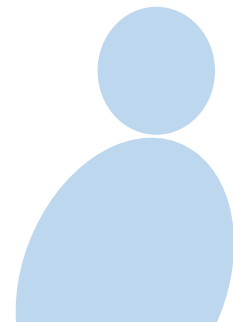


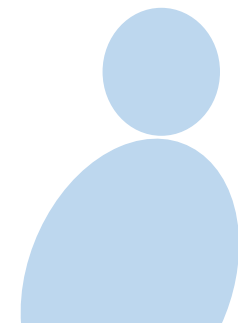
Elective Module:

**Advanced ML
Techniques**



Chapter 1

Linear Regression

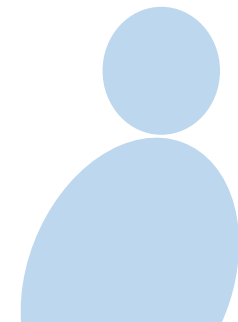


Linear Regression

- Consider the given data
- How to **predict** the SALES, as a function of ADVERTISING

Year	Sales (Million Euro)	Advertising (Million Euro)
1	651	23
2	762	26
3	856	30
4	1,063	34
5	1,190	43
6	1,298	48
7	1,421	52
8	1,440	57
9	1,518	58

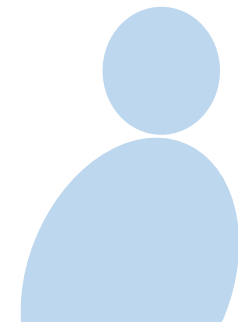
Time t Training Data



Linear Regression

- Consider the given data
- How to **predict** the **Sales**, as a function of **Advertising**

Year	Sales (Million Euro)	Advertising (Million Euro)
1	651	23
2	762	26
3	856	30
4	1,063	34
5	1,190	43
6	1,298	48
7	1,421	52
8	1,440	57
9	1,518	58



Linear Regression

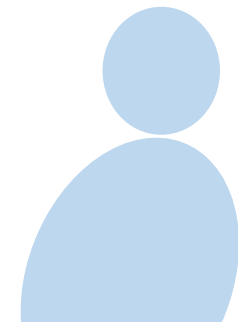
- REGRESSION is an ML technique...

Machine Learning

- which **precisely addresses** this problem

How to predict sales given advertising

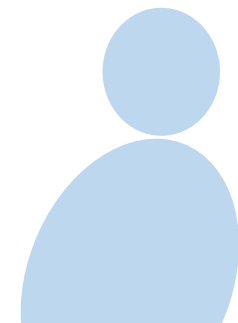
Year	Sales (Million Euro)	Advertising (Million Euro)
1	651	23
2	762	26
3	856	30
4	1,063	34
5	1,190	43
6	1,298	48
7	1,421	52
8	1,440	57
9	1,518	58



Linear Regression

- **Regression** is an ML technique...
 - which **precisely addresses** this problem

Year	Sales (Million Euro)	Advertising (Million Euro)
1	651	23
2	762	26
3	856	30
4	1,063	34
5	1,190	43
6	1,298	48
7	1,421	52
8	1,440	57
9	1,518	58

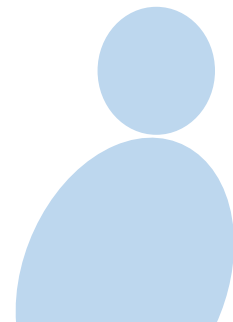


Linear Regression

- Regression: Algorithm to **predict a** RESPONSE variable...
- based on a set of REGRESSOR or EXPLANATORY VARIABLE.

Year	Sales (Million Euro)	Advertising (Million Euro)
1	651	23
2	762	26
3	856	30
4	1,063	34
5	1,190	43
6	1,298	48
7	1,421	52
8	1,440	57
9	1,518	58

Response



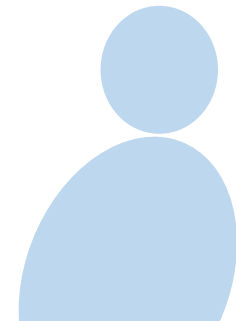
Linear Regression

- Regression: Algorithm to **predict a response** variable...
- based on a set of **regressors** or **explanatory variables**



Year	Sales (Million Euro)	Advertising (Million Euro)
1	651	23
2	762	26
3	856	30
4	1,063	34
5	1,190	43
6	1,298	48
7	1,421	52
8	1,440	57
9	1,518	58

Regressor



Linear Regression

- In general the **regressor** \bar{x} can be an n –dimensional vector
 - x_1 is the cost of **TV advertising**
 - x_2 is the cost of **Radio advertising** and so on....

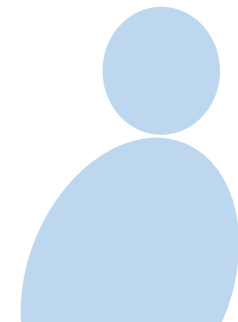
$$\bar{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}$$

Vector of Regressors

	TV	Radio	Newspaper	Sales
0	230.1	37.8	69.2	22.1
1	44.5	39.3	45.1	10.4
2	17.2	45.9	69.3	9.3
3	151.5	41.3	58.5	18.5
4	180.8	10.8	58.4	12.9

x_1
 x_2
 x_3 } Multiple Regressors.

Training instances.



Regression: Other Examples

• Example 1

- $y(k)$ = Price of particular stock at time k

- $x_1(k), x_2(k), \dots, x_N(k)$: Prices of related stocks at time k

$x(k)$ — Price at time k
using $x(k-1), x(k-2), \dots$

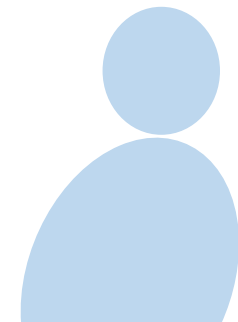
Past values.
AUTO REGRESSION

x_1 — Microsoft
 x_2 — Google
 x_3 — Amazon
...

} Regressors.

Facebook

Response
Time k



Regression: Other Examples

Response
Creta

- Example 2

- $y(k)$ = Sales of SUV at time k

- $x_1(k), x_2(k), \dots, x_N(k)$: Sales of **bikes/ cars** at time k , average income...

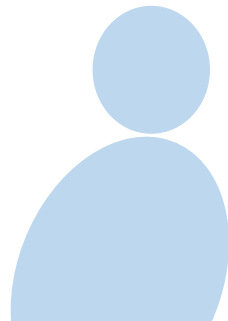
x_1 — Sale of Bikes.

x_2 — Sale of cars.

x_3 — Average income

\vdots

Regressors.



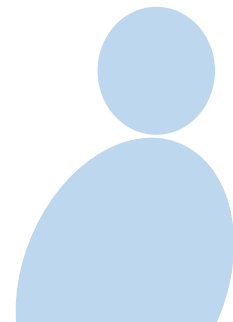
Linear Regression

- The outputs $y(k)$ can be **predicted...**
 - Using a linear combination of **regressors** or **explanatory variables** $x_i(k)$

Linear combination

$$y(k) = h_0 + h_1 x_1(k) + h_2 x_2(k) + \dots$$

Linear model.



Linear Regression Model

$$y(k) =$$

$$= \underbrace{h_0}_{\text{Bias: constant offset}} + h_1 x_1(k) + h_2 x_2(k) + \dots + h_n x_n(k) + \epsilon(k)$$

Response

$$= \underbrace{\begin{bmatrix} 1 & x_1(k) & x_2(k) & \dots & x_n(k) \end{bmatrix}}_{\bar{x}^T(k)} \underbrace{\begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_n \end{bmatrix}}_{\bar{h}} + \epsilon(k)$$

Noise Error

$$y(k) = \bar{x}^T(k) \bar{h} + \epsilon(k)$$

Linear Regression Model

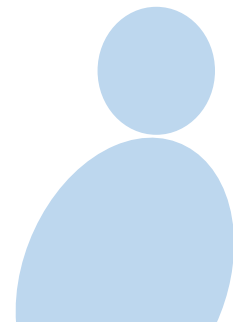
$$\bar{x}(k) = \begin{bmatrix} 1 \\ x_1(k) \\ x_2(k) \\ \vdots \\ x_n(k) \end{bmatrix}$$

$$\bar{h} = \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_n \end{bmatrix}$$

$$\bar{x}^T(k) = [1 \ x_1(k) \ x_2(k) \ \dots \ x_n(k)]$$

$$\underline{y(k) = \bar{x}^T(k) \bar{h} + \epsilon(k)}$$

Model for
Time k



Linear Regression Model

$$y(k)$$

$$= h_0 + h_1 x_1(k) + \cdots + h_n x_n(k) + \epsilon(k)$$

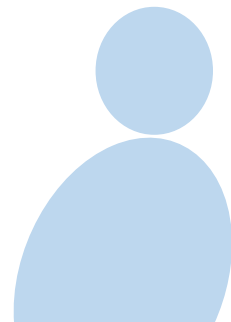
$$= \underbrace{[1 \quad x_1(k) \quad x_2(k) \quad \cdots \quad x_n(k)]}_{n+1} \underbrace{\begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_n \end{bmatrix}}_{\substack{n+1 \text{ dim vector} \\ \text{Regression} \\ \text{Coefficients.}}} + \epsilon(k)$$

model error

$$y(k) = \bar{\mathbf{x}}^T(k) \bar{\mathbf{h}} + \epsilon(k)$$

Bias

Response at
Time k .



Linear Regression

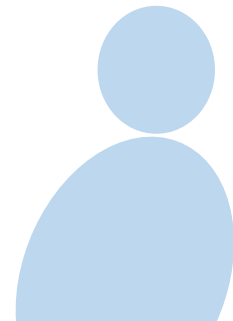
- This is termed Linear Regression

- w_0, h_1, \dots, h_n are termed the

Regression coefficients

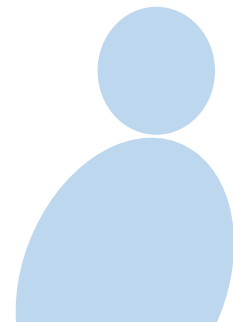
"Learn" Regression coefficients -

Machine Learning (ML).



Linear Regression

- This is termed ***Linear Regression***
- h_0, h_1, \dots, h_n are termed the **Regression coefficients**

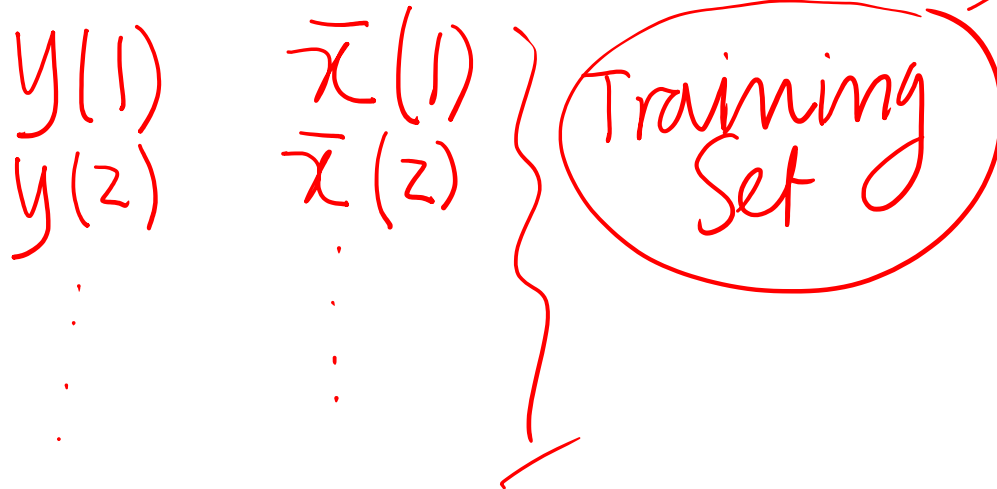


Training Data

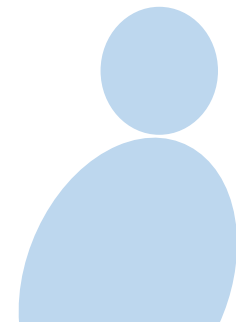
- The **regression coefficients** can be computed as follows

- Consider the availability of Training Pairs; $(y(k), \bar{x}(k))$

- for $k = 1, 2, \dots, M$



Very large Training Set.



Training Data

- The **regression coefficients** can be computed as follows

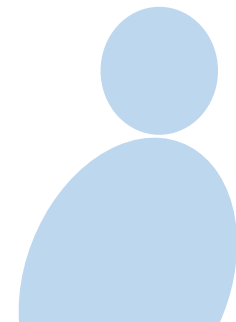
- Consider the availability of **training pairs** $(y(k), \bar{\mathbf{x}}(k))$

- for $k = 1, 2, \dots, M$

$$y(k) = \bar{\mathbf{x}}^T(k) \bar{\mathbf{h}} + \epsilon(k)$$

Training pairs = M .

Regression coefficients · error



Model Computation

$$\bar{y} = X \bar{h} + \bar{\epsilon}$$

- The **training set** can be expressed as

$$\underbrace{\begin{bmatrix} \bar{y} \\ y(1) \\ y(2) \\ \vdots \\ y(M) \end{bmatrix}}_{M \times 1} = \underbrace{\begin{bmatrix} \bar{x}^T(1) \\ \bar{x}^T(2) \\ \vdots \\ \bar{x}^T(M) \end{bmatrix}}_{\substack{X \\ M \times 1}} \bar{h} + \underbrace{\begin{bmatrix} \epsilon(1) \\ \epsilon(2) \\ \vdots \\ \epsilon(M) \end{bmatrix}}_{M \times 1}$$

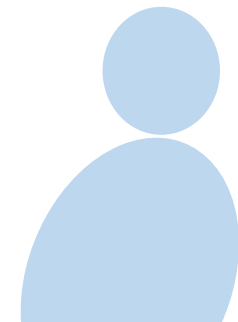
Model Computation

$$\bar{y} = X \bar{h} + \bar{\epsilon}$$

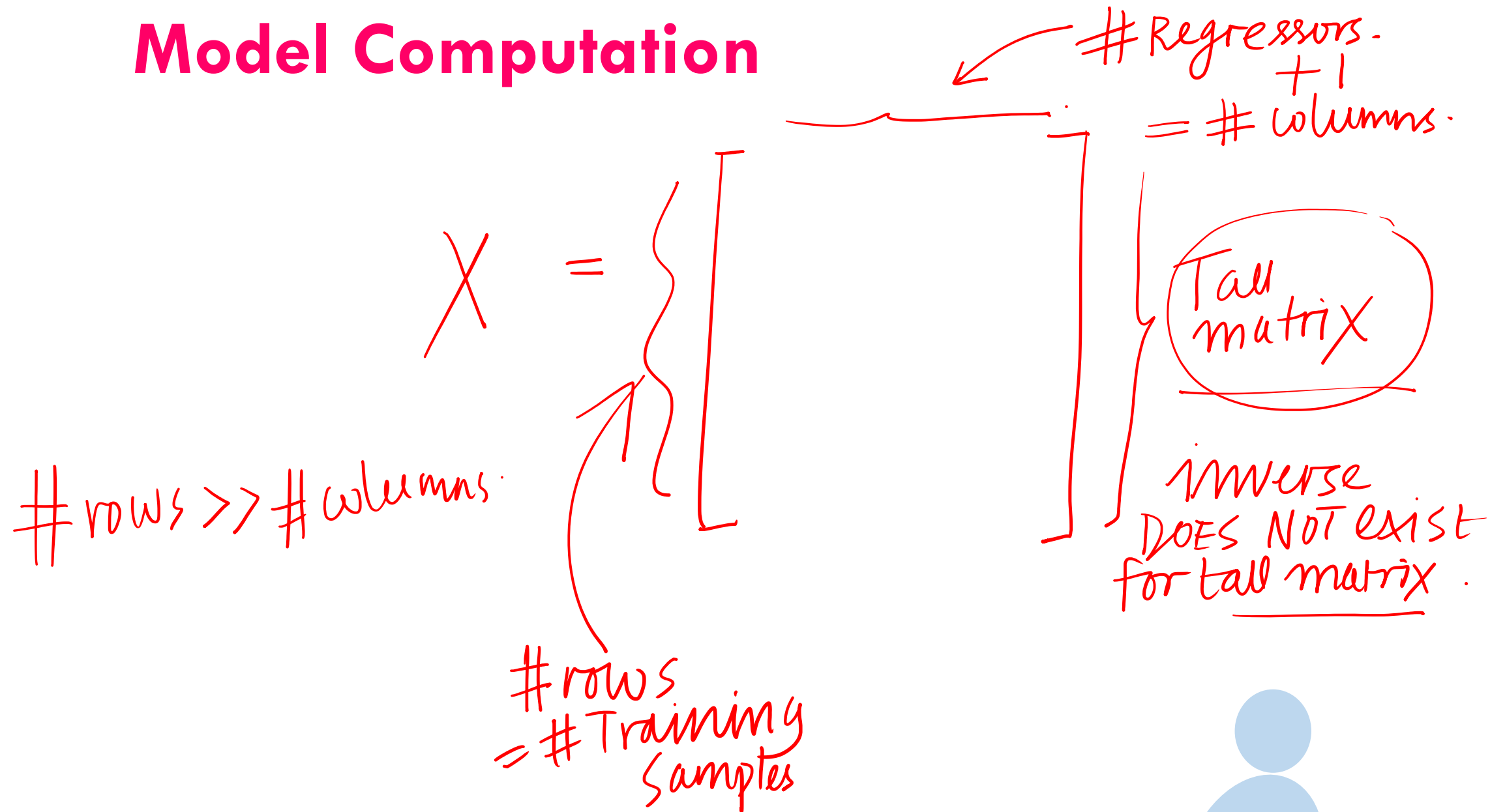
- The **training set** can be expressed as

$$\underbrace{\begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(M) \end{bmatrix}}_{\substack{\bar{y} \\ M \times 1}} = \underbrace{\begin{bmatrix} \bar{\mathbf{x}}^T(1) \\ \bar{\mathbf{x}}^T(2) \\ \vdots \\ \bar{\mathbf{x}}^T(M) \end{bmatrix}}_{\substack{\bar{\mathbf{X}} \\ M \times (n+1) \\ M \gg n+1}} \bar{\mathbf{h}} + \underbrace{\begin{bmatrix} \epsilon(1) \\ \epsilon(2) \\ \vdots \\ \epsilon(M) \end{bmatrix}}_{\substack{\bar{\epsilon} \\ M \times 1}}$$

How to determine \bar{h} ?



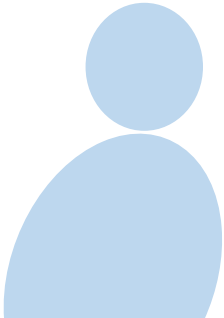
Model Computation



Model Computation

- To determine **regression coefficients** $\bar{\mathbf{h}}$, solve the problem

Find $\bar{\mathbf{h}}$ which gives best approximation to responses $\bar{\mathbf{y}}$.

$$\begin{aligned}\bar{\mathbf{e}} &= \bar{\mathbf{y}} - \mathbf{X}\bar{\mathbf{h}} \\ \text{min error} &= \min \|\bar{\mathbf{e}}\|^2 \\ &= \min \epsilon^2(1) + \epsilon^2(2) + \dots + \epsilon^2(M) \\ &= \min \|\bar{\mathbf{y}} - \mathbf{X}\bar{\mathbf{h}}\|^2\end{aligned}$$


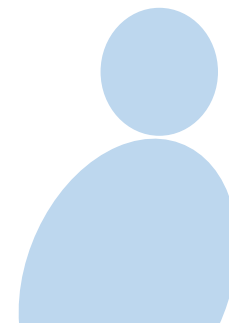
Model Computation

- To determine **regression coefficients** $\bar{\mathbf{h}}$, solve the problem

$$\min \left\| \underbrace{\bar{\mathbf{y}} - \mathbf{X} \bar{\mathbf{h}}}_{\bar{\boldsymbol{\epsilon}}} \right\|^2$$

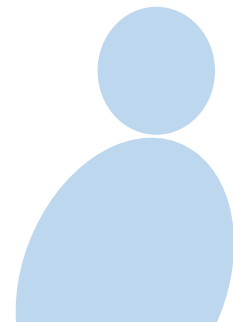
Least
Squares (LS)
problem

minimize square of error
Least Square error



Model Computation

- This is termed the least-squares (LS) problem

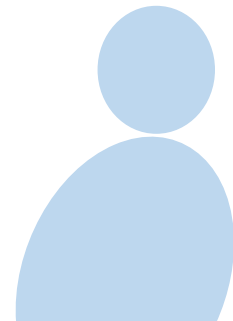


Model Computation

- The **regression coefficients** are given as

$$\bar{h} = (X^T X)^{-1} X^T \bar{y}$$

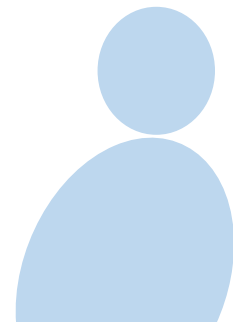
Formula for Regression
coefficients.



Model Computation

- The *regression coefficients* are given as

$$\bar{\mathbf{h}} = \underline{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \bar{\mathbf{y}}}$$



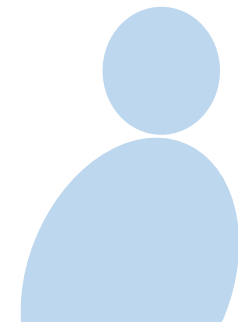
Model Computation

- The matrix $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ is termed as the pseudo-inverse of \mathbf{X} , since

$(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} = \mathbf{I}$

\mathbf{X} is Tall \Rightarrow NOT Square
 \Rightarrow #rows $>$ #columns.
 \Rightarrow \mathbf{X} does NOT have inverse

\Rightarrow Acts like an inverse of \mathbf{X}
 \Rightarrow Hence termed as pseudo inverse!

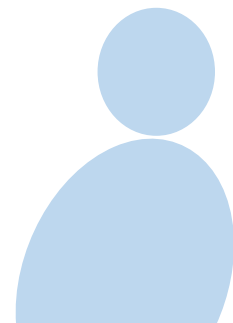


Model Computation

- The matrix $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ is termed as the pseudo-inverse of \mathbf{X} , since

$$(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} = \mathbf{I}$$

$$\bar{h} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \bar{y}$$



Regression Example

- Solve the **LS problem**

$$\min \left\| \begin{bmatrix} -1 \\ 2 \\ -3 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} h_0 \\ h_1 \end{bmatrix} \right\|^2$$

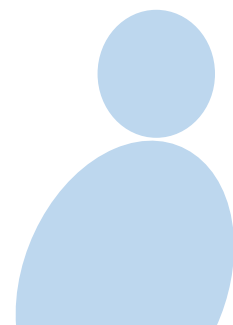
\bar{y}

4 x 1

$\Rightarrow M=4$
Training
Examples.

2 Regression
Coefficients.

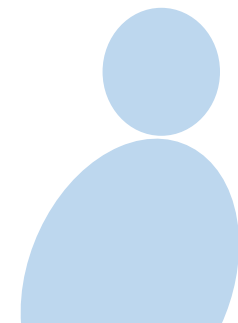
X
4 x 2
4 > 2
Tall matrix



Regression Example

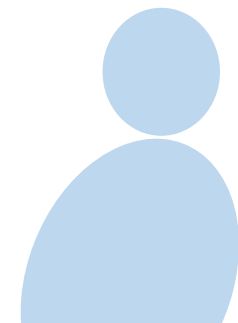
$$\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \quad \bar{\mathbf{y}} = \begin{bmatrix} -1 \\ 2 \\ -3 \\ 1 \end{bmatrix}$$

$\bar{h} = ?$



Regression Example

$$\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \quad \bar{\mathbf{y}} = \begin{bmatrix} -1 \\ 2 \\ -3 \\ 1 \end{bmatrix}$$



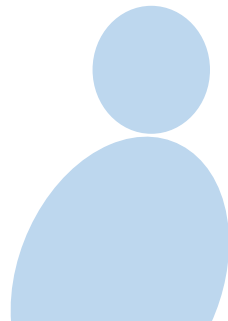
Regression Example

$$\bar{\mathbf{h}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \bar{\mathbf{y}}$$

$$\mathbf{X}^T \mathbf{X} =$$

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}$$

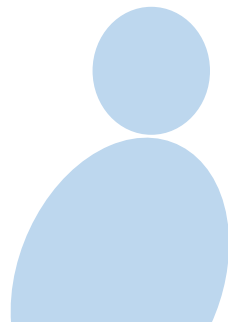
$$= \begin{bmatrix} 4 & 10 \\ 10 & 30 \end{bmatrix}$$



Regression Example

$$\bar{\mathbf{h}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \bar{\mathbf{y}}$$

$$\mathbf{X}^T \mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}$$
$$= \begin{bmatrix} 4 & 10 \\ 10 & 30 \end{bmatrix}$$



Regression Example

$$(\mathbf{X}^T \mathbf{X})^{-1} = \begin{bmatrix} 4 & 10 \\ 10 & 30 \end{bmatrix}^{-1}$$

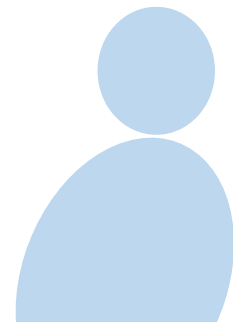
$$= \frac{1}{20} \begin{bmatrix} 30 & -10 \\ -10 & 4 \end{bmatrix} = (\mathbf{X}^T \mathbf{X})^{-1}$$

2x2 inverse

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Regression Example

$$(\mathbf{X}^T \mathbf{X})^{-1} = \frac{1}{20} \begin{bmatrix} 30 & -10 \\ -10 & 4 \end{bmatrix}$$



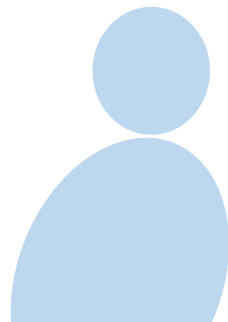
Regression Example

$$\bar{\mathbf{h}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \bar{\mathbf{y}} =$$

$$= \frac{1}{20} \begin{bmatrix} 30 & -10 \\ -10 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \\ -3 \\ 1 \end{bmatrix}$$

$$= \frac{1}{20} \begin{bmatrix} 20 & 10 & 0 & -10 \\ -6 & -2 & 2 & 6 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \\ -3 \\ 1 \end{bmatrix}$$

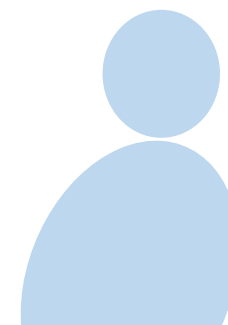
$$= \frac{1}{20} \begin{bmatrix} -10 \\ 2 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{10} \end{bmatrix}$$



Regression Example

$$(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \bar{\mathbf{y}}$$

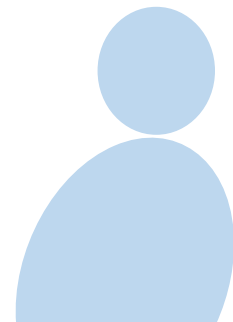
$$\begin{aligned} &= \frac{1}{20} \underbrace{\begin{bmatrix} 30 & -10 \\ -10 & 4 \end{bmatrix}}_{(\mathbf{X}^T \mathbf{X})^{-1}} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \\ -3 \\ 1 \end{bmatrix} \\ &= \frac{1}{20} \underbrace{\begin{bmatrix} 30 & -10 \\ -10 & 4 \end{bmatrix}}_{(\mathbf{X}^T \mathbf{X})^{-1}} \begin{bmatrix} -1 \\ -2 \end{bmatrix} = \frac{1}{20} \begin{bmatrix} -10 \\ 2 \end{bmatrix} \end{aligned}$$



Regression Example

$$\mathbf{h} = \begin{bmatrix} 1 \\ -\frac{1}{2} \\ 1 \\ \frac{1}{10} \end{bmatrix}$$

Regression coefficient
vector



Instructors may use this white area (14.5 cm / 25.4 cm) for the text.
Three options provided below for the font size.

Font: Avenir (Book), Size: 32, Colour: Dark Grey

Font: Avenir (Book), Size: 28, Colour: Dark Grey

Font: Avenir (Book), Size: 24, Colour: Dark Grey

Do not use the space below.

