Fondamenti di Machine Learning

La Regressione

Introduzione alla regressione lineare semplice

presentato da
Giuseppe Gullo

PROFESSION AI

La regressione lineare semplice

Ci permette di trovare una relazione lineare tra
una variabile indipendente (feature)
e una variabile dipendente (target).

Metri quadri	Valore in €
80	160.000
150	300.000
30	120.000
50	100.000
120	240.000
60	180.000
110	200.000
110	250.000
70	???
60	???

La regressione lineare

Dataset di monolocali a Milano

La regressione lineare

Dataset di monolocali a Milano

Metri quadri	Valore in €
80	160.000
150	300.000
30	120.000
50	100.000
120	240.000
60	180.000
110	200.000
110	250.000
70	???
60	???

Quanto valgono questi appartamenti?

x	y
80	160.000
150	300.000
30	120.000
50	100.000
120	240.000
60	180.000
110	200.000
110	250.000
70	???
60	???

La regressione lineare

Dataset di monolocali a Milano

$$x$$

La regressione lineare

Dataset di monolocali a Milano

x	y
80	160.000
150	300.000
30	120.000
50	100.000
120	240.000
60	180.000
110	200.000
110	250.000
70	???
60	???

$$x \qquad\qquad y$$

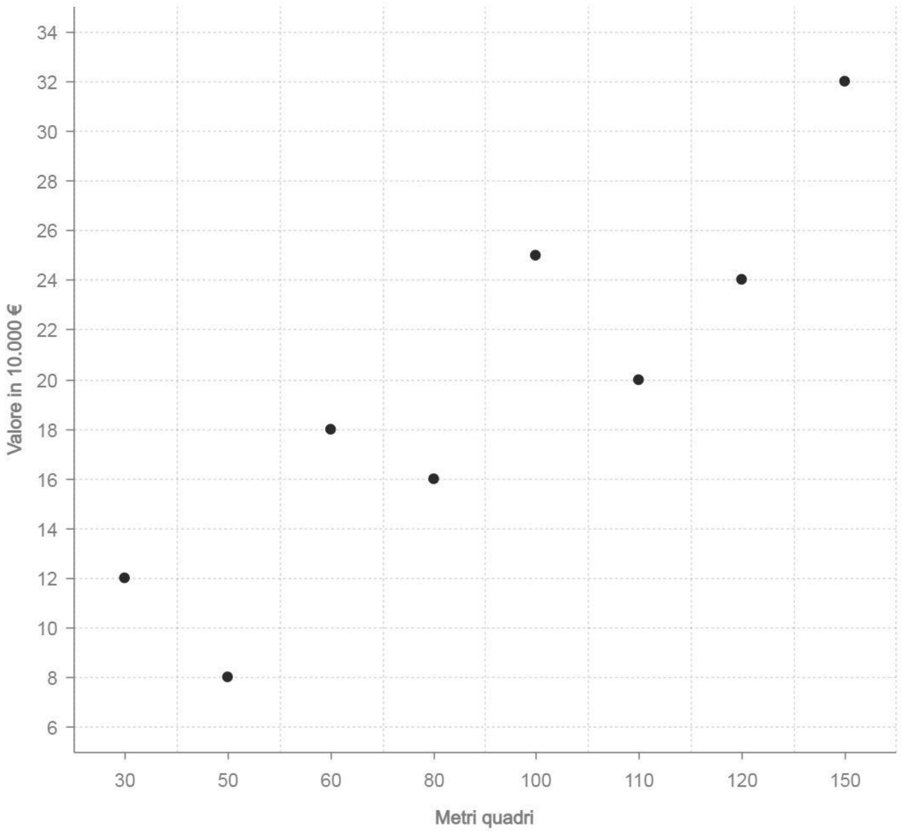
Metri quadri	Valore in €
80	160.000
150	300.000
30	120.000
50	100.000
120	240.000
60	180.000
110	200.000
110	250.000
70	???
60	???

La regressione lineare

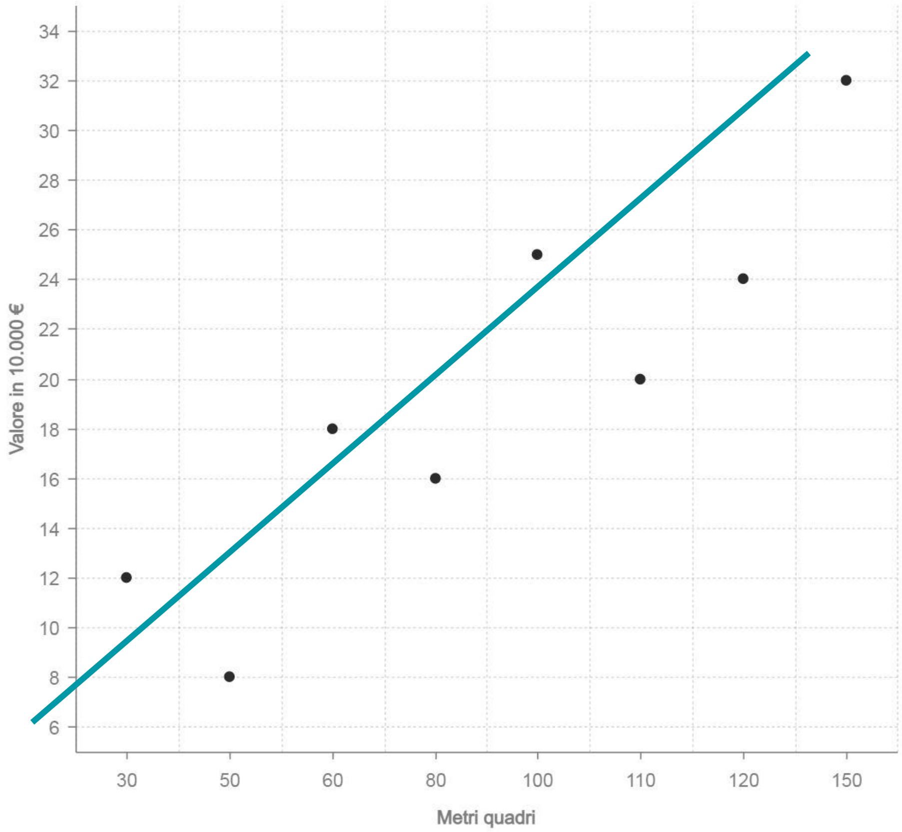
Dataset di monolocali a Milano

$$f(x) = y$$

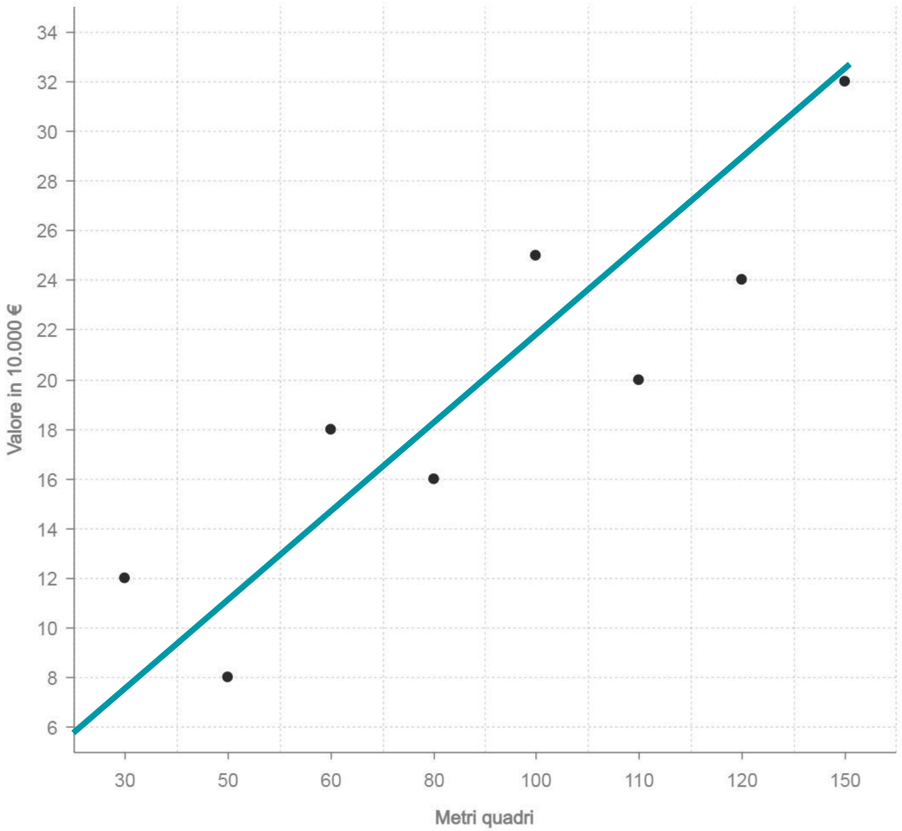
Rappresentazione grafica del dataset



Rappresentazione grafica del dataset



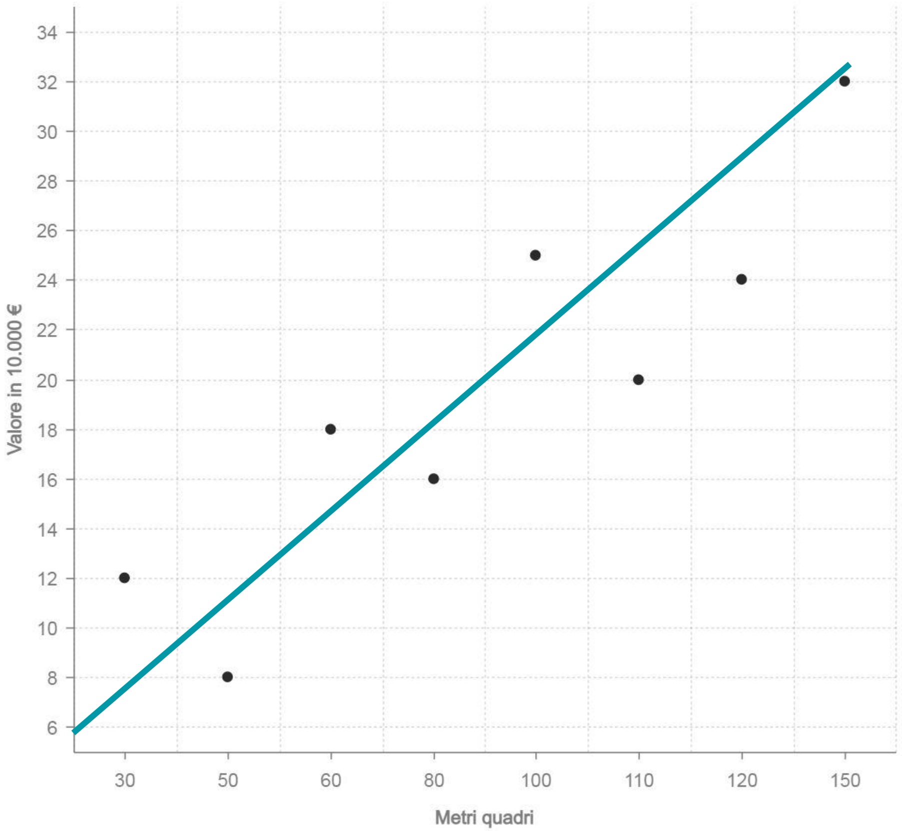
Rappresentazione grafica del dataset



Equazione di una retta
(forma esplicita)

$$y = q + mx$$

Rappresentazione grafica del dataset

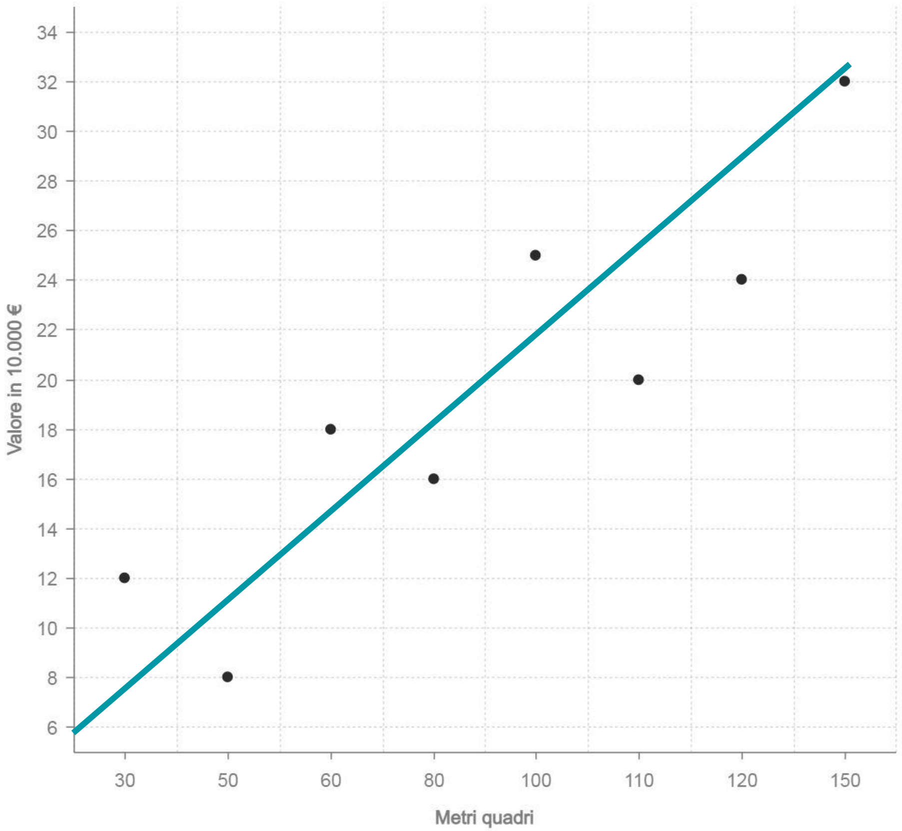


Equazione di una retta
(forma esplicita)

$$y = q + mx$$

- y: VARIABILE DIPENDENTE
- q: INTERCETTA
- m: COEFFICIENTE (ANGOLARE)
- x: VARIABILE INDIPENDENTE

Rappresentazione grafica del dataset



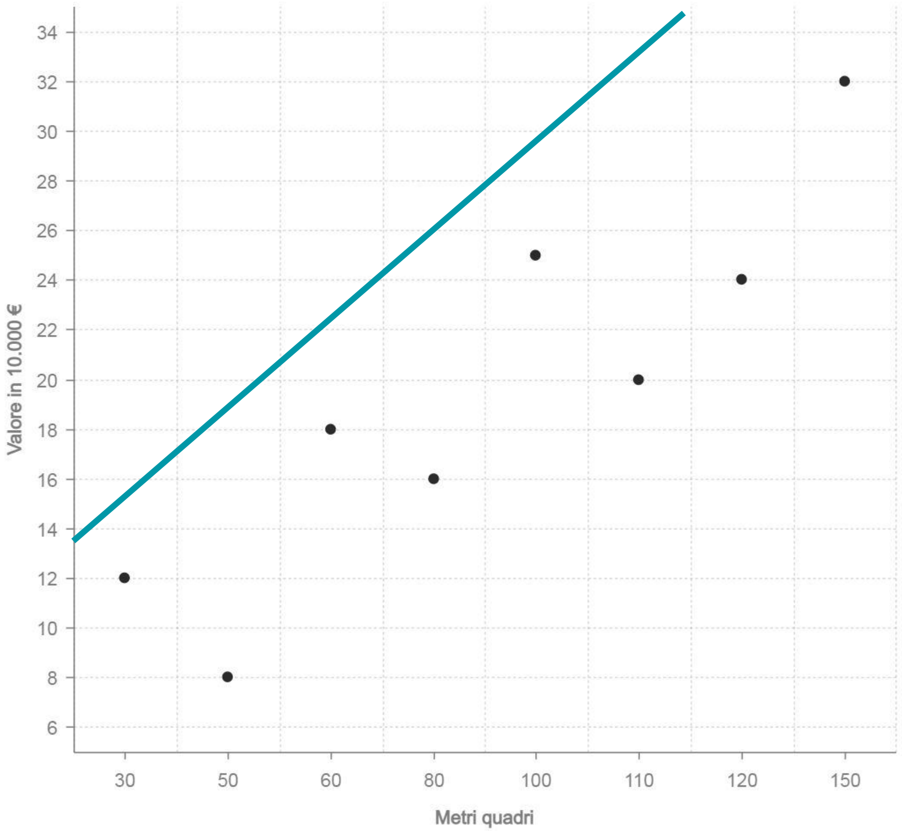
Equazione di una retta
(forma esplicita)

$$y = \underline{q} + mx$$

L'intercetta stabilisce la distanza della retta dall'origine.

Quindi funge da **bias**.

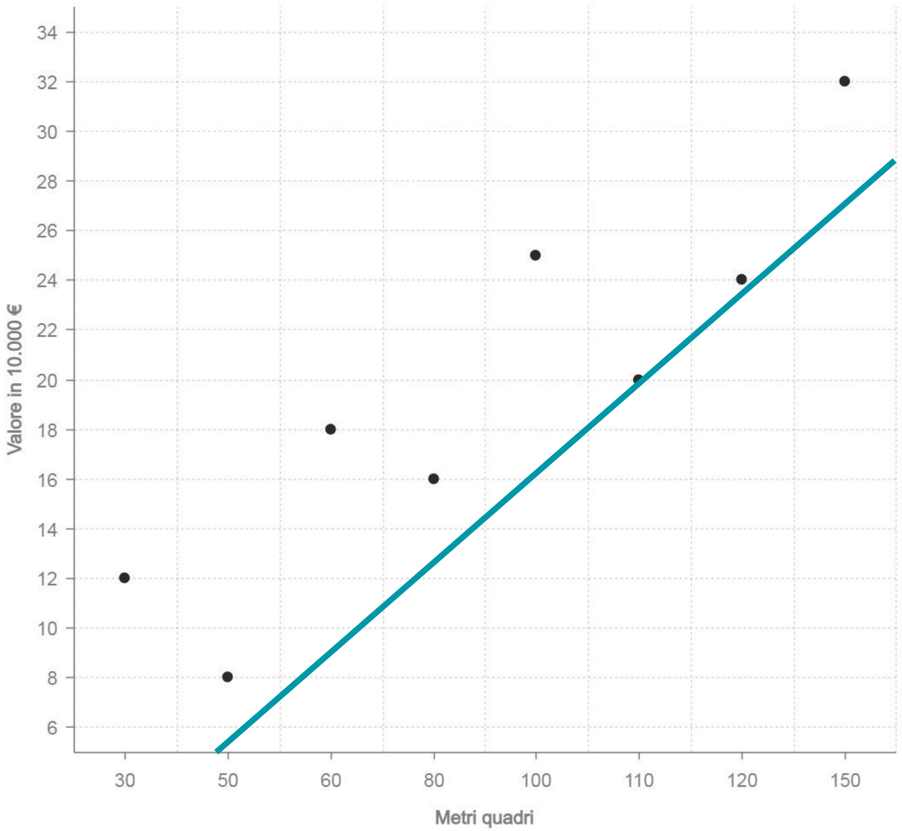
Rappresentazione grafica del dataset



Equazione di una retta
(forma esplicita)

$$y = \underline{q} + mx$$

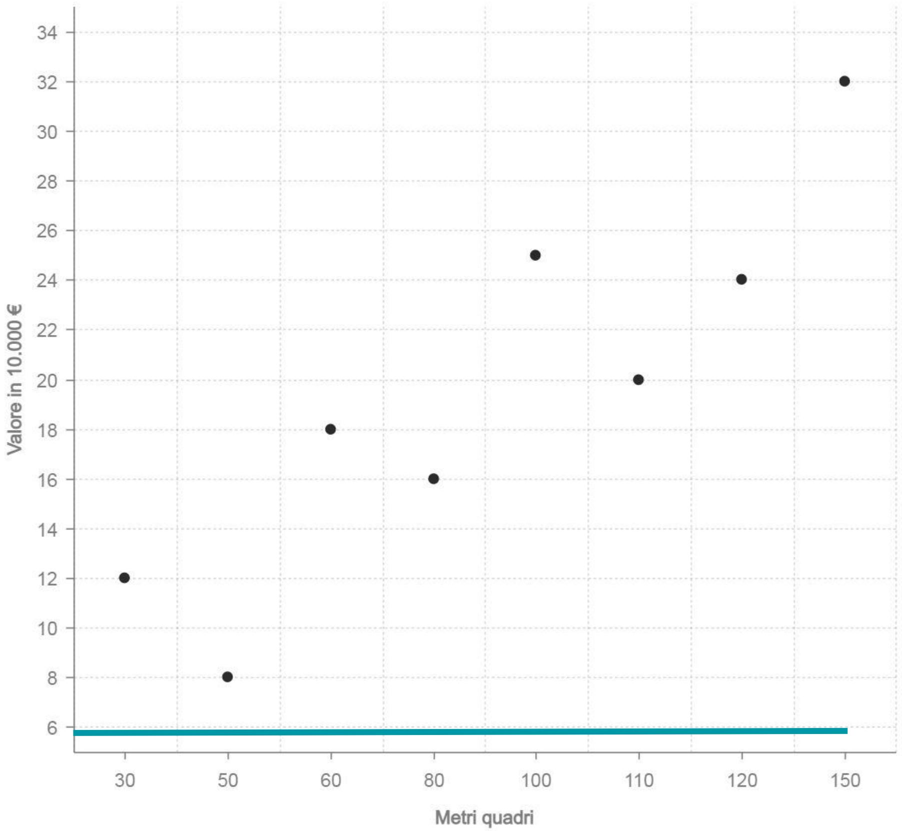
Rappresentazione grafica del dataset



Equazione di una retta
(forma esplicita)

$$y = \underline{q} + mx$$

Rappresentazione grafica del dataset

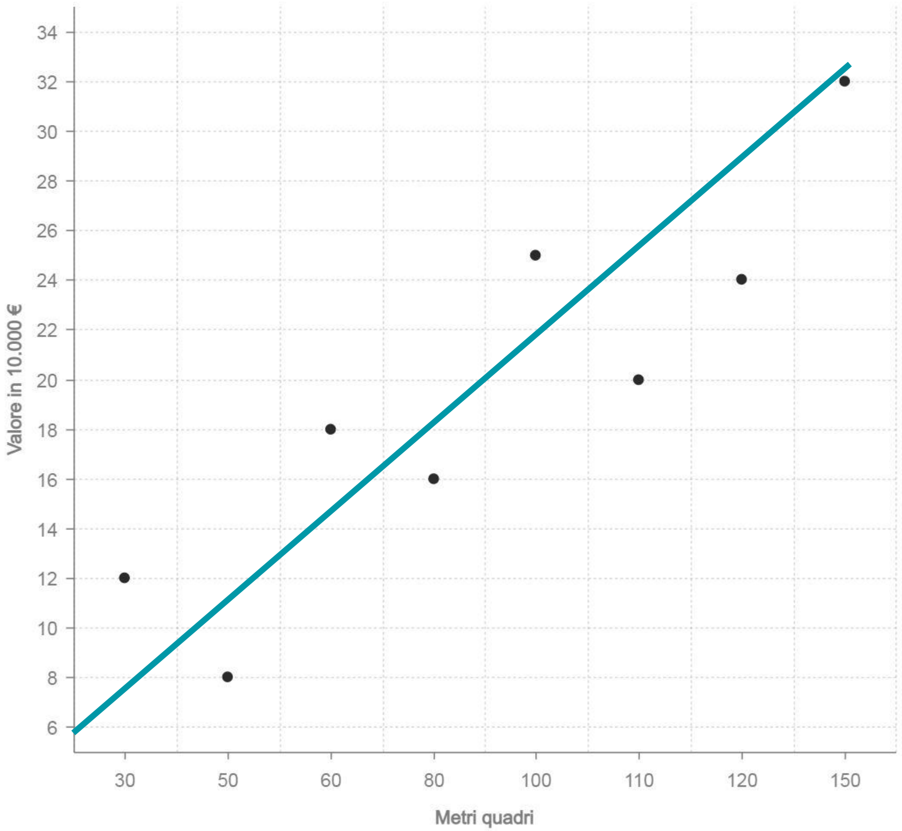


Equazione di una retta
(forma esplicita)

$$y = q$$

Se il coefficiente angolare è pari a 0, vuol dire che la feature non ha nessun impatto sul target

Rappresentazione grafica del dataset



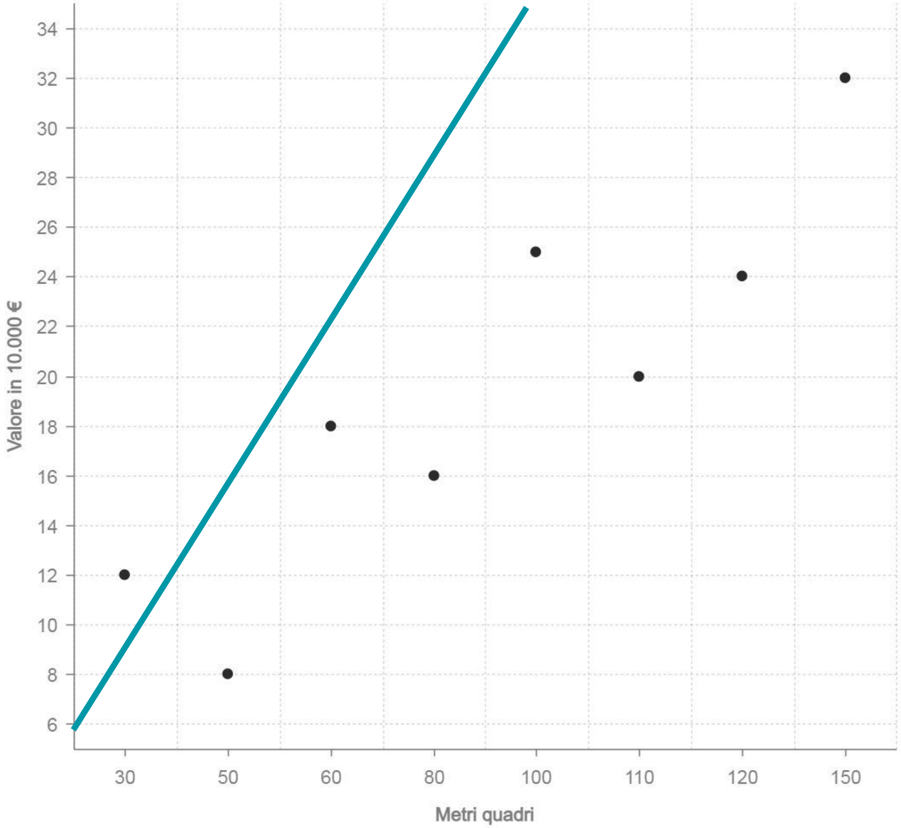
Equazione di una retta
(forma esplicita)

$$y = q + \underline{m}x$$

Il coefficiente angolare stabilisce l'impatto che una variazione della variabile indipendente ha sulla variabile dipendente.

Quindi funge da **peso**

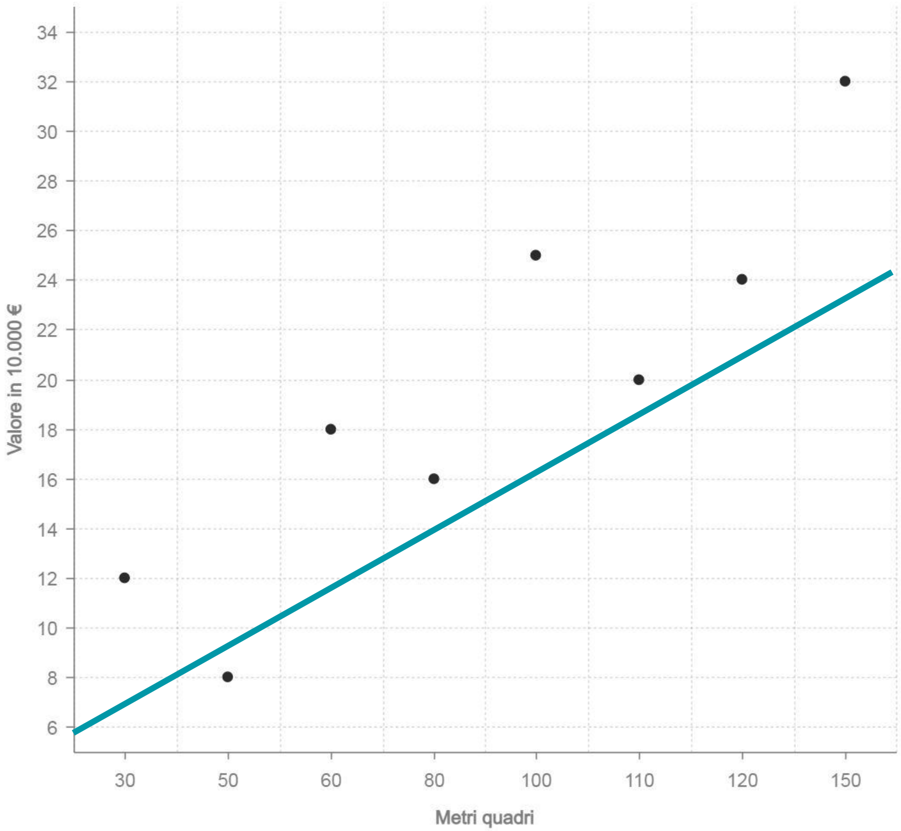
Rappresentazione grafica del dataset



Equazione di una retta
(forma esplicita)

$$y = q + \underline{m}x$$

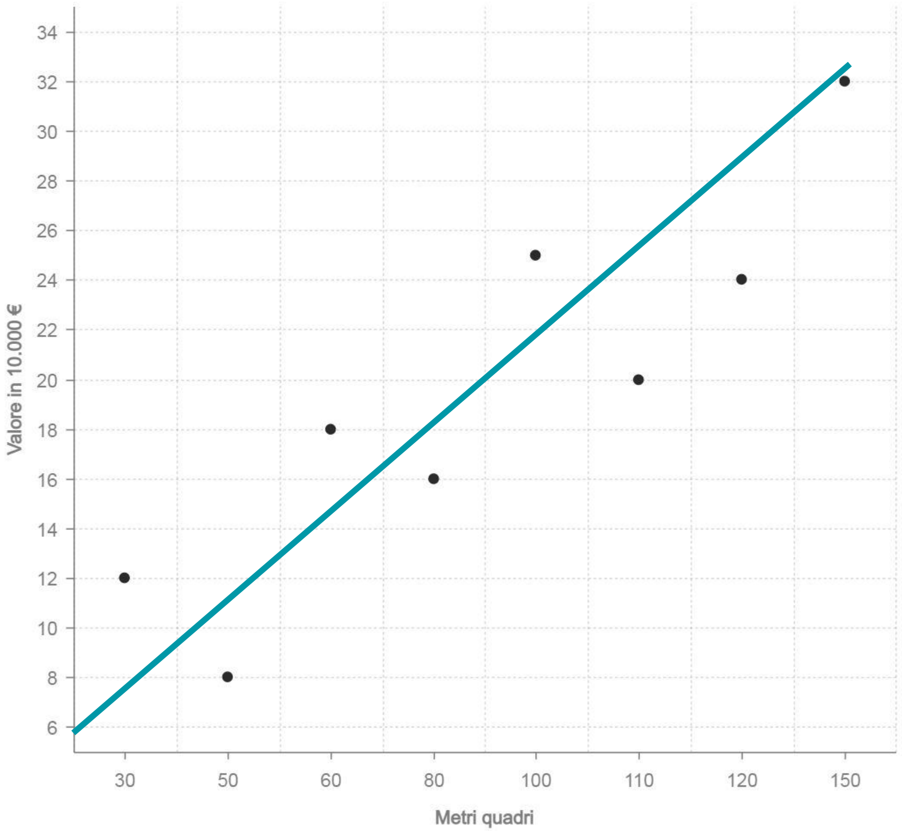
Rappresentazione grafica del dataset



Equazione di una retta
(forma esplicita)

$$y = q + \underline{m}x$$

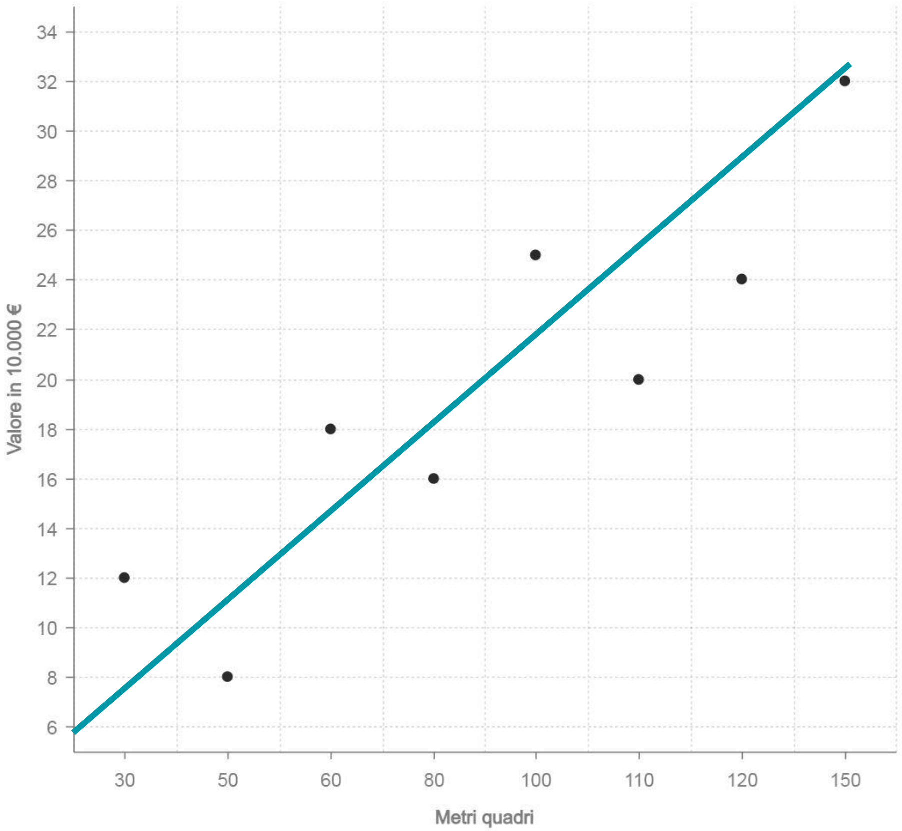
Rappresentazione grafica del dataset



Equazione di una retta
(forma esplicita)

$$\underset{\text{PREVISIONE}}{y} = \overset{\text{BIAS}}{b} + \underset{\text{PESO}}{w}\overset{\text{FEATURE}}{x}$$

Rappresentazione grafica del dataset



Equazione di una retta

(forma esplicita)

$$y = 5.5 + 0.16x$$

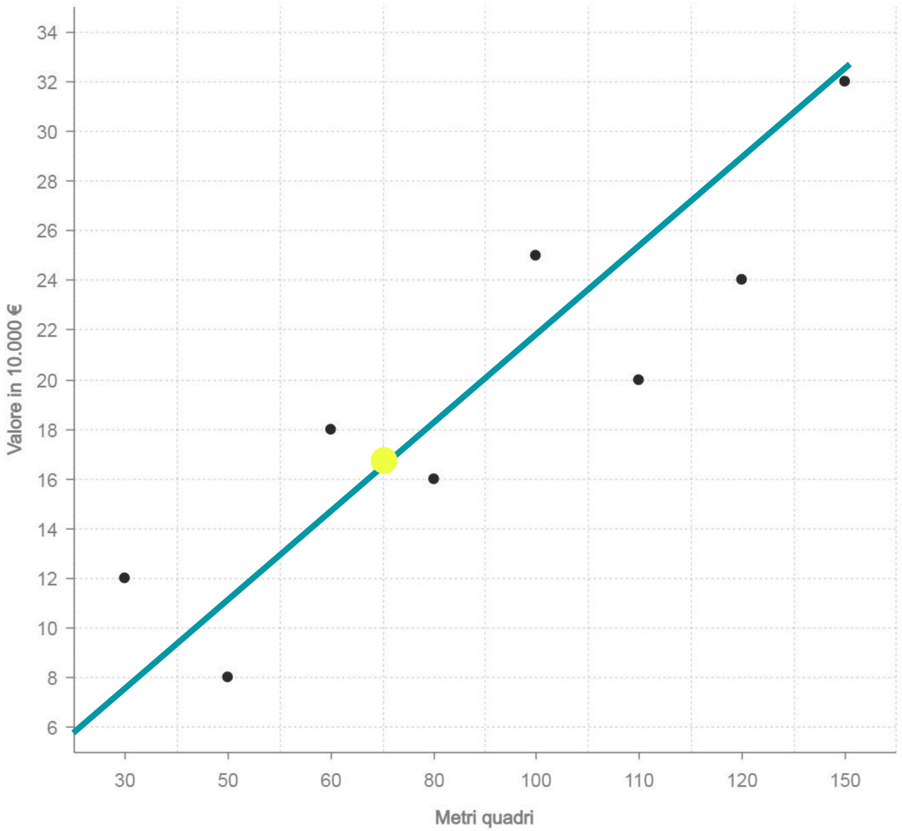
Metri quadri	Valore in €
80	160.000
150	300.000
30	120.000
50	100.000
120	240.000
60	180.000
110	200.000
110	250.000
70	???
60	???

La regressione lineare

Dataset di monolocali a Milano

Quanto valgono questi appartamenti?

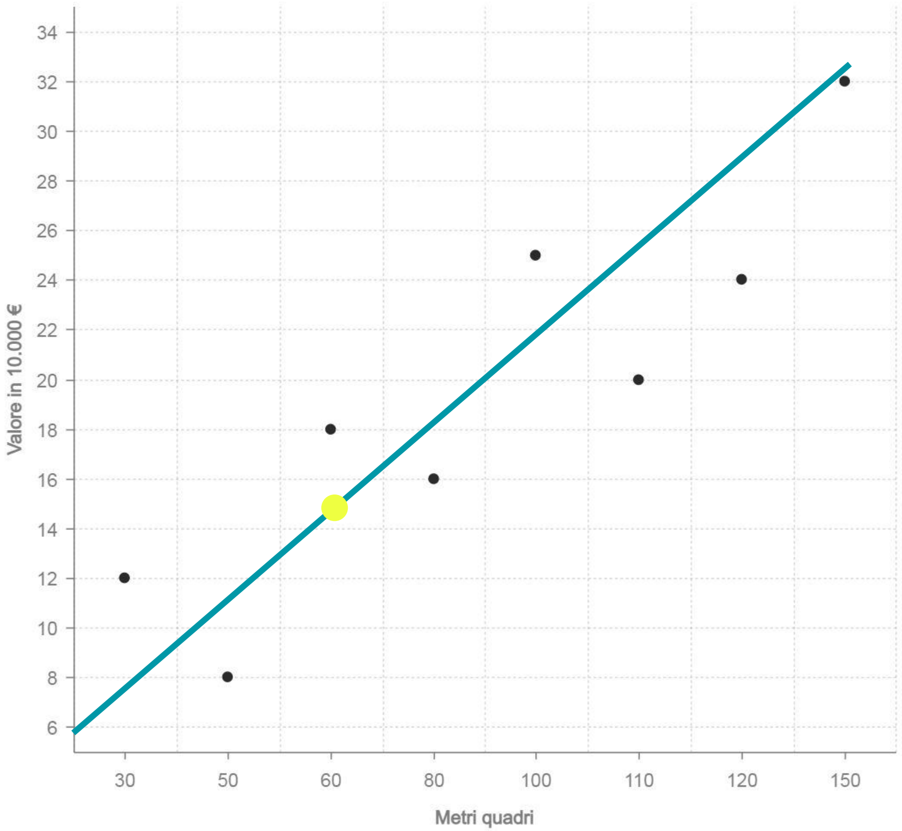
Rappresentazione grafica del dataset



Equazione di una retta
(forma esplicita)

$$5.5 + 0.16 \cdot 70 = 16.7$$

Rappresentazione grafica del dataset



Equazione di una retta
(forma esplicita)

$$5.5 + 0.16 \cdot 60 = 15.1$$

Fondamenti di Machine Learning

La Regressione

il Metodo dei Minimi Quadrati

presentato da
Giuseppe Gullo

PROFESSION AI

il Metodo dei Minimi Quadrati

E’ un metodo numerico che ci permette di stimare
il valore dei coefficienti di un modello di regressione lineare.

il Metodo dei Minimi Quadrati

E' un metodo numerico che ci permette di stimare
il valore dei coefficienti di un modello di regressione lineare.

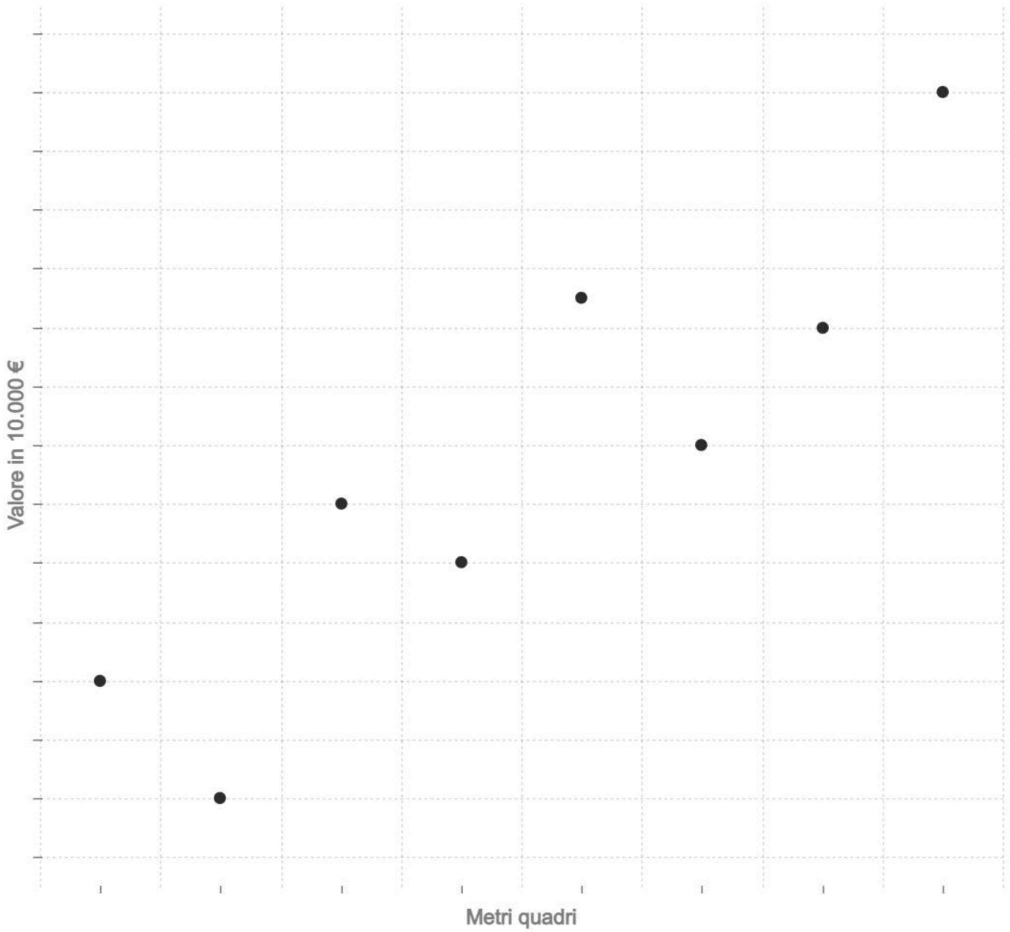
Lo fa trovando i coefficienti che minimizzano l'errore del modello.

Set di dati

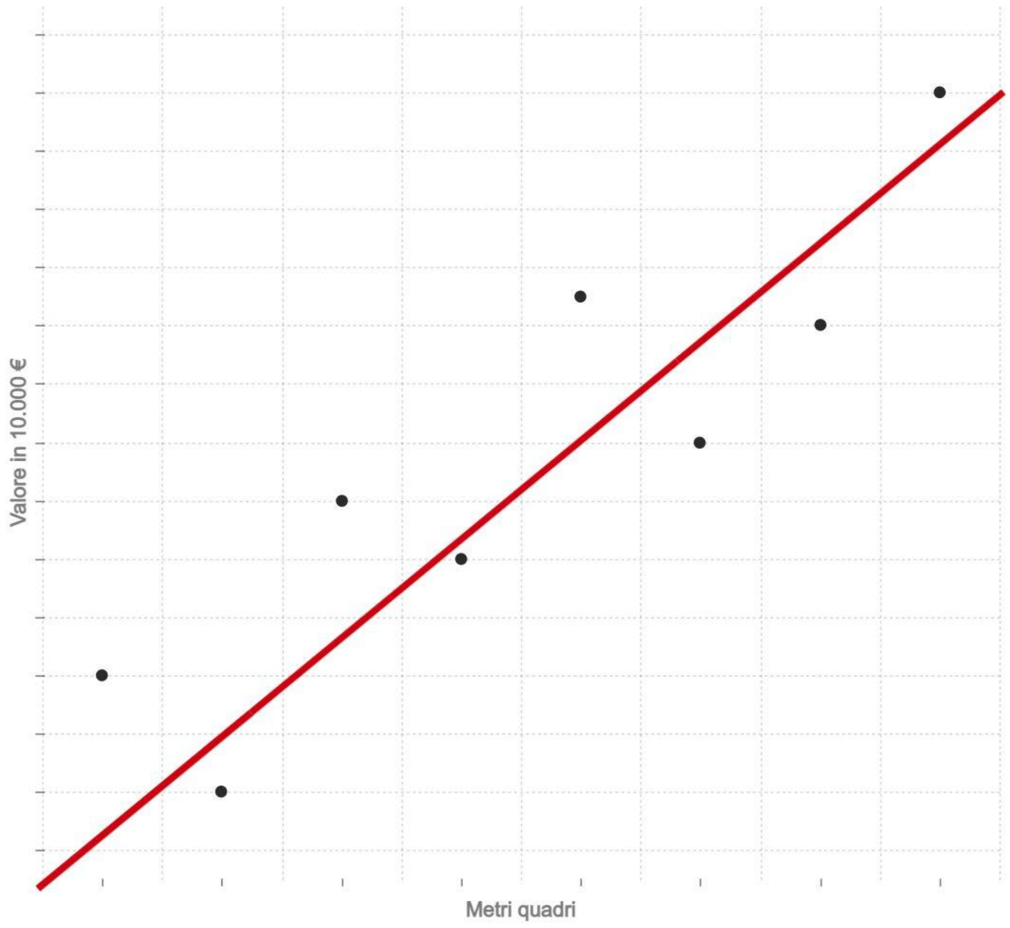
Metri quadri	Valore in €
80	160.000
150	300.000
30	120.000
50	100.000
120	240.000
60	180.000
110	200.000
110	250.000
70	???
60	???

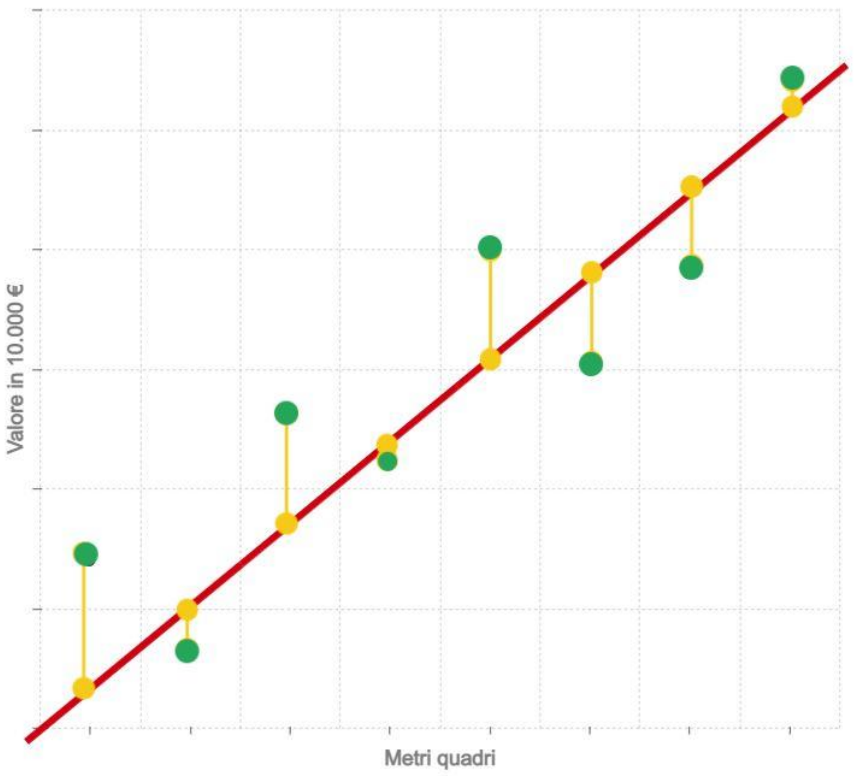
Dati di addestramento

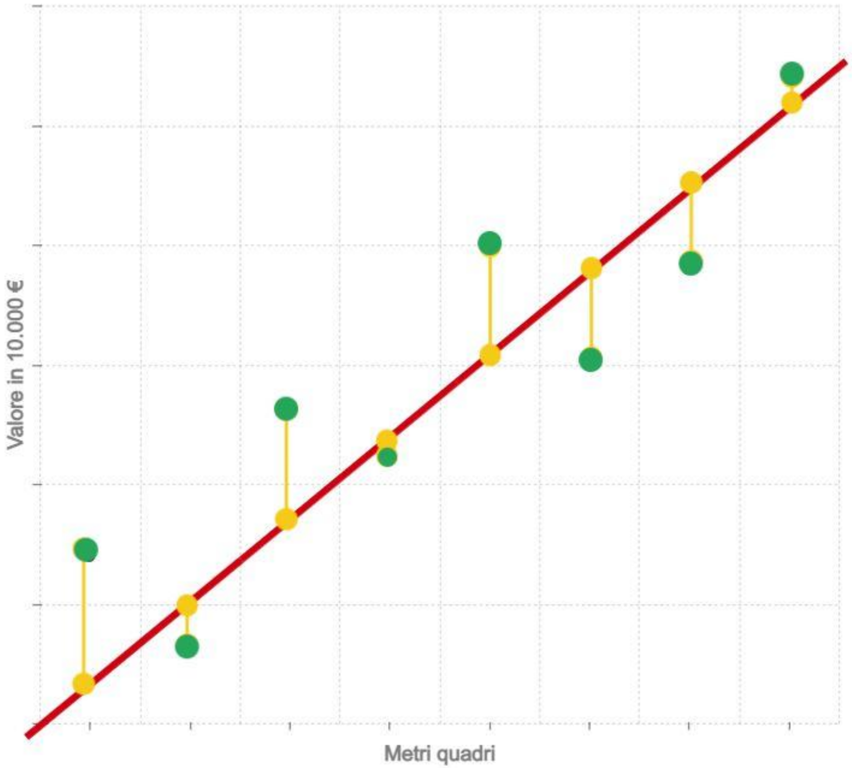
Metri quadri	Valore in €
80	160.000
150	300.000
30	120.000
50	100.000
120	240.000
60	180.000
110	200.000
110	250.000



Valore in 10.000 €
Metri quadri

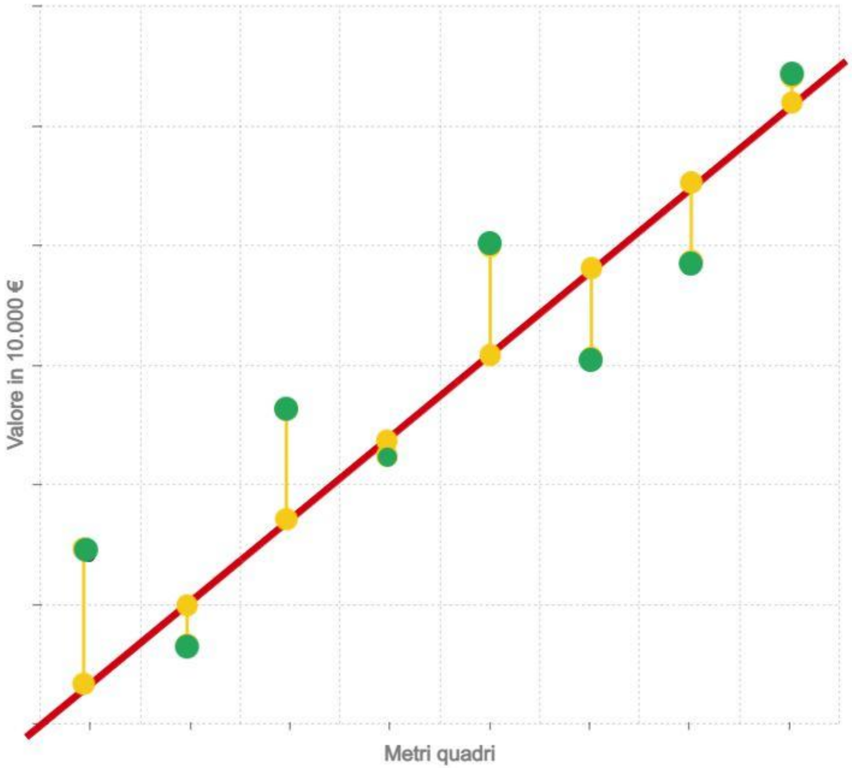

Valore in 10.000 €
Metri quadri


Valore in 10.000 €
Metri quadri



I residui

x	y	previsione
80	16	18.3
150	30	29.5
30	12	10.3
50	10	13.5
120	24	24.7
60	18	15.1
110	20	23.1
110	25	23.1



I residui

x	y	previsione	residui
80	16	18.3	-2.3
150	30	29.5	0.5
30	12	10.3	1.7
50	10	13.5	-3.5
120	24	24.7	-0.7
60	18	15.1	2.9
110	20	23.1	-3.1
110	25	23.1	1.9

Funzione di costo (Cost Function)

Somma dei quadrati residui (RSS)

$$RSS = \sum_{i=1}^{N} (y_i - \hat{y_i})^2$$

Funzione di costo (Cost Function)

Somma dei quadrati residui (RSS)

$$RSS = \sum_{i=1}^{N} (y_i - f(x))^2$$

Funzione di costo (Cost Function)

Somma dei quadrati residui (RSS)

$$RSS = \sum_{i=1}^{N} (y_i - wx + b)^2$$

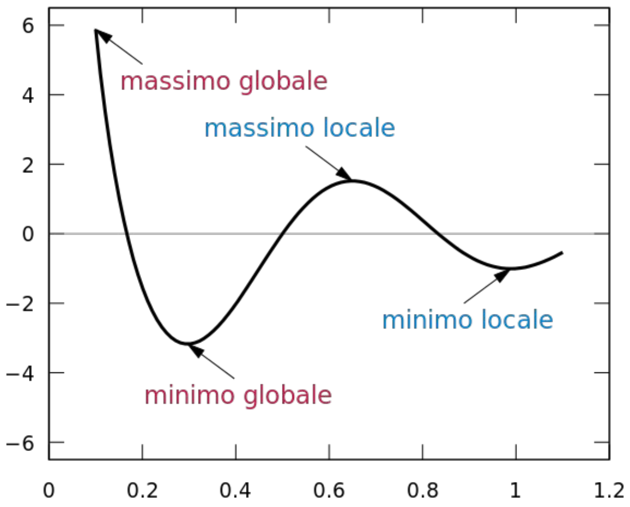
Funzione obiettivo (Objective Function)

$$L(w,b) = \sum_{i=1}^{N} (y_i - \underline{w}x + \underline{b})^2$$

Dobbiamo trovare i valori di w e b
che minimizzano la somma dei quadrati residui.

Ottimizzazione matematica

L'ottimizzazione è una branca della matematica applicata
che studia teoria e metodi per la ricerca
dei punti di massimo e minimo di una funzione matematica
all'interno di un dominio specificato.



Dati di addestramento

x	y
80	160.000
150	300.000
30	120.000
50	100.000
120	240.000
60	180.000
110	200.000
110	250.000

$$wx + b = y$$

il Metodo dei Minimi Quadrati in Pratica

x	y	xy	x^2
80	16	1280	6400
150	30	4500	22500
30	12	360	900
50	10	500	2500
120	24	2880	14400
60	18	1080	3600
110	20	2200	12100
110	25	2750	12100

il Metodo dei Minimi Quadrati in Pratica

x	y	xy	x^2
80	16	1280	6400
150	30	4500	22500
30	12	360	900
50	10	500	2500
120	24	2880	14400
60	18	1080	3600
110	20	2200	12100
110	25	2750	12100

$$\sum x = 710$$

$$\sum y = 155$$

$$\sum xy = 15550$$

$$\sum x^2 = 74500$$

$$n = 8$$

il Metodo dei Minimi Quadrati in Pratica

Peso

$$\underline{w}x + b = y$$

$$w = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - \left(\sum x\right)^2}$$

il Metodo dei Minimi Quadrati in Pratica

Peso

$$\underline{w}x + b = y$$

$$w = \frac{8 \cdot 15550 - 710 \cdot 155}{8 \cdot 74500 - (710)^2}$$

il Metodo dei Minimi Quadrati in Pratica

Peso

$$\underline{w}x + b = y$$

$$w = 0.16$$

il Metodo dei Minimi Quadrati in Pratica

Bias

$$0.16x + \underline{b} = y$$

$$b = \frac{\sum y - w \sum x}{n}$$

il Metodo dei Minimi Quadrati in Pratica

Bias

$$0.16x + \underline{b} = y$$

$$b = \frac{155 - 0.16 \cdot 710}{8}$$

il Metodo dei Minimi Quadrati in Pratica

L'equazione lineare del modello

$$0.16x + 5.5 = y$$

il Metodo dei Minimi Quadrati in Pratica

L’equazione lineare del modello

$$0.16x + 5.5 = y$$

Metri quadri	Valore in €
80	160.000
150	300.000
30	120.000
50	100.000
120	240.000
60	180.000
110	200.000
110	250.000
70	???
60	???

$$0.16 \cdot 70 + 5.5 = 16.7$$

Metri quadri	Valore in €
80	160.000
150	300.000
30	120.000
50	100.000
120	240.000
60	180.000
110	200.000
110	250.000
70	167.000
60	???

$$0.16 \cdot 70 + 5.5 = 16.7$$

Metri quadri	Valore in €
80	160.000
150	300.000
30	120.000
50	100.000
120	240.000
60	180.000
110	200.000
110	250.000
70	167.000
60	151.000

$$0.16 \cdot 60 + 5.5 = 15.1$$

Come ricavare le formule?

METODO DEI MINIMI QUADRATI
STEP BY STEP

OBIETTIVO

Trovare i valori di pesi e bias che minimizzano l'RSS.

Nel punto di minimo la derivata rispetto a w o b vale 0.

$$RSS = \sum_{i=1}^{N} (y_i - (wx_i + b))^2$$

PASSAGGI

1) Esprimiamo l'RSS in funzione di w e b

$$L(w,b) = \sum_{i=1}^{N} (y_i - (wx_i + b))^2$$

QUADRATO DI UN TRINOMIO

$$= \sum_{i=1}^{N} (w^2x_i^2 + 2bwx_i + b^2 - 2wx_iy_i - 2by_i + y_i^2)$$

$$= w^2 \Sigma x^2 + 2bw \Sigma x + nb^2 - 2w \Sigma xy - 2b \Sigma y + \Sigma y^2$$

2) Calcoliamo le derivate parziali rispetto w e b

$$\frac{\partial L}{\partial w} = \cancel{2} w \Sigma x^2 + \cancel{2} b \Sigma x - \cancel{2} \Sigma xy = w \Sigma x^2 + b \Sigma x - \Sigma xy$$

$$\frac{\partial L}{\partial b} = \cancel{2} w \Sigma x + \cancel{2} nb - \cancel{2} \Sigma y = w \Sigma x + nb - \Sigma y$$

3) Poniamo le equazioni ottenute a 0 e ricaviamo w e b.

$$w \Sigma x + nb - \Sigma y = 0$$

$$\Rightarrow \boxed{b = \frac{\Sigma y - w \Sigma x}{n}}$$

. Sostituiamo b in $\frac{\partial L}{\partial w}$ e risolviamo.

$$w \Sigma x^2 + \frac{(\Sigma y - w \Sigma x)}{n} \Sigma x - \Sigma xy = 0$$

$$\Rightarrow nw \Sigma x^2 + (\Sigma y - w \Sigma x) \Sigma x - n \Sigma xy = 0$$

$$\Rightarrow nw \Sigma x^2 + \Sigma x \Sigma y - w(\Sigma x)^2 - n \Sigma xy = 0$$

$$\Rightarrow nw \Sigma x^2 - w(\Sigma x)^2 = n \Sigma xy - \Sigma x \Sigma y$$

$$\Rightarrow w(n \Sigma x^2 - (\Sigma x)^2) = n \Sigma xy - \Sigma x \Sigma y$$

$$\Rightarrow \boxed{w = \frac{n \Sigma xy - \Sigma x \Sigma y}{n \Sigma x^2 - (\Sigma x)^2}}$$

Come ricavare le formule?

Se vuoi approfondire ed hai delle buone basi di analisi matematica, guarda l'approfondimento in allegato con tutti i passaggi.

Come ricavare le formule?

$$w = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - \left(\sum x\right)^2} \qquad b = \frac{\sum y - w\sum x}{n}$$

Come ricavare le formule?

1. Partiamo dalla somma dei quadrati residui

$$RSS = \sum_{i=1}^{N} (y_i - wx + b)^2$$

Come ricavare le formule?

2. Calcoliamo le derivate parziali rispetto a peso e bias

$$\delta w = w \sum x^2 + b \sum x - \sum xy$$

$$\delta b = w \sum x + bn - \sum y$$

Come ricavare le formule?

3. Risolviamo le equazioni per *w* e *b*.

$$w = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - \left(\sum x\right)^2}$$

$$b = \frac{\sum y - w\sum x}{n}$$

Fondamenti di Machine Learning

La Regressione

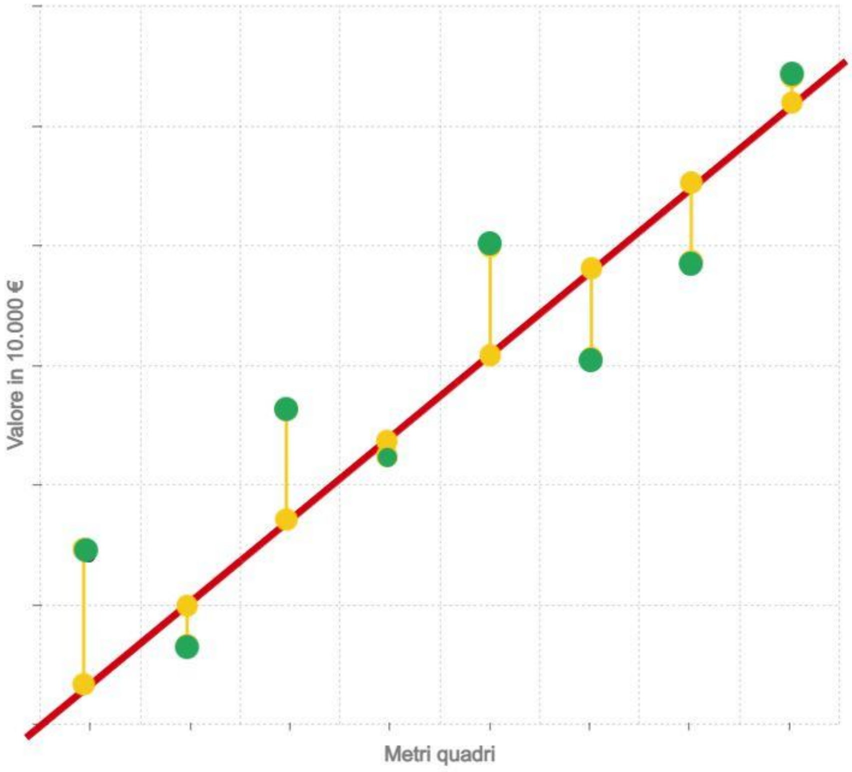
Le funzioni di costo

presentato da
Giuseppe Gullo

PROFESSION AI

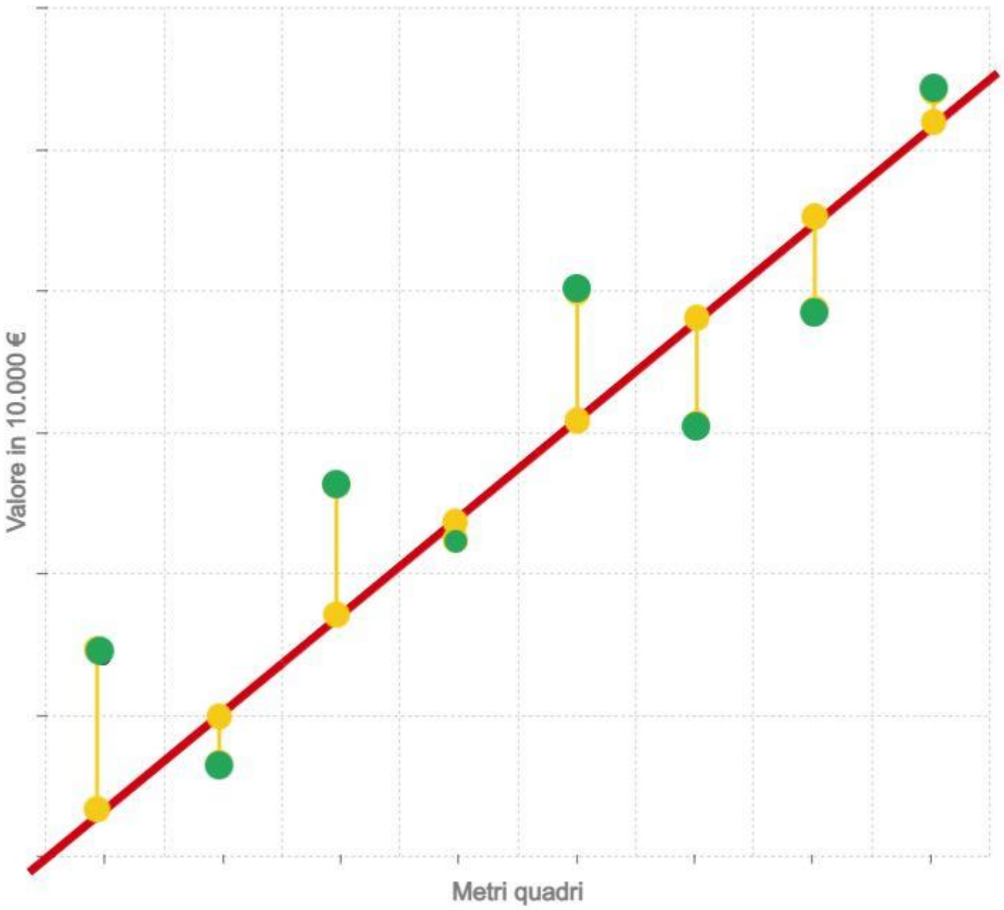
Le funzioni di costo

Permettono di quantificare la capacità predittiva del modello



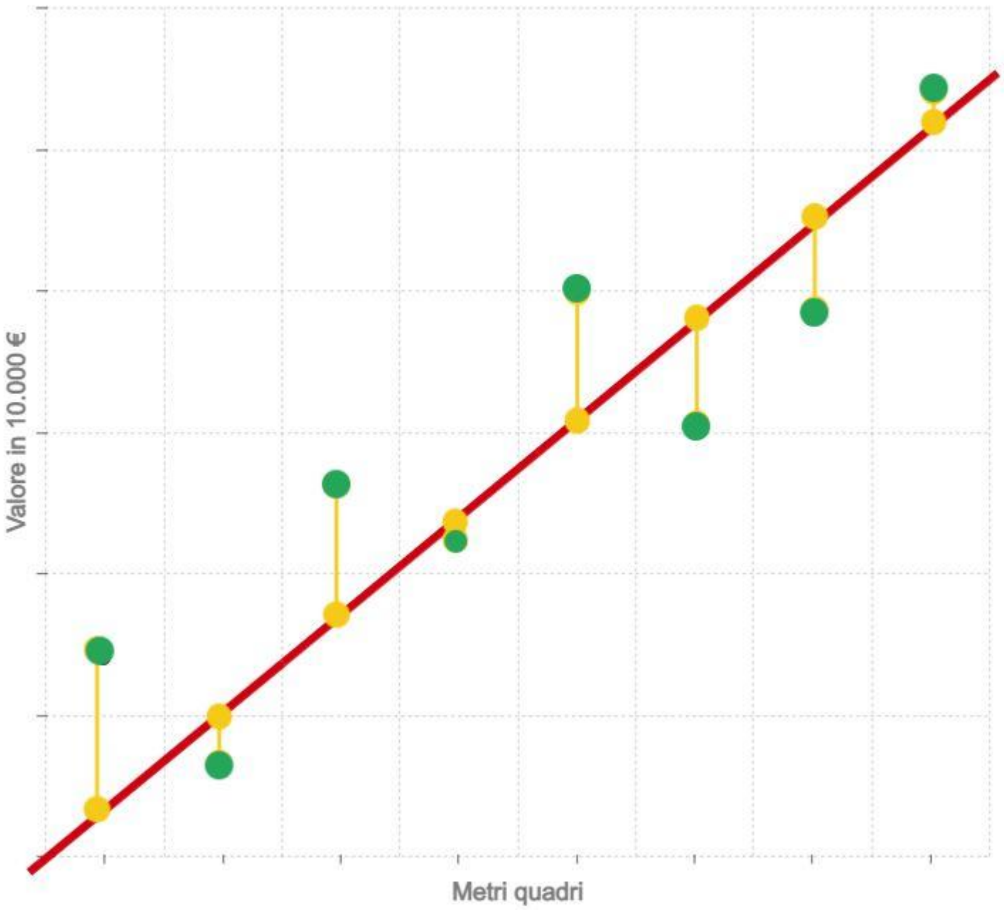
I residui

x	y	previsione	residui
80	16	18.3	-2.3
150	30	29.5	0.5
30	12	10.3	1.7
50	10	13.5	-3.5
120	24	24.7	-0.7
60	18	15.1	2.9
110	20	23.1	-3.1
110	25	23.1	1.9



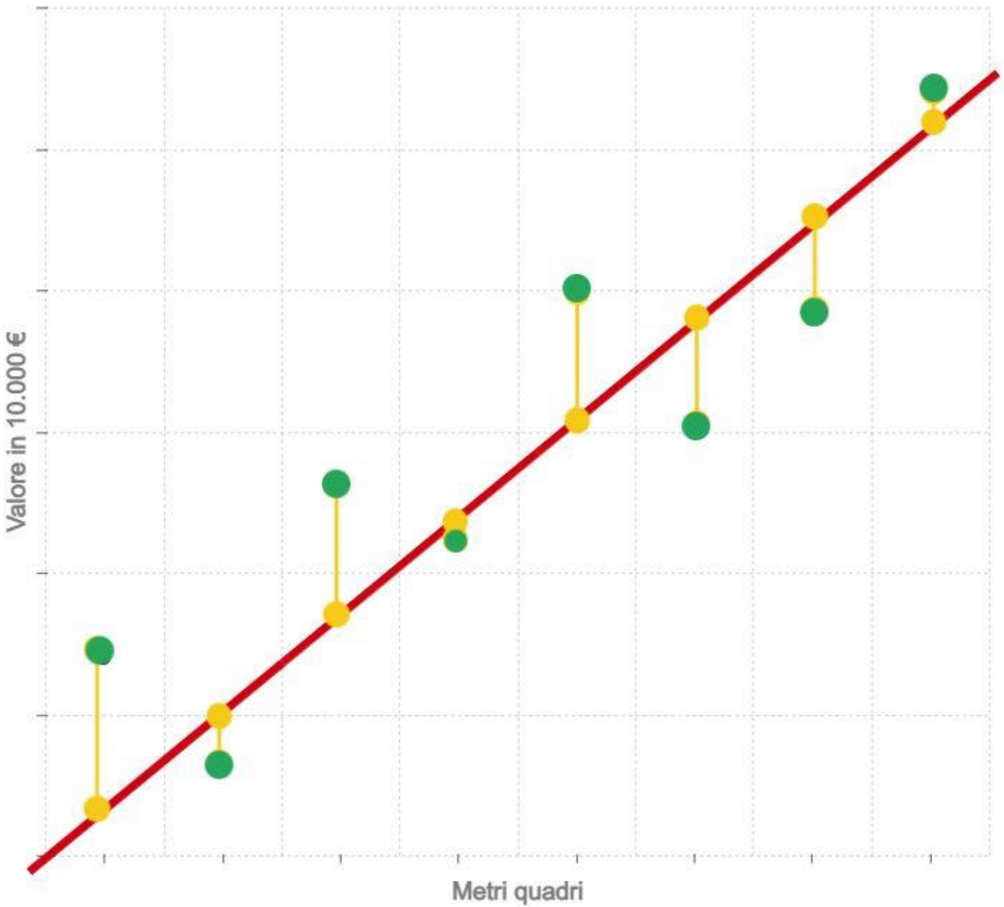
Errore assoluto medio (MAE)

Il MAE è definito come
la media della somma
del valore assoluto dei residui



Errore assoluto medio (MAE)

$$MAE = \frac{1}{n}\sum_{i=1}^{N}|y_i - f(x_i)|$$

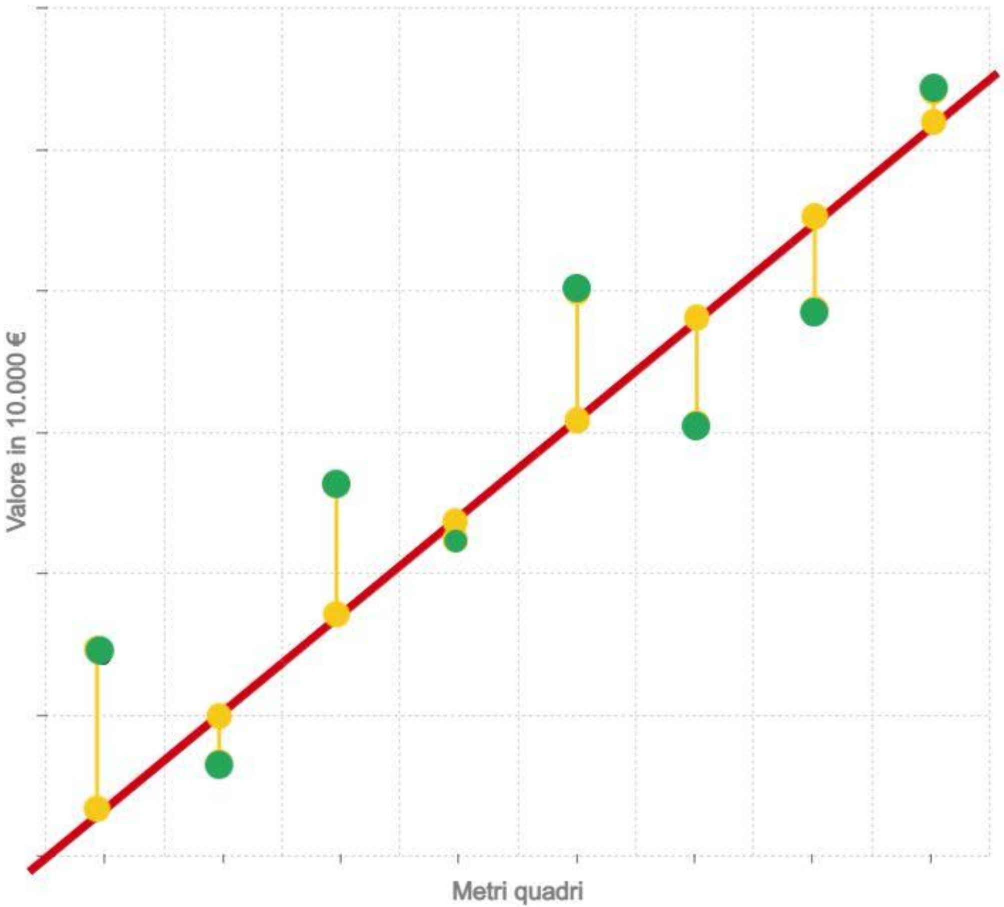


Errore assoluto medio (MAE)

$$MAE = \frac{1}{n}\sum_{i=1}^{N}|y_i - f(x_i)|$$

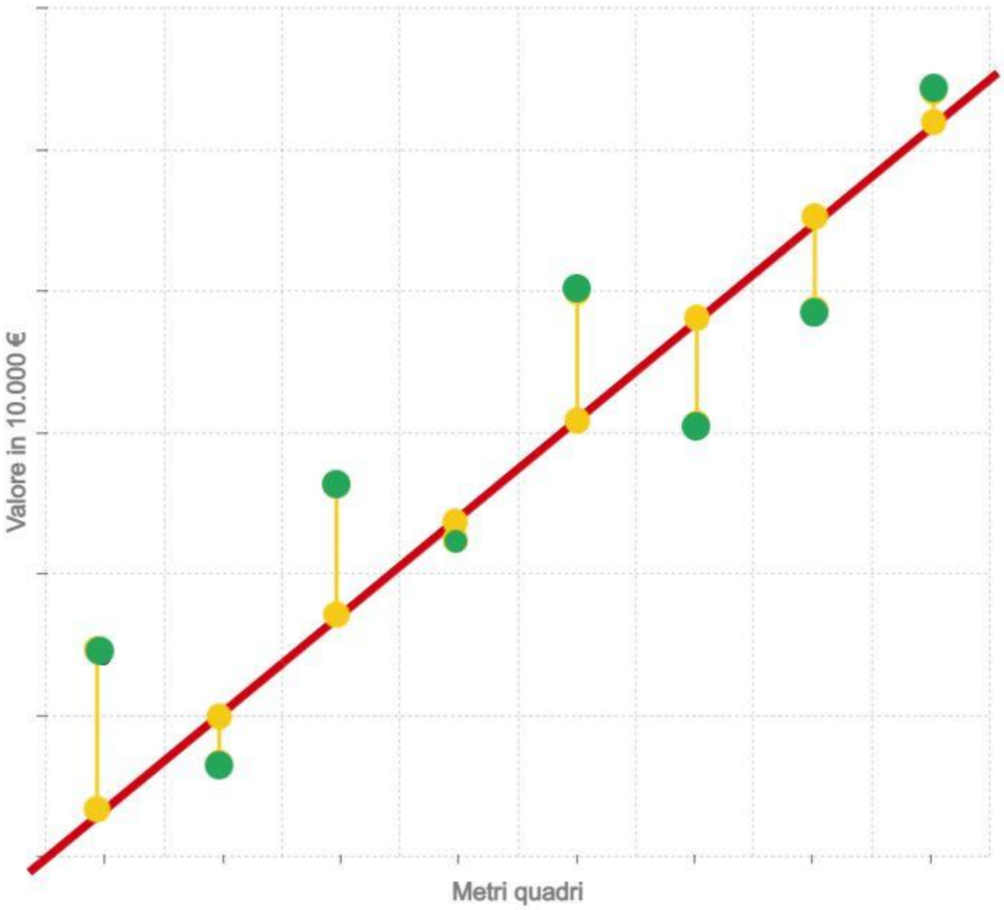
Valore corretto

Previsione del modello



Errore assoluto medio (MAE)

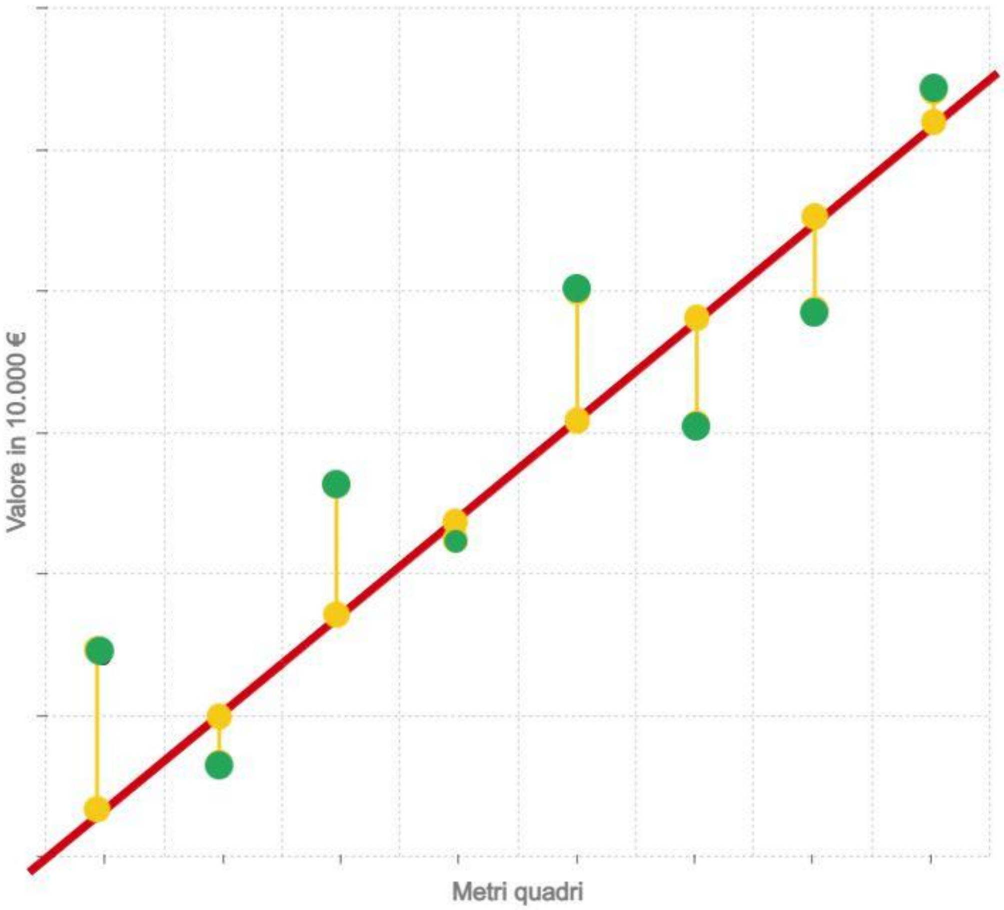
Fornisce un'indicazione di quanto mediamente il modello si è sbagliato



Errore assoluto medio (MAE)

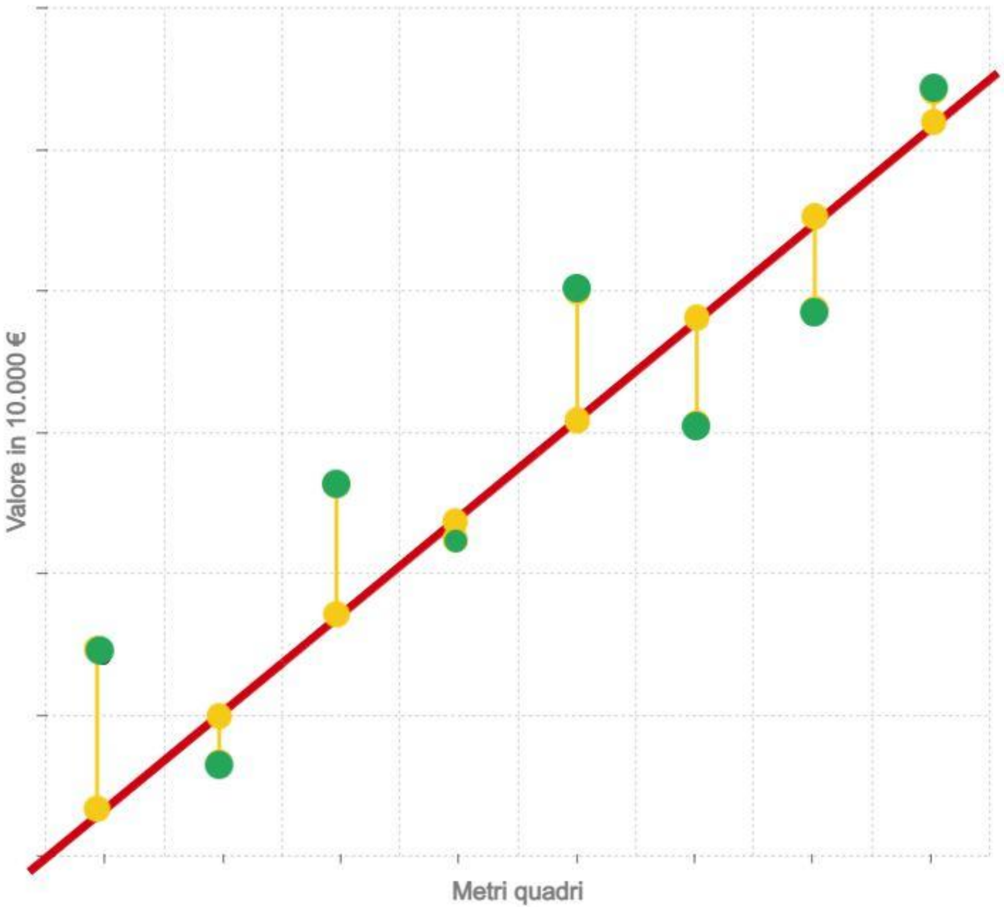
$$MAE = 2.075$$

Mediamente il modello sbaglia di 20.750 €



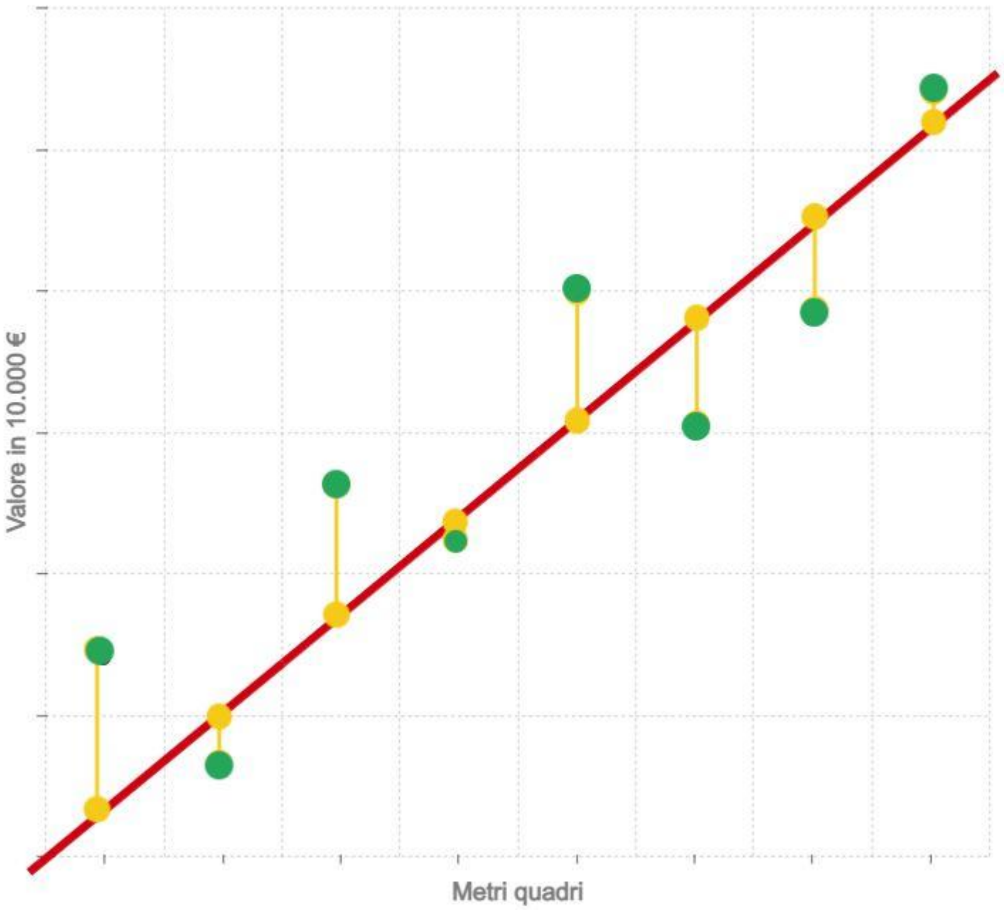
Errore quadratico medio (MSE)

Il MSE è definito come
la media della somma
dei quadrati dei residui



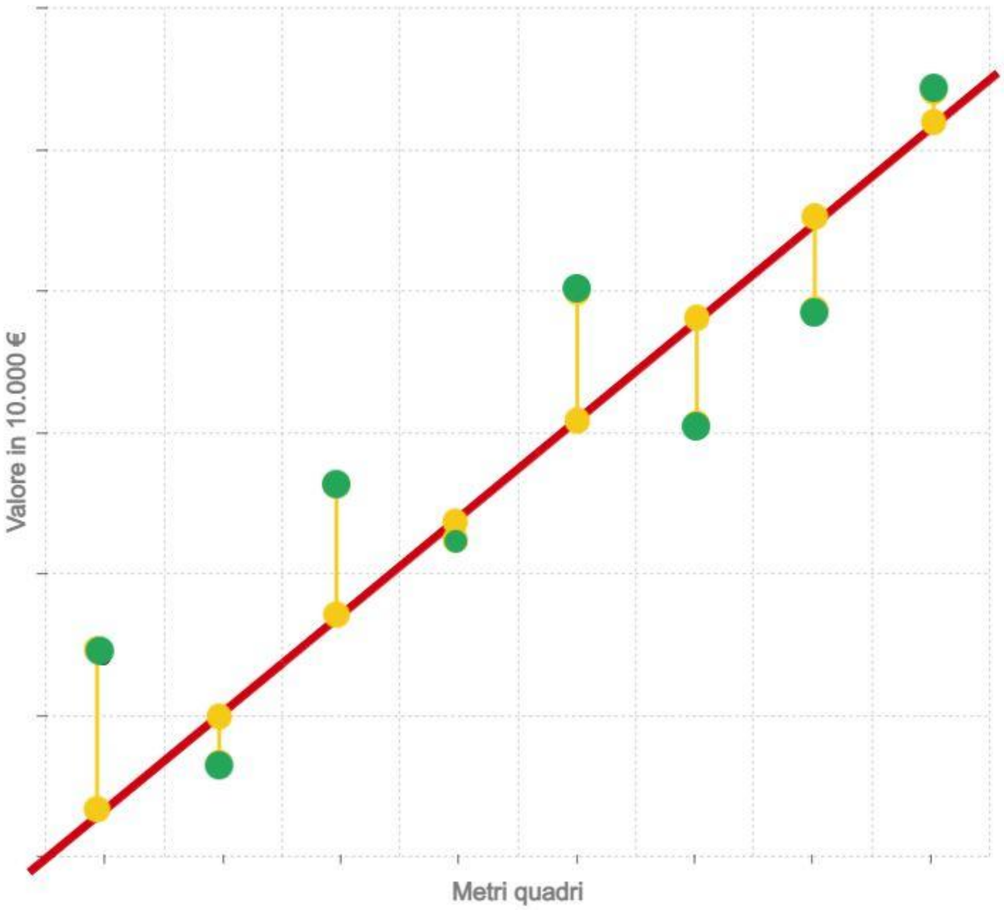
Errore quadratico medio (MSE)

$$MSE = \frac{1}{n}\sum_{i=1}^{N}\left(y_i - f(x_i)\right)^2$$



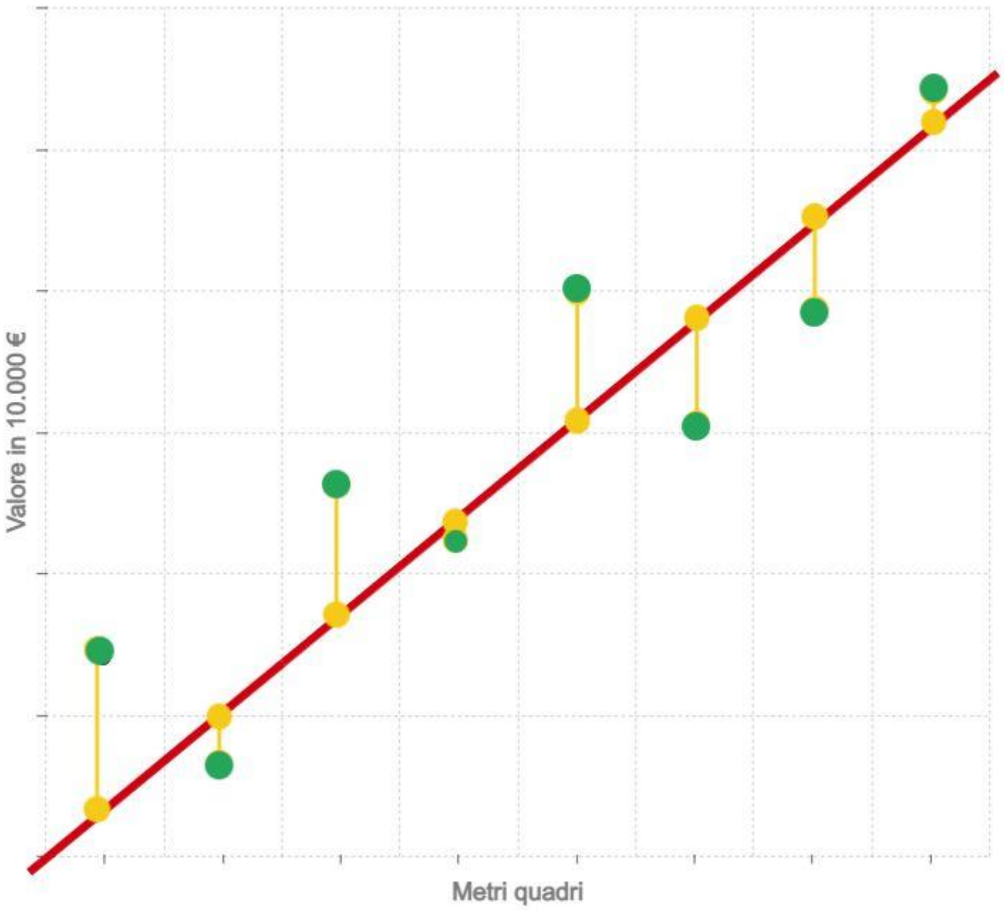
Errore quadratico medio (MSE)

$$MSE = 5.35$$



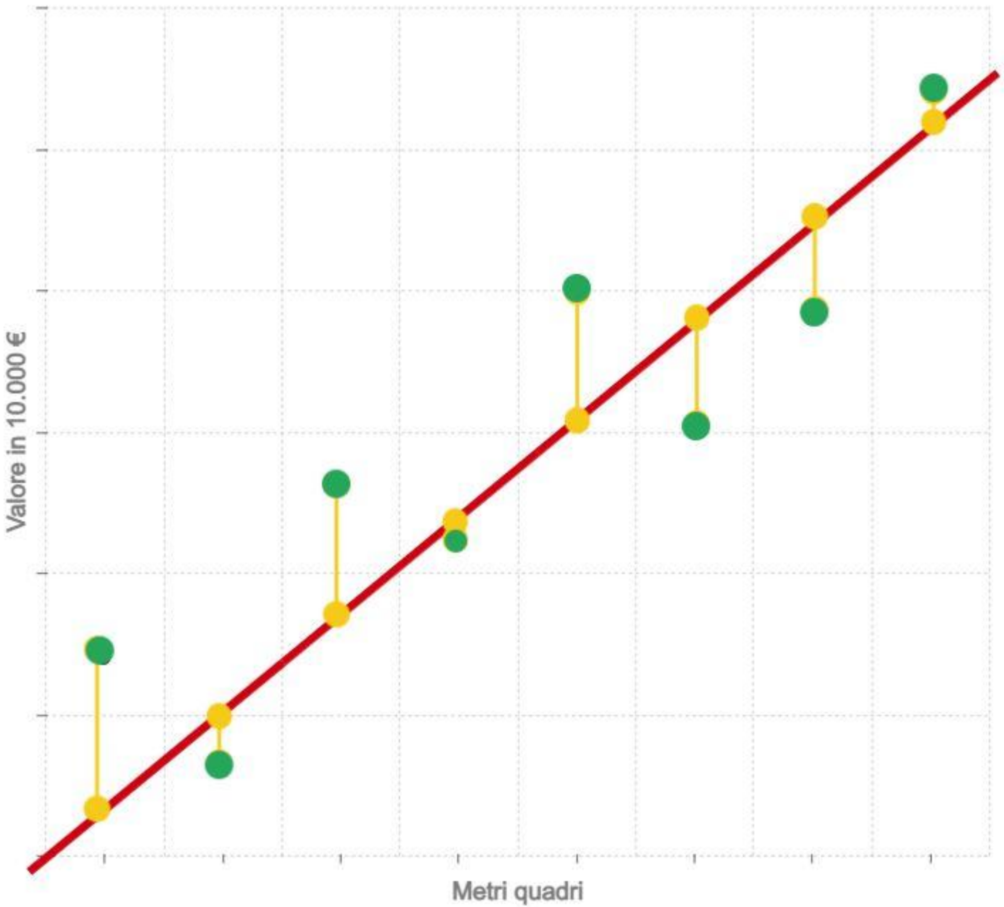
Errore quadratico medio (MSE)

Da un peso maggiore ai residui maggiori, cioè agli errori più grandi del modello.



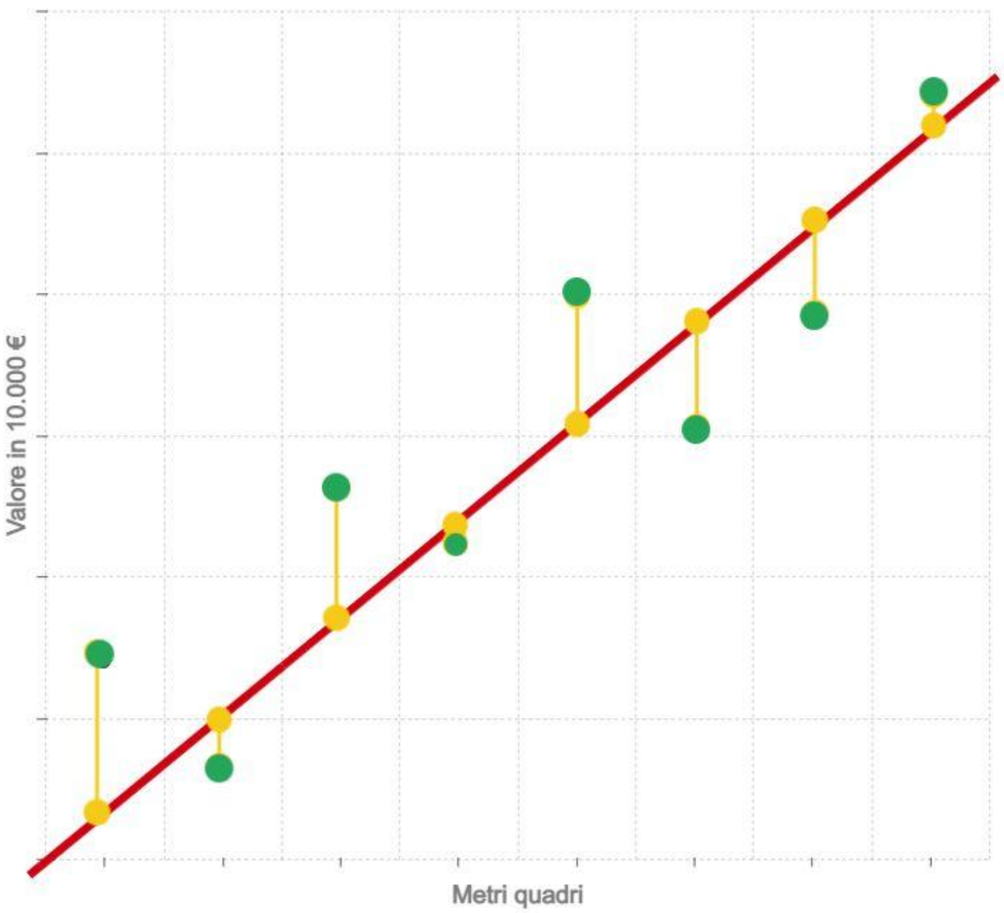
Somma dei quadrati residui (RSS)

La RSS è definita come
la somma degli errori al quadrato
per ogni punto del nostro dataset



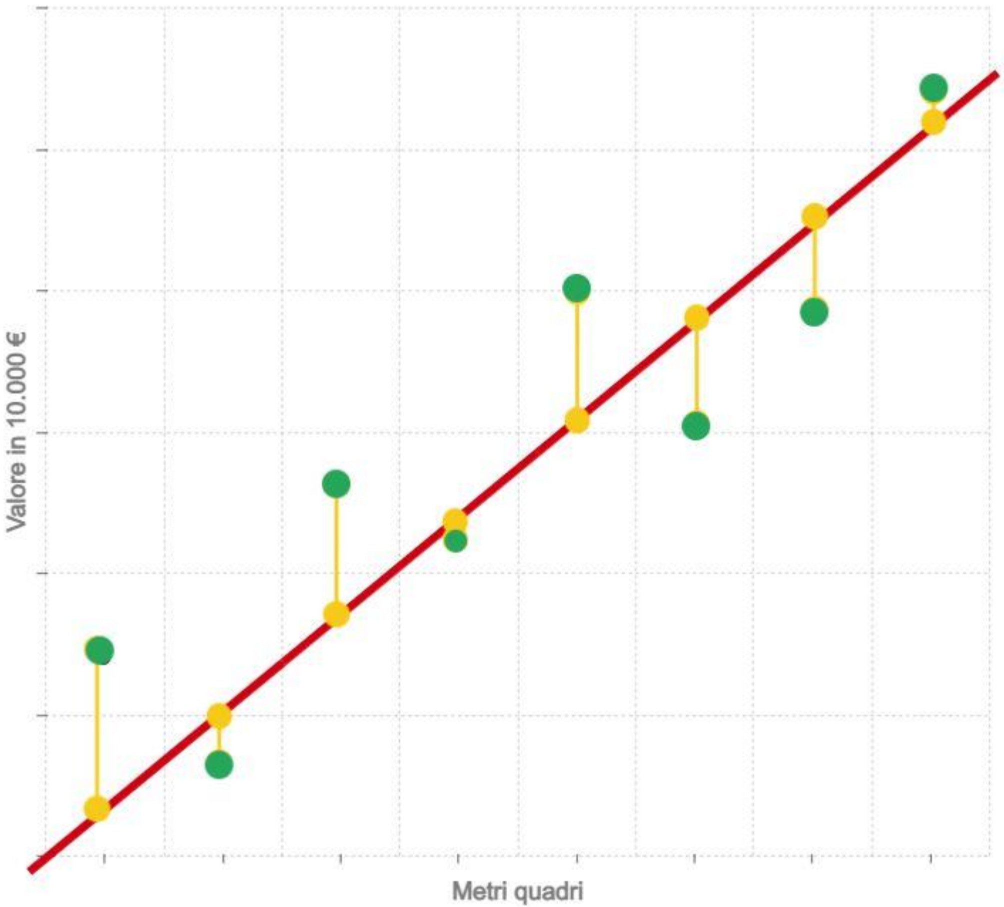
Somma dei quadrati residui (RSS)

$$RSS = \sum_{i=1}^{N} (y_i - f(x_i))^2$$



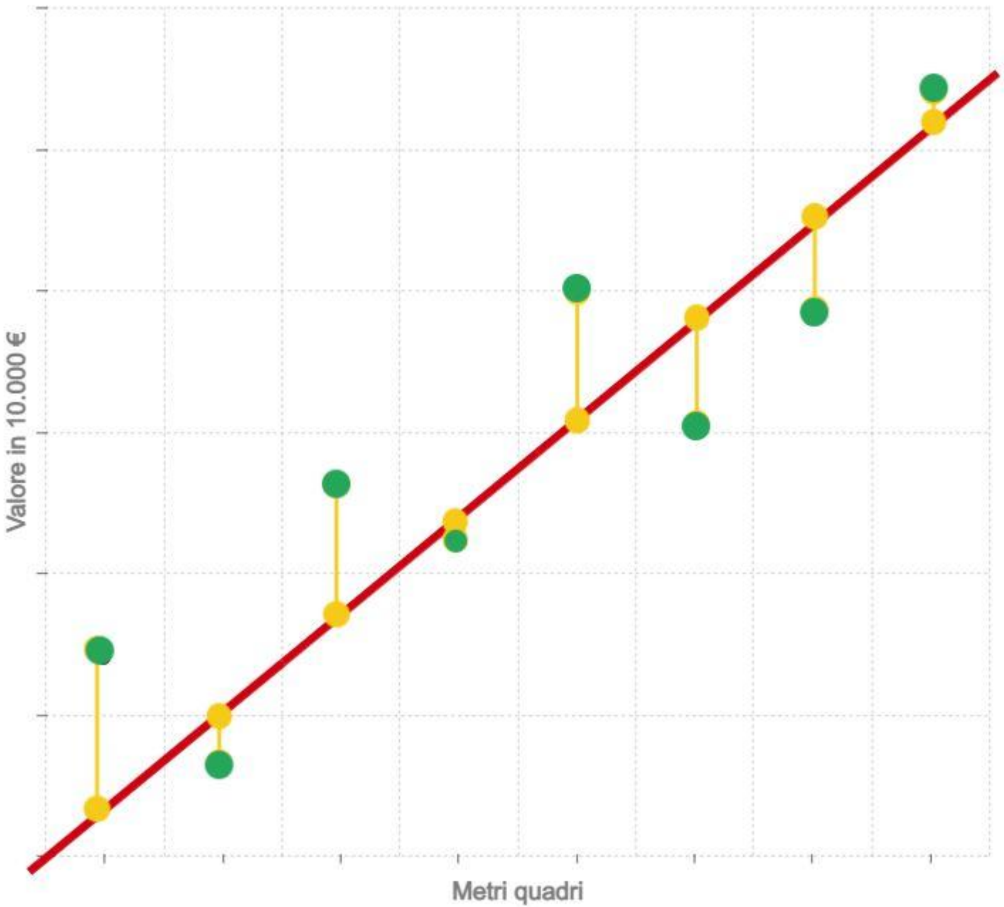
Radice dell’errore quadratico medio (RMSE)

E’ la radice quadrata del MSE.



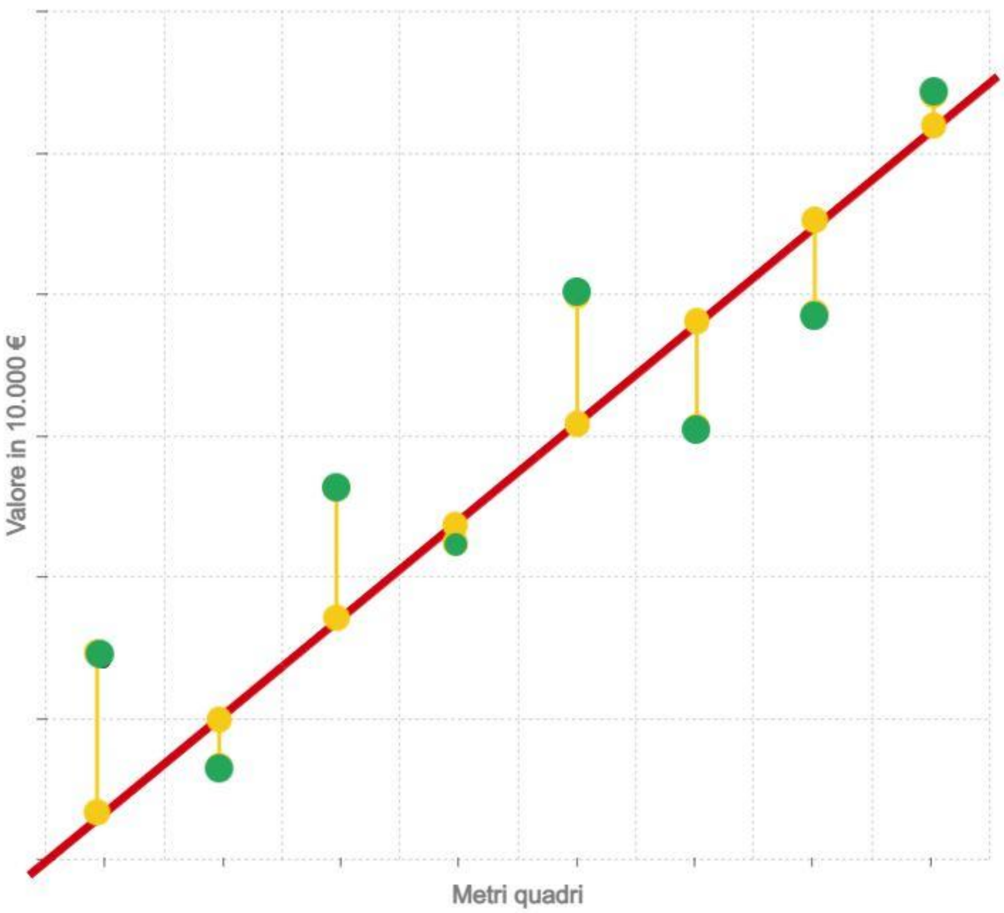
Radice dell'errore quadratico medio (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{N}(y_i - f(x_i))^2}$$



Radice dell'errore quadratico medio (RMSE)

$$RMSE = 2.31$$



Radice dell'errore quadratico medio (RMSE)

Estraendo la radice riportiamo il valore sulla stessa scala dei residui.

Fondamenti di Machine Learning

La Regressione

Il Coefficiente di Determinazione

presentato da
Giuseppe Gullo

PROFESSION AI

Le metriche studiate finora
ritornano valori che vanno sempre rapportati
alla magnitudine della variabile target.

Coefficiente di determinazione (R2)

Ritorna un valore **standardizzato.**

Le metriche studiate finora
ritornano valori che vanno sempre rapportati
alla magnitudine della variabile target.

$$MAE = 2.075 \qquad \bar{y} = 19.37$$

$$MSE = 5.35 \qquad y_{\max} = 30$$

$$RMSE = 2.31 \qquad y_{\min} = 10$$

Coefficiente di determinazione (R2)

Somma dei quadrati residui (RSS)

$$RSS = \sum_{i=1}^{N} (y_i - f(x))^2$$

Coefficiente di determinazione (R2)

Somma dei quadrati totali (SST)

$$SST = \sum_{i=1}^{N} (y_i - \bar{y})^2$$

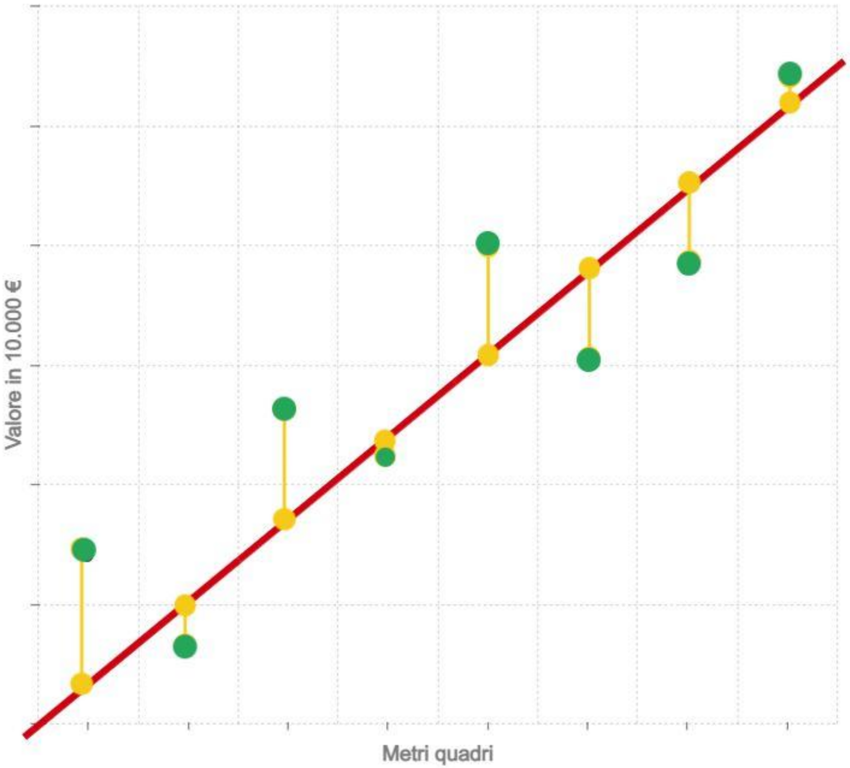
Coefficiente di determinazione (R2)

$$R^2 = 1 - \frac{RSS}{SST}$$

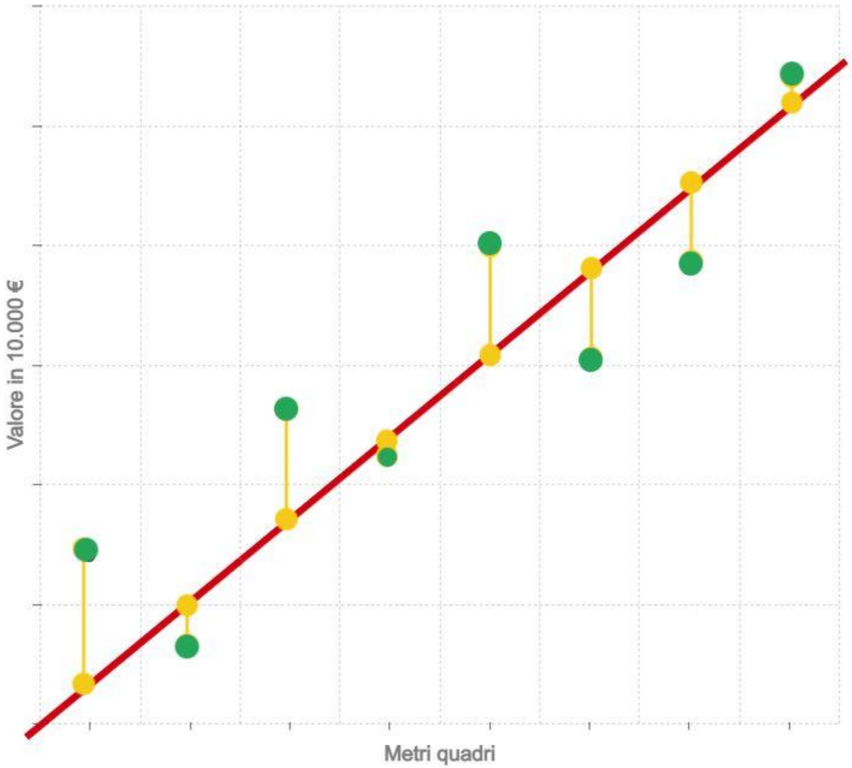
Coefficiente di determinazione (R2)

$$R^2 = 1 - \frac{RSS}{SST}$$

- **R2< 0.3** il modello è inutile.
- **0.3 < R2 < 0.5** il modello è scarso.
- **0.5 < R2 < 0.7** il modello è discreto.
- **0.7 < R2 < 0.9** il modello è buono.
- **0.9 < R2 < 1** il modello è ottimo.
- **R2 = 1** molto probabilmente c'è un errore nel modello.



x	y	previsione
80	16	18.3
150	30	29.5
30	12	10.3
50	10	13.5
120	24	24.7
60	18	15.1
110	20	23.1
110	25	23.1



x	y	previsione
80	16	18.3
150	30	29.5
30	12	10.3
50	10	13.5
120	24	24.7
60	18	15.1
110	20	23.1
110	25	23.1

$$\bar{y} = 19.37$$

x	y	previsione
80	16	18.3
150	30	29.5
30	12	10.3
50	10	13.5
120	24	24.7
60	18	15.1
110	20	23.1
110	25	23.1

$$\bar{y} = 19.37$$

$$RSS = 42.8$$

$$SST = 321.87$$

$$R^2 = 0.87$$

Fondamenti di Machine Learning

La Regressione

La Regressione Lineare Multipla

presentato da

Giuseppe Gullo

PROFESSION AI

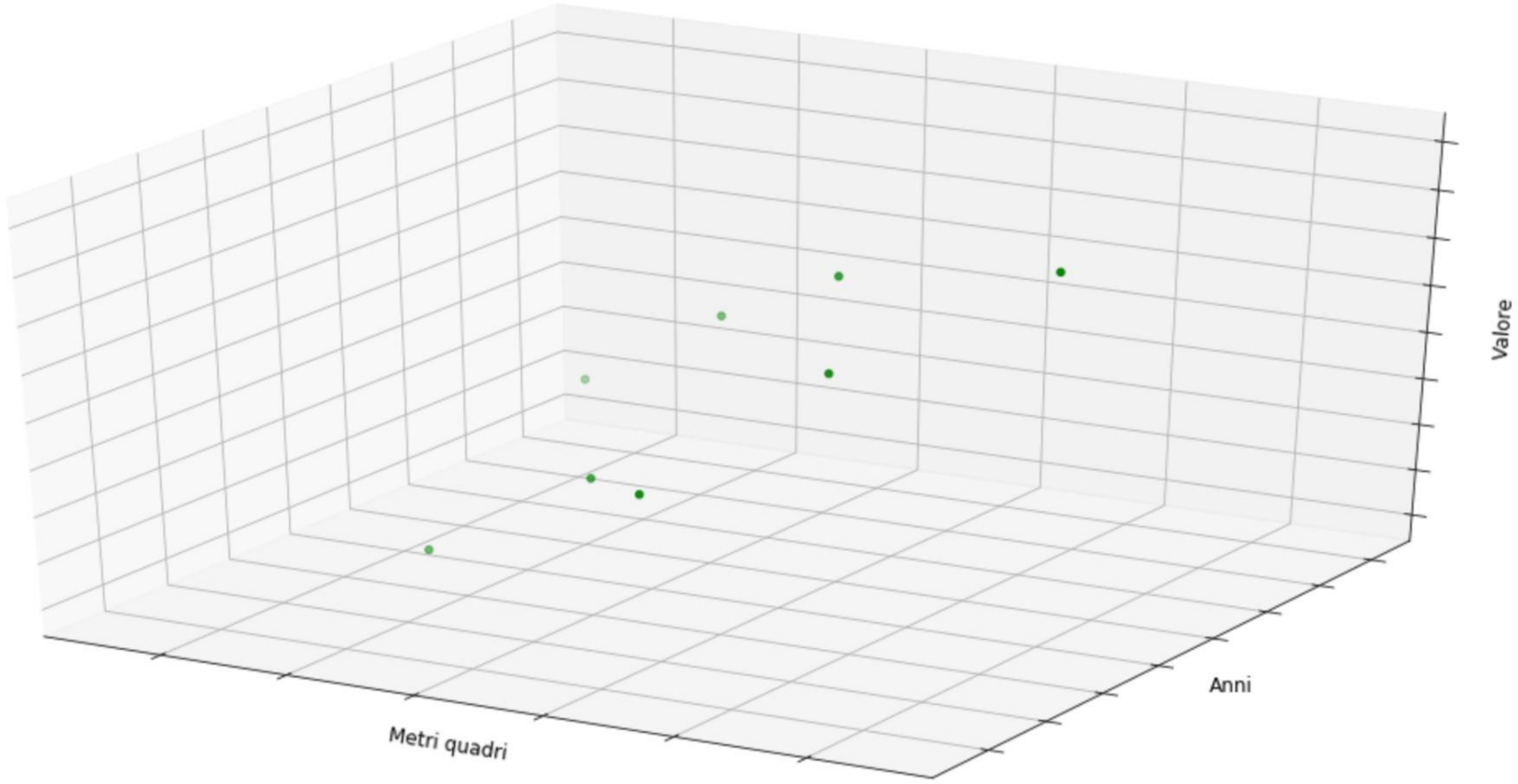
La regressione lineare multipla

Ci permette di trovare una relazione lineare tra
due o più variabili indipendenti (feature)
e una variabile dipendente (target).

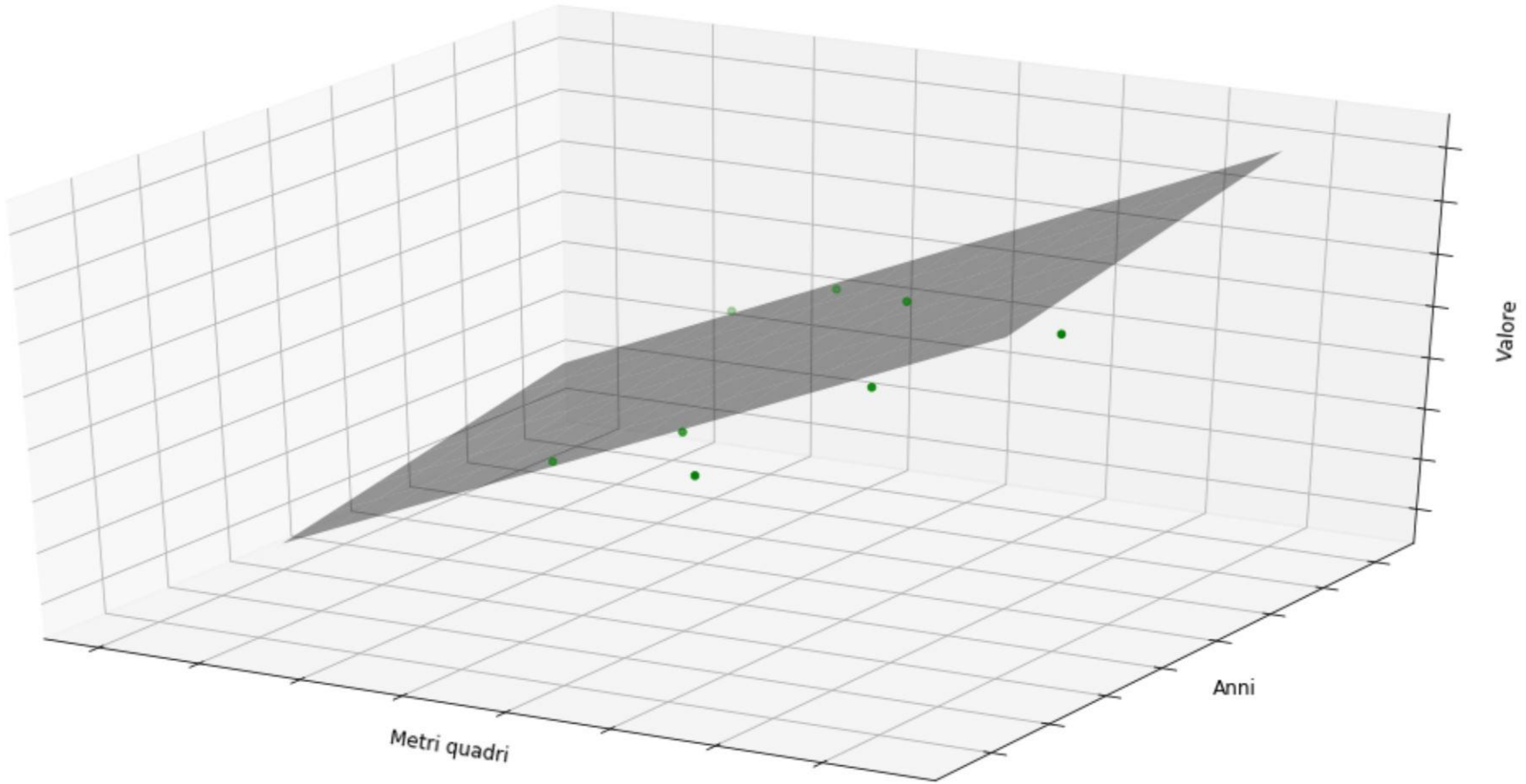
Dataset di monolocali con 2 feature

Metri quadri	Anno di costruzione	Valore in €
80	1995	160.000
150	1995	300.000
30	2008	120.000
50	1996	100.000
120	1994	240.000
60	2006	180.000
110	1989	200.000
110	2000	250.000

Dataset di monolocali con 2 feature



Dataset di monolocali con 2 features



La regressione lineare multipla

2 feature

$$y = w_1x_1 + w_2x_2 + b$$

Parametri del modello

$$w_1,\ w_2,\ b$$

La regressione lineare multipla

3 feature

$$y = w_1x_1 + w_2x_2 + w_3x_3 + b$$

Parametri del modello

$$w_1,\ w_2,\ w_3,\ b$$

La regressione lineare multipla

Visualizzare un dataset con 3 feature

La regressione lineare multipla

N feature

$$y = w_1x_1 + w_2x_2 + w_3x_3 + \ldots + w_Nx_N + b$$

Parametri del modello

$$w_1,\ w_2,\ w_3, \ldots, w_N\ , b$$

La regressione lineare multipla

Rappresentazione compatta

$$y = \sum_{i=1}^{N} w_i x_i + b$$

La regressione lineare multipla

Rappresentazione vettoriale

$$y = W \cdot X + b$$

$$W = [w_1, w_2, w_3, \ldots, w_N]$$

$$X = [x_1, x_2, x_3, \ldots, x_N]$$

La regressione lineare multipla

Rappresentazione vettoriale

$$y = \boxed{W \cdot X} + b$$

PRODOTTO SCALARE

Prodotto Scalare

Dot Product

Il prodotto scalare tra due vettori
che hanno la stessa dimensione
si esegue sommando il prodotto
degli elementi alla stessa posizione
dei due vetori

$$\begin{bmatrix} 3 & 5 & 2 & 7 & 1 \end{bmatrix} \bullet \begin{bmatrix} 4 \\ 2 \\ 6 \\ 1 \\ 5 \end{bmatrix} = 3*4+5*2+2*6+7*1+1*5 = 46$$

La regressione lineare multipla

Altra rappresentazione vettoriale

$$y = W \cdot X$$

$$W = [w_0, w_1, w_2, w_3, \ldots, w_N]$$

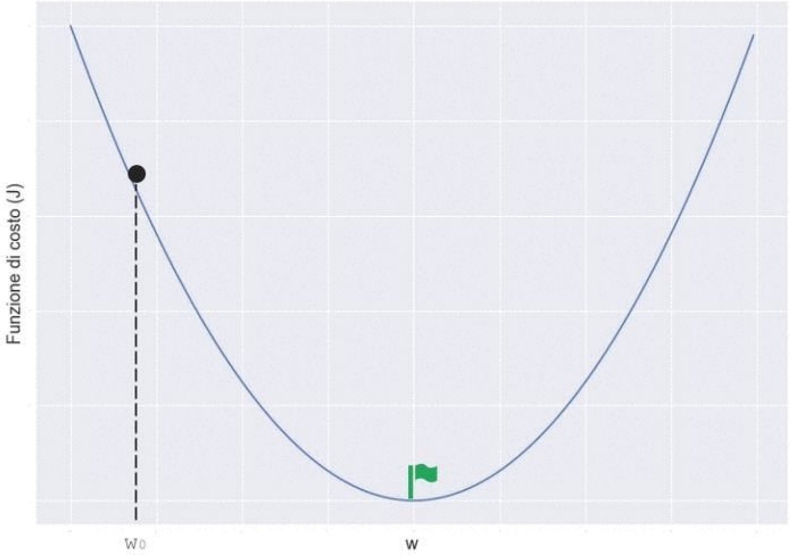
$$X = [1, x_1, x_2, x_3, \ldots, x_N]$$

Come si trovano i coefficienti di una regressione lineare multipla?

Possiamo usare il metodo dei minimi quadrati,
ma diventa computazionalmente dispendioso
al crescere del numero delle feature.

Algoritmi di ottimizzazione iterativi

Minimizzano una funzione tramite una serie di step,
in cui i parametri vengono spinti verso i valori
che permettono di ottenere il minimo.




```
Algoritmo Gradient Descent

w = rand()
```

Algoritmi di ottimizzazione iterativi

Minimizzano una funzione tramite una serie di step,
in cui i parametri vengono spinti verso i valori
che permettono di ottenere il minimo.

Gradient Descent

Stochastic Gradient Desecent

ADAM

Adadelta

RMSprop

Adagrad

Adadelta

Fondamenti di Machine Learning

La Regressione

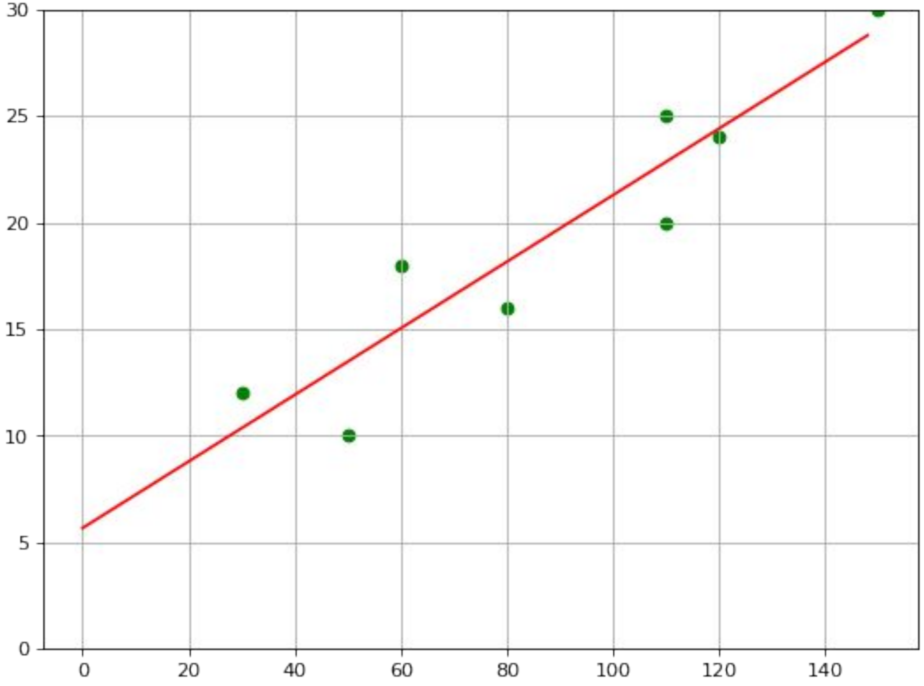
La Regressione Polinomiale

presentato da
Giuseppe Gullo

PROFESSION AI

La regressione lineare polinomiale

Ci permette di trovare una **relazione non lineare** tra
una o più variabili indipendenti (feature)
e una variabile dipendente (target)
tramite un **polinomio**.

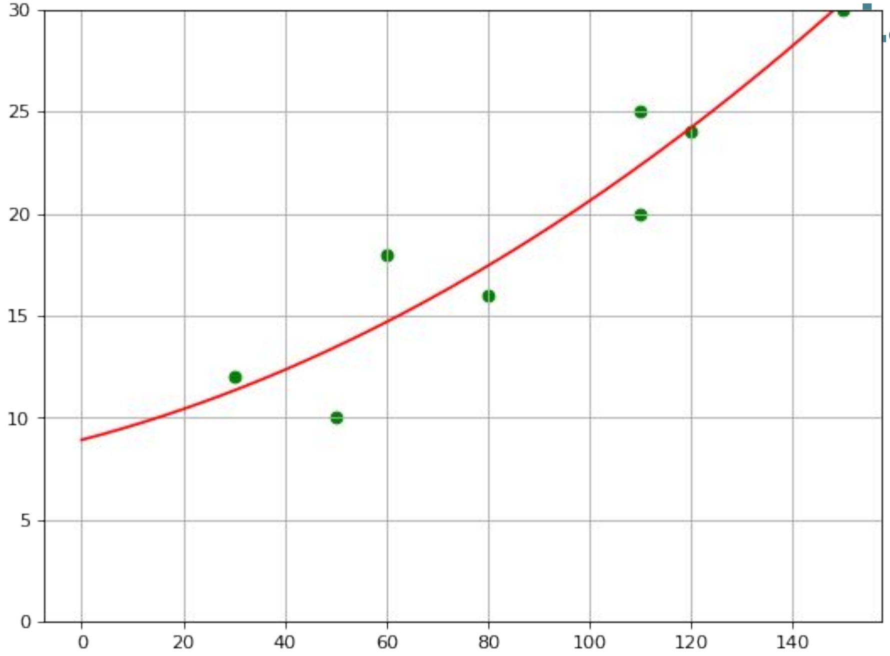


La regressione lineare semplice

Polinomio di primo grado

$$y = wx + b$$

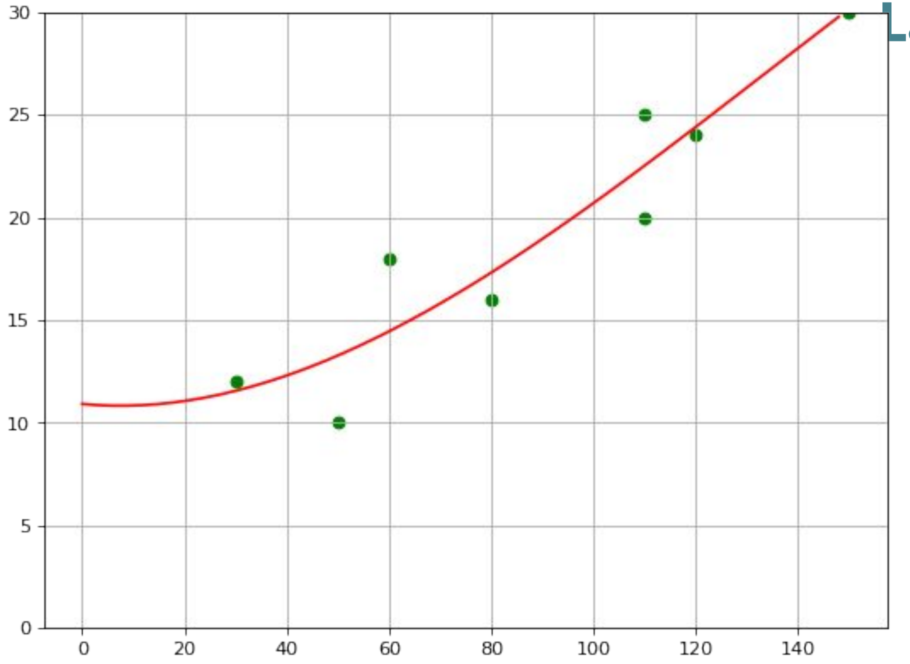
La regressione lineare polinomiale



Polinomio di secondo grado

$$y = w_1 x + w_2 x^2 + b$$

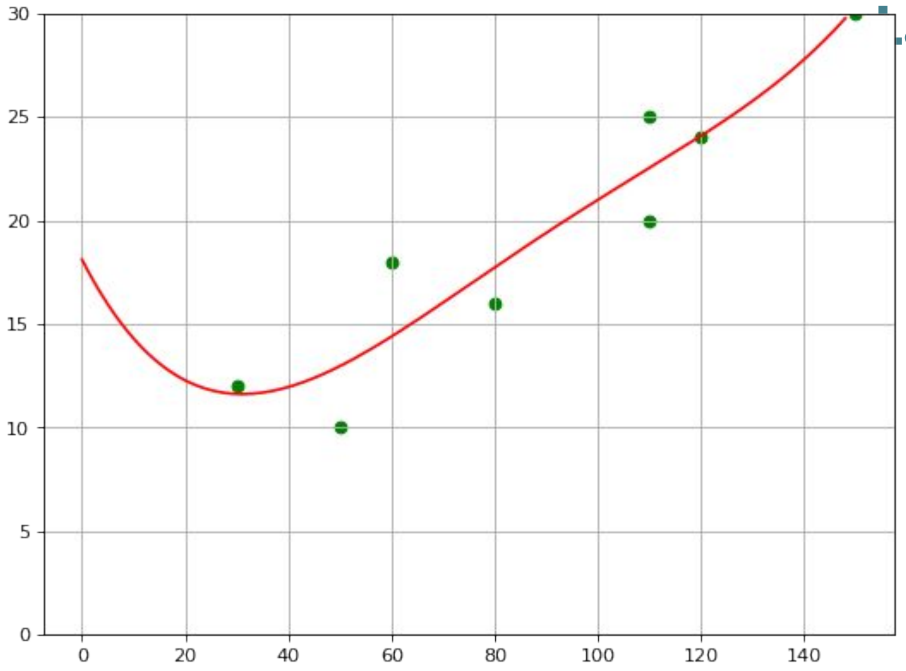
La regressione lineare polinomiale



Polinomio di terzo grado

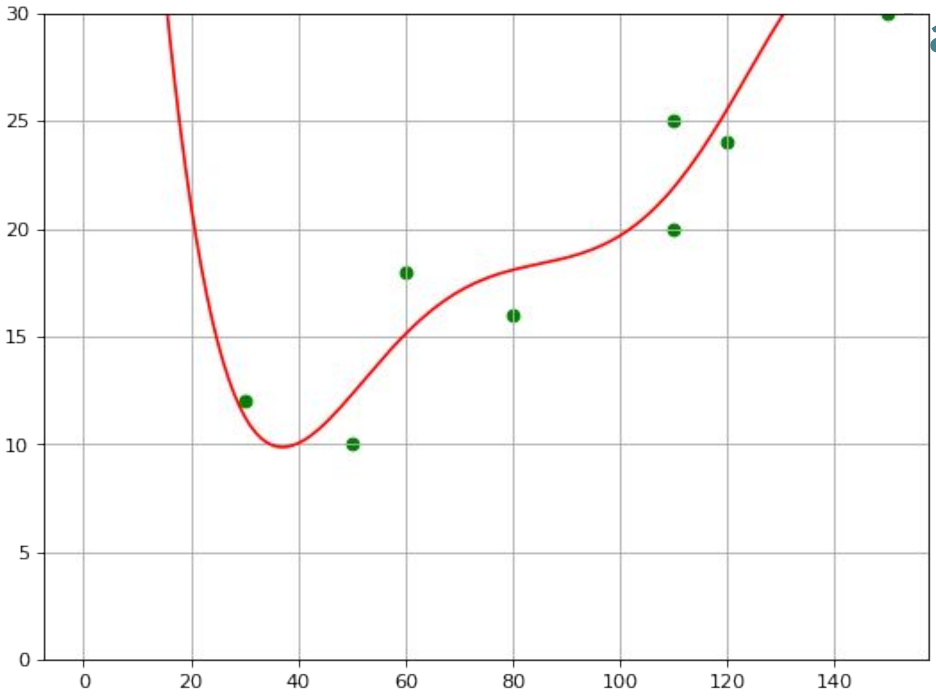
$$y = w_1 x + w_2 x^2 + w_3 x^3 + b$$

La regressione lineare polinomiale



Polinomio di quarto grado

$$y = w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4 + b$$

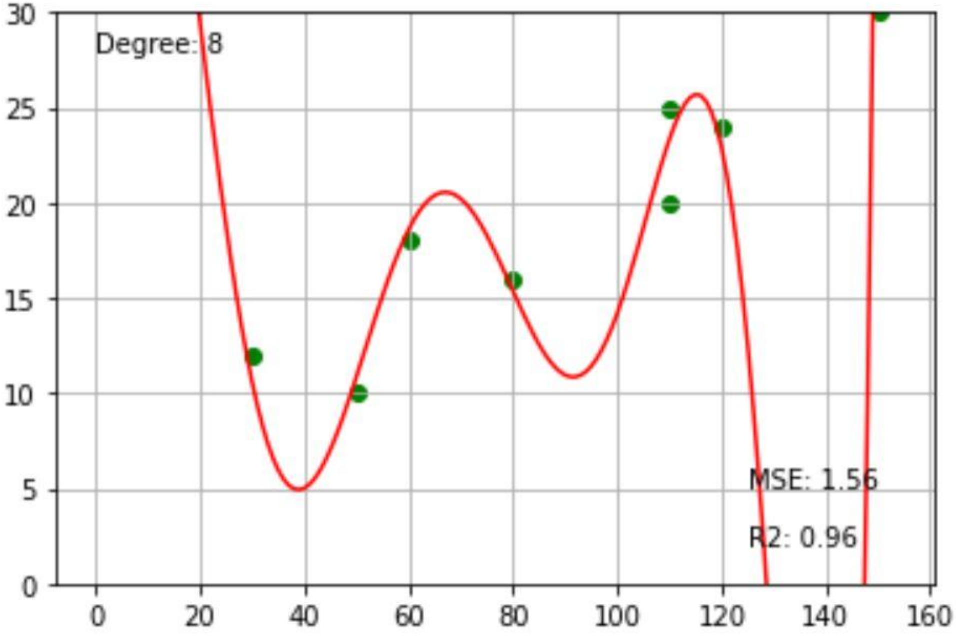


a regressione lineare polinomiale

Polinomio di quinto grado

$$y = w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4 + w_5 x^5 + b$$

a regressione lineare polinomiale



Polinomio di ottavo grado

$$y = \sum_{i=1}^{8} w_i x_i + b$$

Attenzione

Anche se la relazione è non lineare,
il modello polinomiale è comunque lineare.

$$y = w_1 x + w_2 x^2 + b$$

Attenzione

Anche se la relazione è non lineare,
il modello polinomiale è comunque lineare.

$$y = w_1 x + w_2 x^2 + b$$

I coefficienti del modello sono sempre di grado 1.

Attenzione

Anche se la relazione è non lineare,
il modello polinomiale è comunque lineare.

$$y = w_1 x + w_2 x^2 + b$$

Stiamo creando solo nuove features,
che vengono chiamate **features polinomiali**.