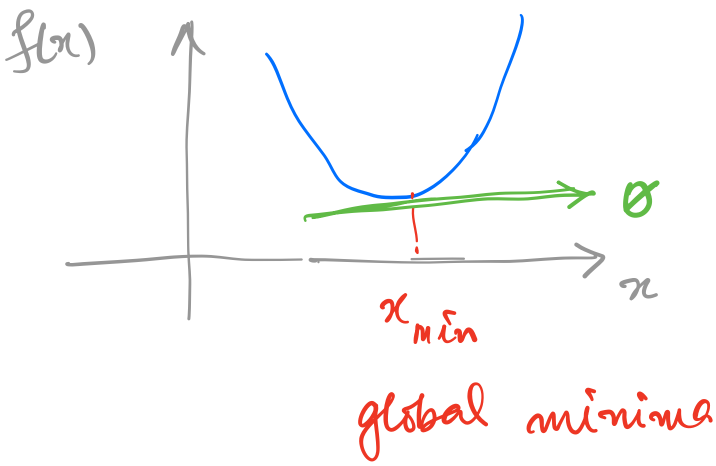
$X \quad \theta \quad Y$

$$e = (Y - X\theta)^T (Y - X\theta)$$

predictions

Q. How do I minimize the error?





SOLVE

$$\nabla_\theta \; (Y - X\theta)^T (Y - X\theta)$$

parameters.

$$= \nabla_\theta \; Y^TY - Y^TX\theta - (X\theta)^TY + (X\theta)^TX\theta$$

$$= \nabla_\theta \; \cancel{Y^TY} + \theta^TX^TX\theta - Y^TX\theta - (X\theta)^TY$$

① real number.

② $\operatorname{tr} a = a, \quad a \in \mathbb{R}$

$$= \nabla_\theta \operatorname{tr}\left(\cancel{Y^T Y}^{\,0} + \theta^T X^T X \theta - Y^T X \theta - \underline{\theta^T X^T Y}\right)$$

$$= \nabla_\theta \operatorname{tr}\left(\theta^T X^T X \theta - \theta^T X^T Y - Y^T X \theta\right)$$

$$= \nabla_\theta \operatorname{tr} \theta^T X^T X \theta - \nabla_\theta \operatorname{tr} \underline{\theta^T X^T Y} - \nabla_\theta \operatorname{tr} Y^T X \theta$$

real number

$$(\theta^T X^T Y)^T = \theta^T X^T Y$$

$$\Rightarrow Y^T X \theta$$

$$= \nabla_\theta \operatorname{tr} \theta^T X^T X \theta - \underline{\nabla_\theta \operatorname{tr} Y^T X \theta - \nabla_\theta \operatorname{tr} Y^T X \theta}$$

$$= \underset{①}{\nabla_\theta \operatorname{tr} \theta^T X^T X \theta} - \underset{②}{2 \nabla_\theta \operatorname{tr} Y^T X \theta}$$

② $\nabla_A \operatorname{tr} AB = B^T$

$\operatorname{tr} AB = \operatorname{tr} BA$

$\operatorname{tr} ABC = \operatorname{tr} CAB$

② $\nabla_\theta \operatorname{tr} Y^T X \theta$

$$= \nabla_{\underbrace{\theta}_{A}} \operatorname{tr} \underbrace{\theta}_{A} \underbrace{Y^T X}_{B}$$

$$= (Y^T X)^T$$

$$= X^T Y$$

$$\nabla_A \operatorname{tr} ABA^TC = CAB + C^TAB^T$$

---

① $\nabla_\theta \operatorname{tr} \theta^T X^T X \theta$

$$= \nabla_\theta \operatorname{tr} \theta \theta^T X^T X$$

$$= \nabla_{\underbrace{\theta}_{A}} \operatorname{tr} \underbrace{\theta}_{A} \underbrace{I}_{B} \underbrace{\theta^T}_{A^T} \underbrace{X^T X}_{C}$$

$$= X^T X \theta I + (X^T X)^T \theta I^T$$

$$= X^T X \theta + \underbrace{(X^T X)^T}_{\text{square symmetric matrix}} \theta$$

square symmetric matrix

$$(X^T X)^T = X^T X$$

$$= X^T X \theta + X^T X \theta$$

$$= 2X^TX\theta$$

$$\nabla_\theta \epsilon = 2X^TX\theta - 2X^TY$$

set it to $\emptyset$

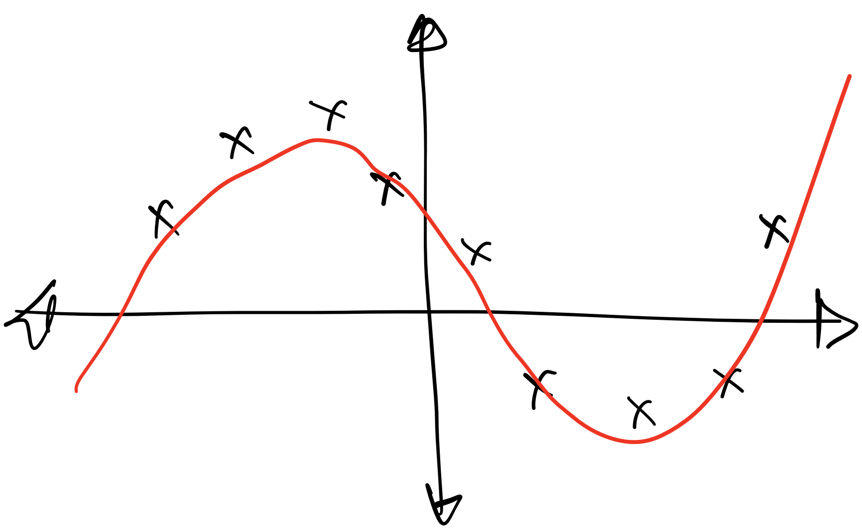
$$\cancel{2}X^TX\theta - \cancel{2}X^TY = 0$$

$$\Rightarrow X^TX\theta - X^TY = 0$$

$$\Rightarrow X^TX\theta = X^TY$$

$$\Rightarrow \theta = \underbrace{(X^TX)^{-1}X^T}_{\text{pseudo-inverse}}Y.$$

Normal Equation.



$$y = x^3 w_3 + x^2 w_2 + x w_1 + w_0$$

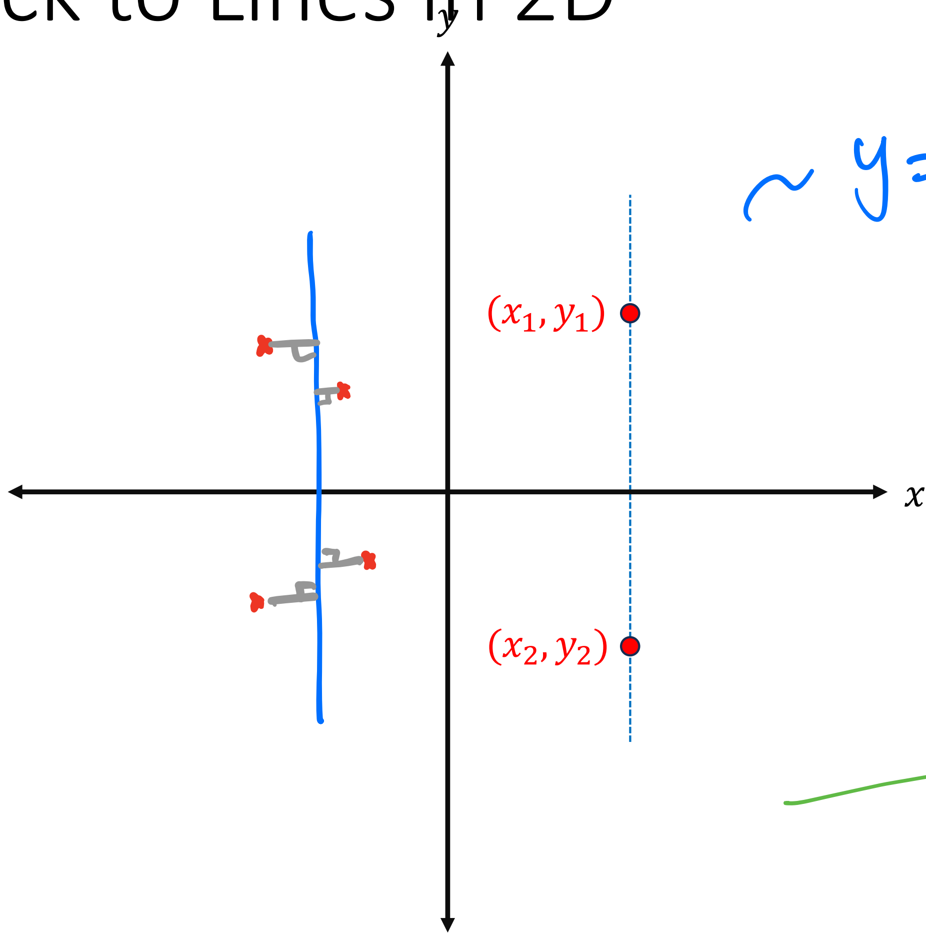
$$\operatorname{tr}\begin{bmatrix} a & & & \xi \\ & b & & \\ \xi & & c & \\ & & & d \end{bmatrix} = a+b+c+d$$

$$\underbrace{\hspace{10em}}_{\text{real number}}$$

Total least squares

As opposed to ordinary least squares

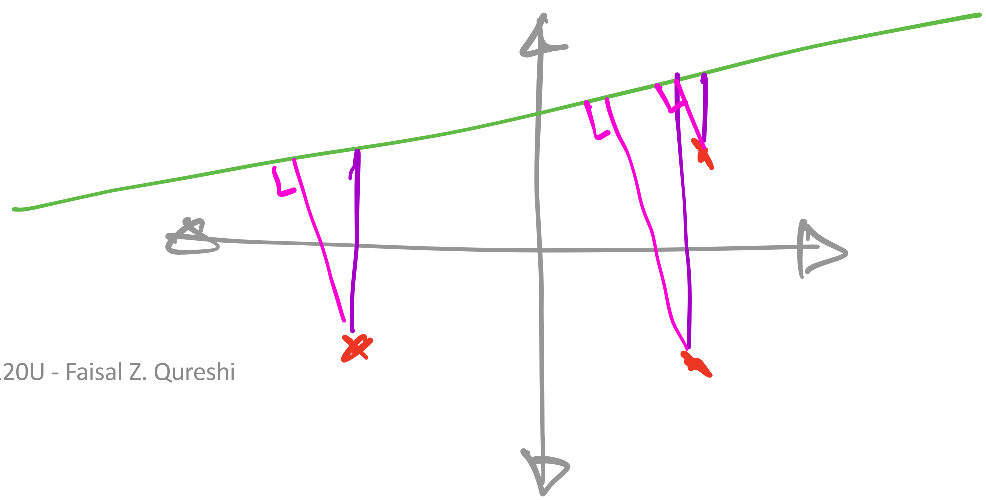
Back to Lines in 2D



$\sim y = mx + c$

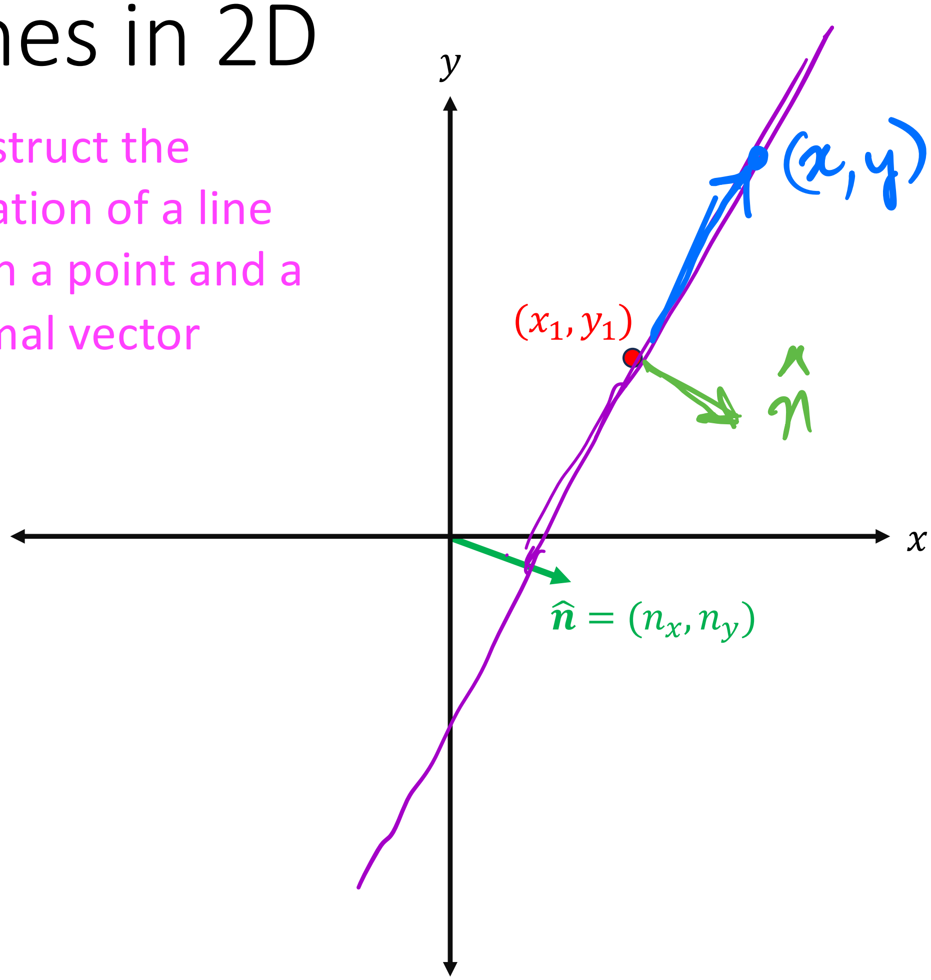
$\downarrow$

infinite (undefined)



Lines in 2D

Construct the equation of a line given a point and a normal vector



$$(x - x_1, y - y_1) \perp \hat{n}$$

$$(x - x_1, y - y_1) \cdot (n_x, n_y) = 0$$

$$\Rightarrow (x - x_1) n_x + (y - y_1) n_y = 0$$

$$\Rightarrow n_x x - x_1 n_x + n_y y - y_1 n_x = 0$$

$$\Rightarrow \underbrace{n_x}_{a} x + \underbrace{n_y}_{b} y \underbrace{- x_1 n_x - y_1 n_y}_{c} = 0$$

$$ax + by + c = 0$$

$$A\vec{x} = 0$$

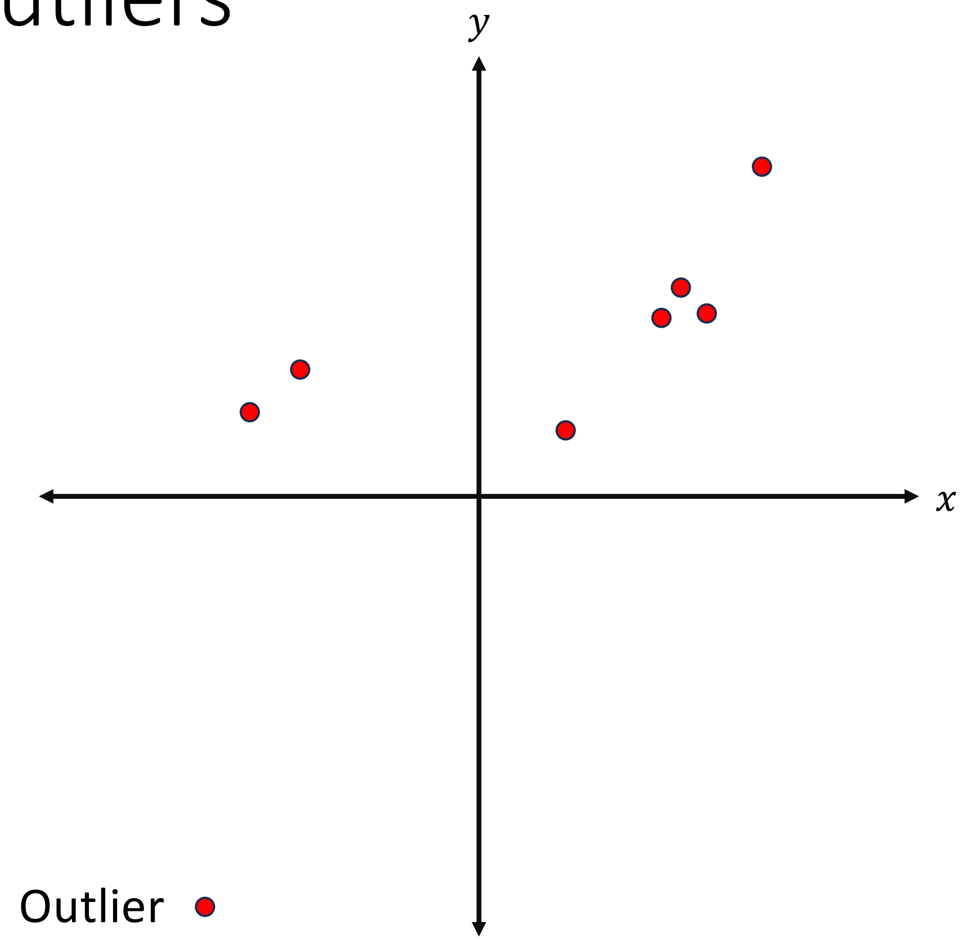
Total Least Squares vs Ordinary Least Squares

Method	Errors in X	Errors in Y	Distance minimized
Ordinary least squares	No	Yes	Vertical
Total least squares	Yes	Yes	Perpendicular

Model fitting

Issues

Outliers

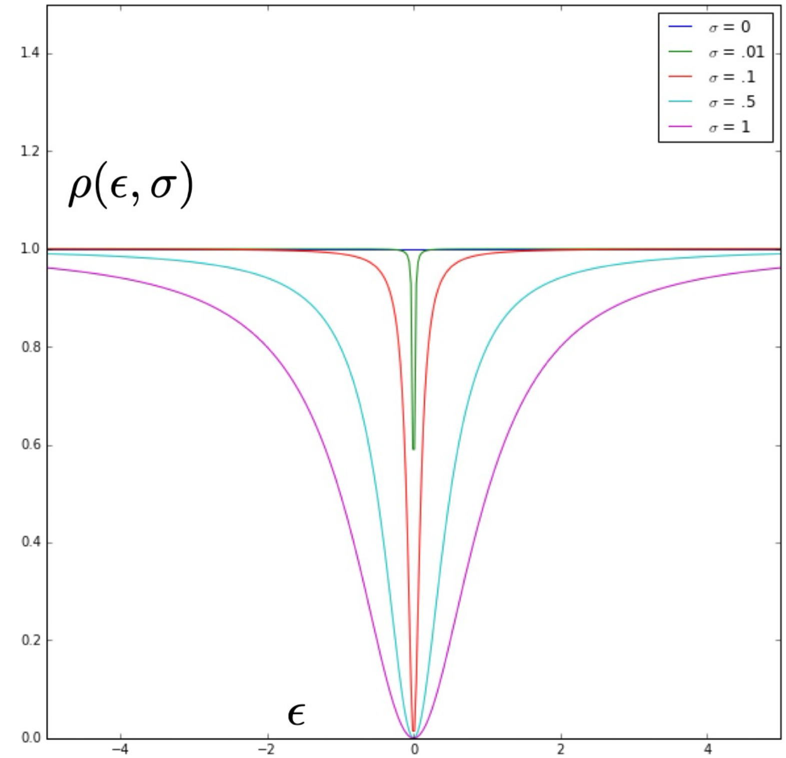


Least squares estimate is sensitive to outliers

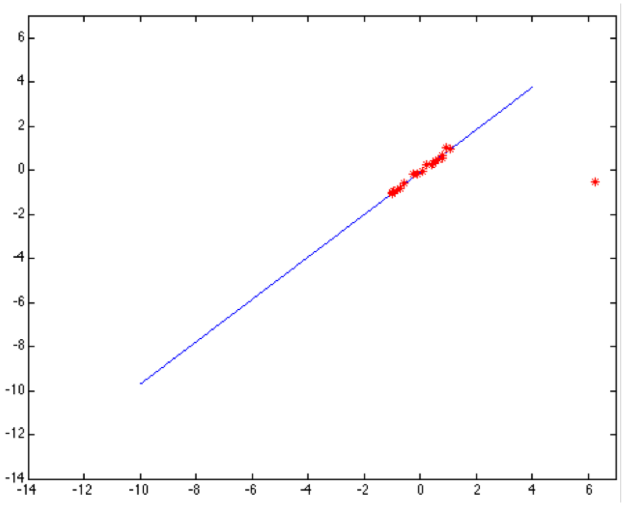
Robust Least Squares

$$\rho(\epsilon, \sigma) = \frac{\epsilon^2}{\epsilon^2 + \sigma^2}$$

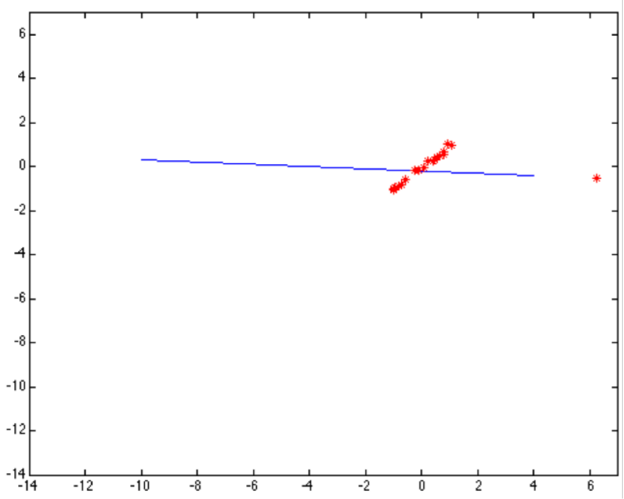
Scale parameter



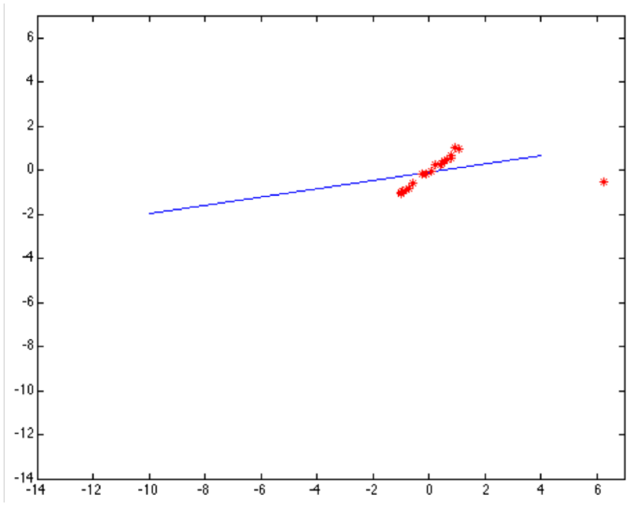
Robust Least Squares



When scale is just right, the effects of outliers are eliminated



When scale is too small, the error value for almost all points is the same. This results in a poor fit.



When scale is too large, the system behaves like good old least squares, i.e., it remains sensitive to outliers.

Incomplete data

Hidden data

Not available at the time of model fitting

Summary

- Line fitting in 2D
- Least squares
- How to deal with outliers?

Minimizing $\epsilon = \sum_{i=1}^{n} \left(y_i - (m\, x_i + c_i)\right)^2$

Re-write $\epsilon = \sum_{i=1}^{n}\left(y_i - (m\, x_i + c_i)\right)^2$ in Matrix Form