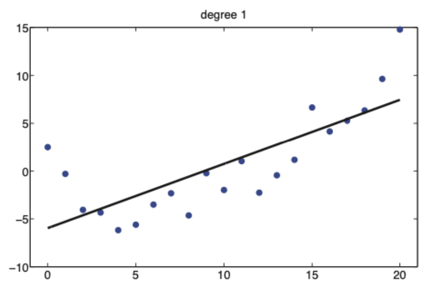
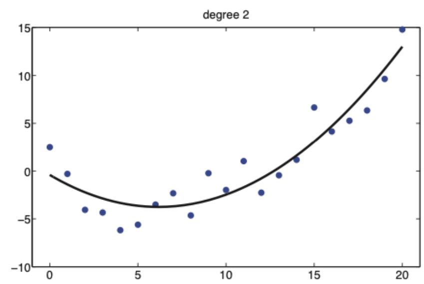
RESESSIONE LINEARE

venerdì 11 ottobre 2024 16:49

LA REGRESSIONE E' PROPRIO COME LA CLASSIFICAZIONE SOLO CHE NELLA REGRESSIONE E' CONTINUA

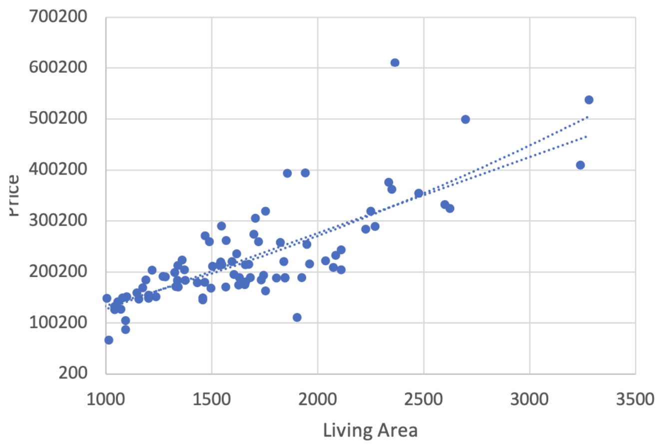


Regressione lineare (a)



Regressione Polinomiale (b)

DATO UN GRAFICO VOGLIAMO PREVEDERE UN VALORE → $x \in \mathbb{R}$



x Living area (feet2)	y Price (1000$s)
1656	215
896	105
1329	172
2110	244
...	...

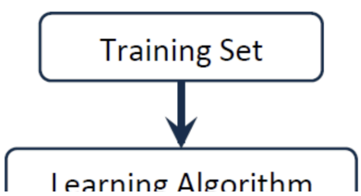
L'ESEMPIO E' IL SEGUENTE (x, y) = COPPIA DI VALORI

m = NUMERI DI ESEMPI DI FORMAZIONE

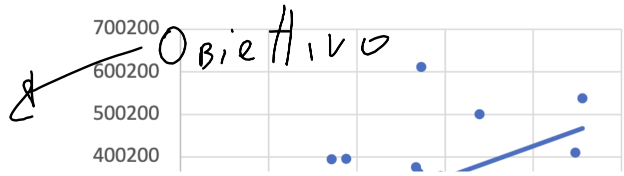
x = INPUT O FEATURE

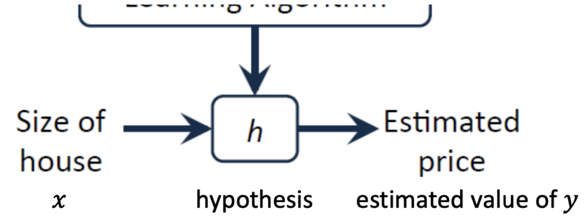
y = OUTPUT O TARGET

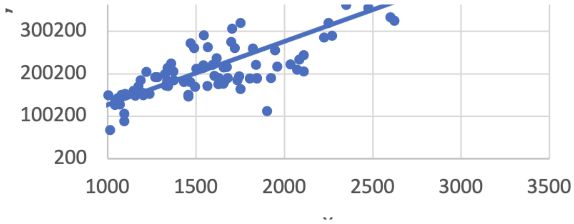
GENERALIZZAZIANDO LA COPPIA → $(x^i, y^i) = i^{th}$



Obiettivo







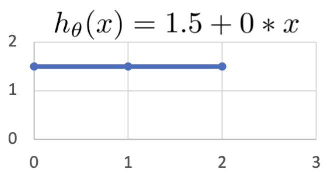
$$\underbrace{h_\theta(x) = \theta_0 + \theta_1 x \triangleq h(x)}_{\text{FUNZIONE IPOTESI}}$$

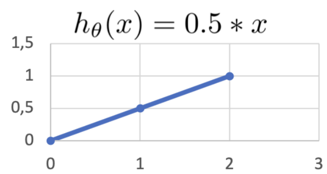
LA DOMANDA E' LA SEGUENTE COME FACCIAMO A CAPIRE CHE θ ?

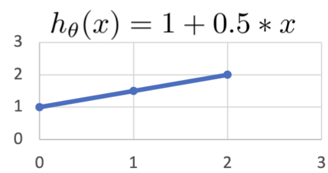
θ_i = Weights / peso

Se $\boxed{h(x)} = \theta_0 + \theta_1 X$ SOSTITUIAMO:

θ_1 COEFFICIENTE ANGOLARE RETTA







COST FUNCTION: INTUITION

FUNZIONE DI COSTO QUADRATICO

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \Big(\overbrace{h_\theta(x^{i})}^{\text{funzione predittiva}} - y^{(i)} \Big)^2$$

m numero degli esempio di allenamento.
Y = valore associato di x (VALORE DELLA CASA)
IL NOSTRO OBIETTIVO
è trovare i valori di θ_1 e θ_2 che minimizzano la funzione di costo:

$$\theta_{min} = \arg\min_{\theta} J(\theta)$$

TALE FUNZIONE DI COSTO SOMIGLIA MOLTO A QUELLA DELLA MSE EQUAZIONE DI ERRORE QUADRATICO MEDIO

$$MSE = \frac{1}{m} \sum_{1}^{m} \left(y_1 - y_2 \right)^2$$

IN DEFINITIVA

#training examples

$$\underset{\theta_0,\theta_1}{argmin} \frac{1}{2m} \sum_{i=1}^{m} \overbrace{(h_\theta(\mathbf{x^{(i)}}) - y^{(i)})^2}^{\text{Squared error function}}$$

$$h_\theta(\mathbf{x^{(i)}}) = \theta_0 + \theta_1 x^{(i)}$$

$$\underset{(CostFunction)}{J(\theta_0, \theta_1)} = J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(\mathbf{x^{(i)}}) - y^{(i)})^2$$

$$\theta_{min} = argmin_\theta(J(\theta_0, \theta_1)) = argmin_\theta(J(\theta)).$$

1.5.1 LR Cost function and some considerations

Defining a cost function is important because we need to estimate how much accurate are our prediction $h_\theta(x) = \theta_0 + \theta_1 x$ compared to the real value y.

Coefficients θ randomly choosed

The easiest approach consist of choosing random coefficients θ hoping to find a good approximation.

Intuition of LR

We search for the minimum parameters such that the average prediction $h_\theta\left(x^{(i)}\right)$ is as much close to the real values $y^{(i)}$. The problem is defined as it follows:

$$\theta^* = \underset{\theta_0,\theta_1}{\operatorname{argmin}} \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2 = \underset{\theta_0,\theta_1}{\operatorname{argmin}} J(\theta_0, \theta_1) = \underset{\theta}{\operatorname{arcmin}} J(\theta)$$

Where every prediction is defined as $h_\theta\left(x^{(i)}\right) = \theta_0 + \theta_1 x^{(i)}$.

The cost function is defined as:

$$J(\theta) = J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

Nota: this form or error is called **Mean Square Error (MSE)**