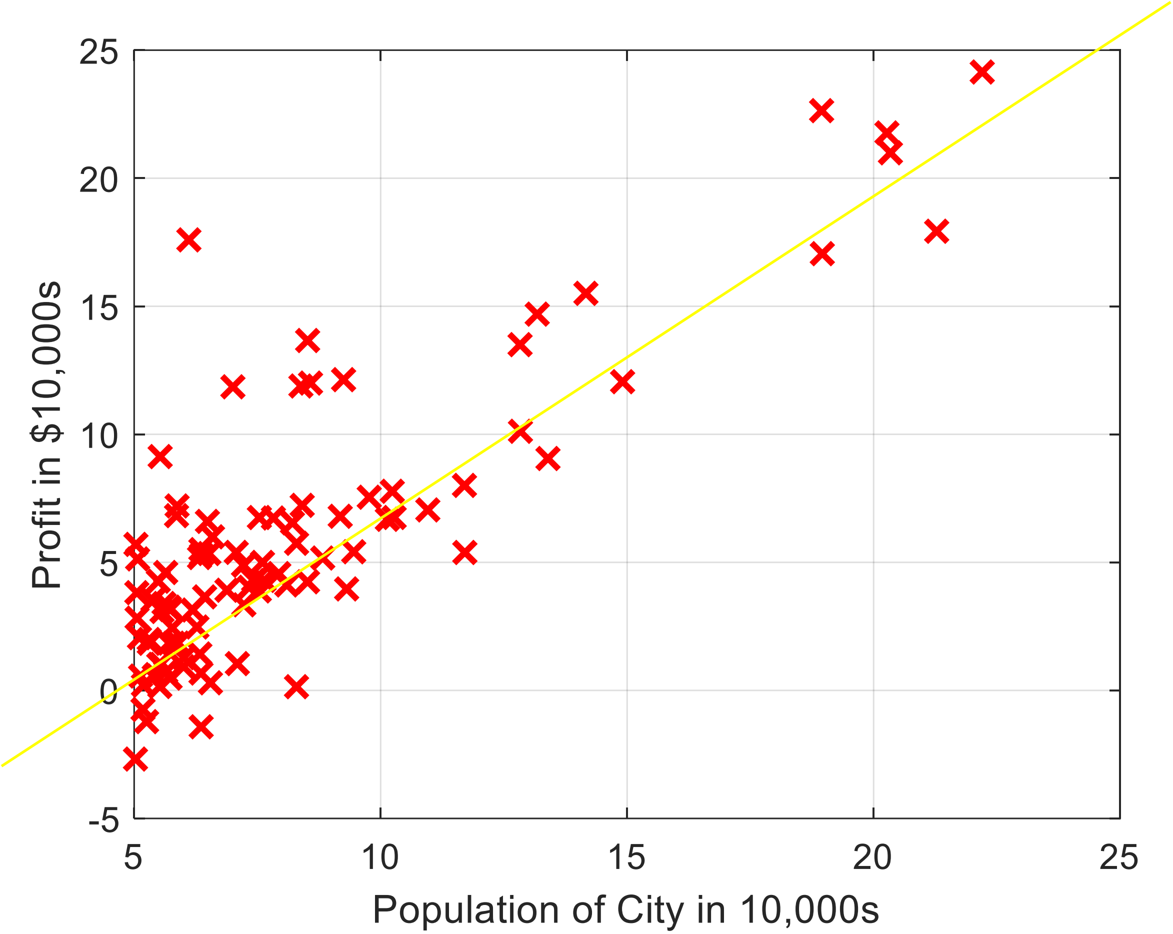
Esercizio 1: Stima dei profitti di un ristorante

Problema: il CEO di un franchising di ristoranti che sta valutando diverse città per l'apertura di un nuovo ristorante.

La catena ha già ristoranti in varie città e sono disponibili dati di profitti e di popolazione di questa città.

L'obiettivo è utilizzare questi dati per selezionare in quale città aprire la nuova attività



- Ogni città è descritta da:
  - ✓ φ_1: Popolazione [in 10000 unità]

✓ L'output y è il profitto [in 10000$]

- Il dataset consiste di $N = 97$ città con $\varphi_1(i)$, e $y(i)$, per $i = 1, \dots, N$

Calcolare e implementare il gradiente

Come calcoliamo il gradiente? Supponiamo che il nostro modello sia

$$\boxed{y = \theta_0 + \theta_1 \cdot \varphi + \epsilon}$$

$$J(\theta_0, \theta_1) = \frac{1}{N} \sum_{i=1}^{N} \big(y(i) - \theta_0 - \theta_1 \cdot \varphi(i)\big)^2 \quad \Rightarrow \quad \underset{2 \times 1}{\nabla J(\theta_0, \theta_1)} = \left[\frac{\partial J(\theta_0, \theta_1)}{\partial \theta_0} \quad \frac{\partial J(\theta_0, \theta_1)}{\partial \theta_1}\right]^\top$$

$$\frac{\partial J(\theta_0, \theta_1)}{\partial \theta_0} = \frac{2}{N} \sum_{i=1}^{N} \big(y(i) - \theta_0 - \theta_1 \cdot \varphi(i)\big) \cdot (-1) = -\frac{2}{N} \underset{1 \times N}{X(:,1)^\top} \cdot \underset{N \times 1}{(Y - X\boldsymbol{\theta})}$$

$$\frac{\partial J(\theta_0, \theta_1)}{\partial \theta_1} = \frac{2}{N} \sum_{i=1}^{N} \big(y(i) - \theta_0 - \theta_1 \cdot \varphi(i)\big) \cdot \big(-\varphi(i)\big) = -\frac{2}{N} \underset{1 \times N}{X(:,2)^\top} \cdot \underset{N \times 1}{(Y - X\boldsymbol{\theta})}$$

Attività 1 - Regressione Lineare

es 1: φ_1 : popolazione [in 10'000 unità] ; Y : profitto [in 10'000 \$] } Dati , $N = 97$ città

modello lineare

$$Y(i) = \theta_0 + \theta_1 \varphi_1 + \varepsilon$$

Funzione di costo

$$J(\bar{\theta}) = \frac{1}{N} \Sigma \left(Y(i) - \theta_0 - \theta_1 \varphi_1 \right)^2$$

$$J(\theta_0, \theta_1) = \frac{1}{N} \Sigma \left(Y(i) - \theta_0 - \theta_1 \cdot \varphi(i) \right)^2$$

Gradiente

$$\nabla J(\theta_0, \theta_1) = \left[\frac{dJ(\theta_0,\theta_1)}{d\theta_0} \quad \frac{dJ(\theta_0,\theta_1)}{d\theta_1} \right]^T$$

$$\frac{dJ(\theta_0,\theta_1)}{d\theta_0} = \frac{2}{N} \Sigma \left(Y(i) - \theta_0 - \theta_1 \cdot \varphi_1(i) \right) \cdot (-1) = -\frac{2}{N} \Sigma \left(Y(i) - X\bar{\theta} \right)$$

$$\frac{dJ(\theta_0,\theta_1)}{d\theta_1} = \frac{2}{N} \Sigma \left(Y(i) - \theta_0 - \theta_1 \cdot \varphi_1(i) \right) \cdot \left(-\varphi(i) \right) = -\frac{2}{N} X(:,2)^T (Y - X\bar{\theta})$$

con $X = \begin{bmatrix} 1 & \varphi_1(1) \\ \vdots & \vdots \\ 1 & \varphi_1(N) \end{bmatrix}$ e $\bar{\theta} = [\theta_0 \ \ \theta_1]$ e $Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_N \end{bmatrix}$

stima

$$\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} - \alpha \cdot \nabla J(\bar{\theta}) \Big|_{\theta = \hat{\theta}^{(k)}}$$

con α = learning rate

Gradient descent ... in MatLab

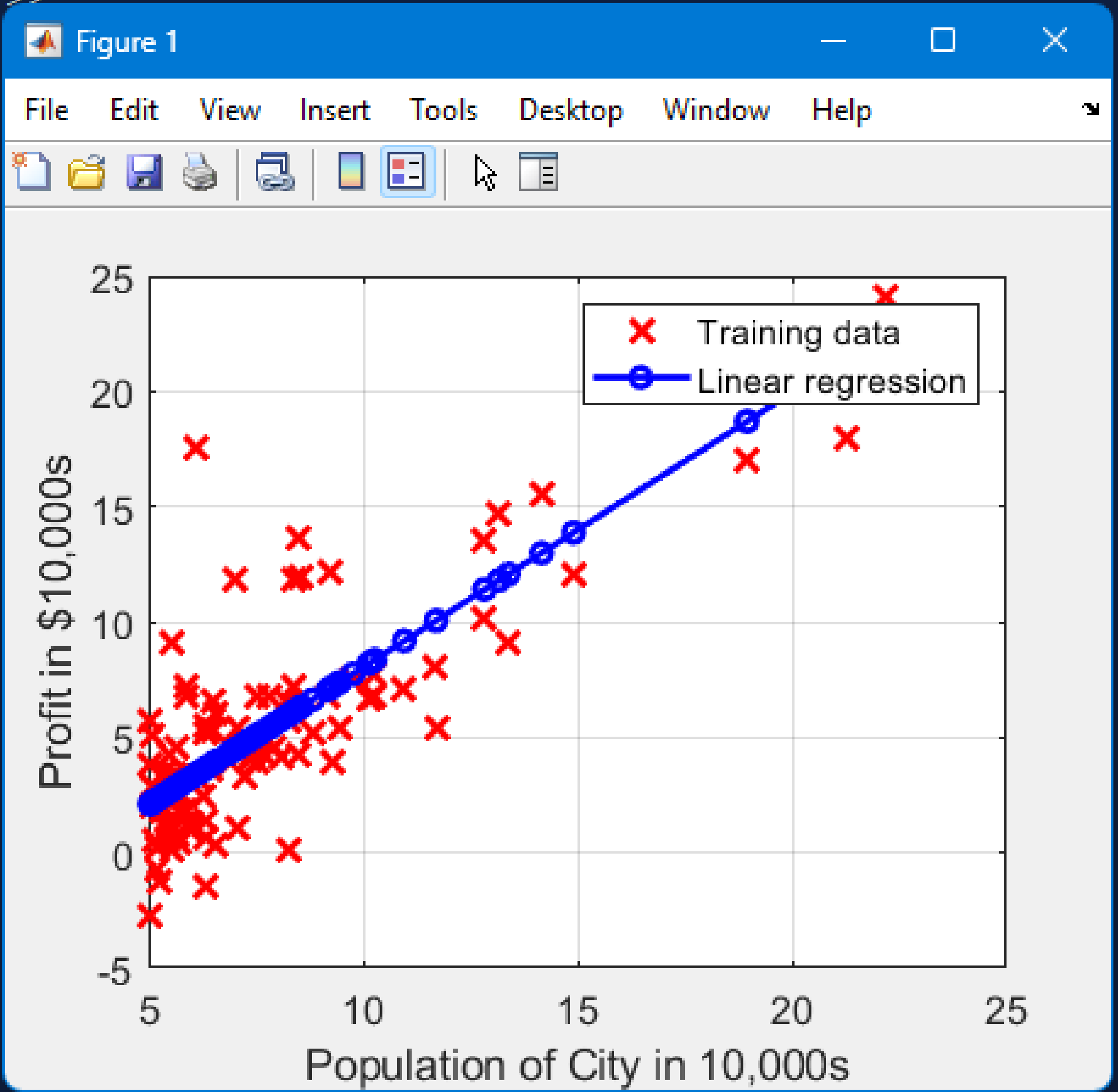
? quanto vale la stima del vettore $\hat{\theta}$ ⟶ $\hat{\theta}_0 = -3{,}87$, $\hat{\theta}_1 = 1{,}$

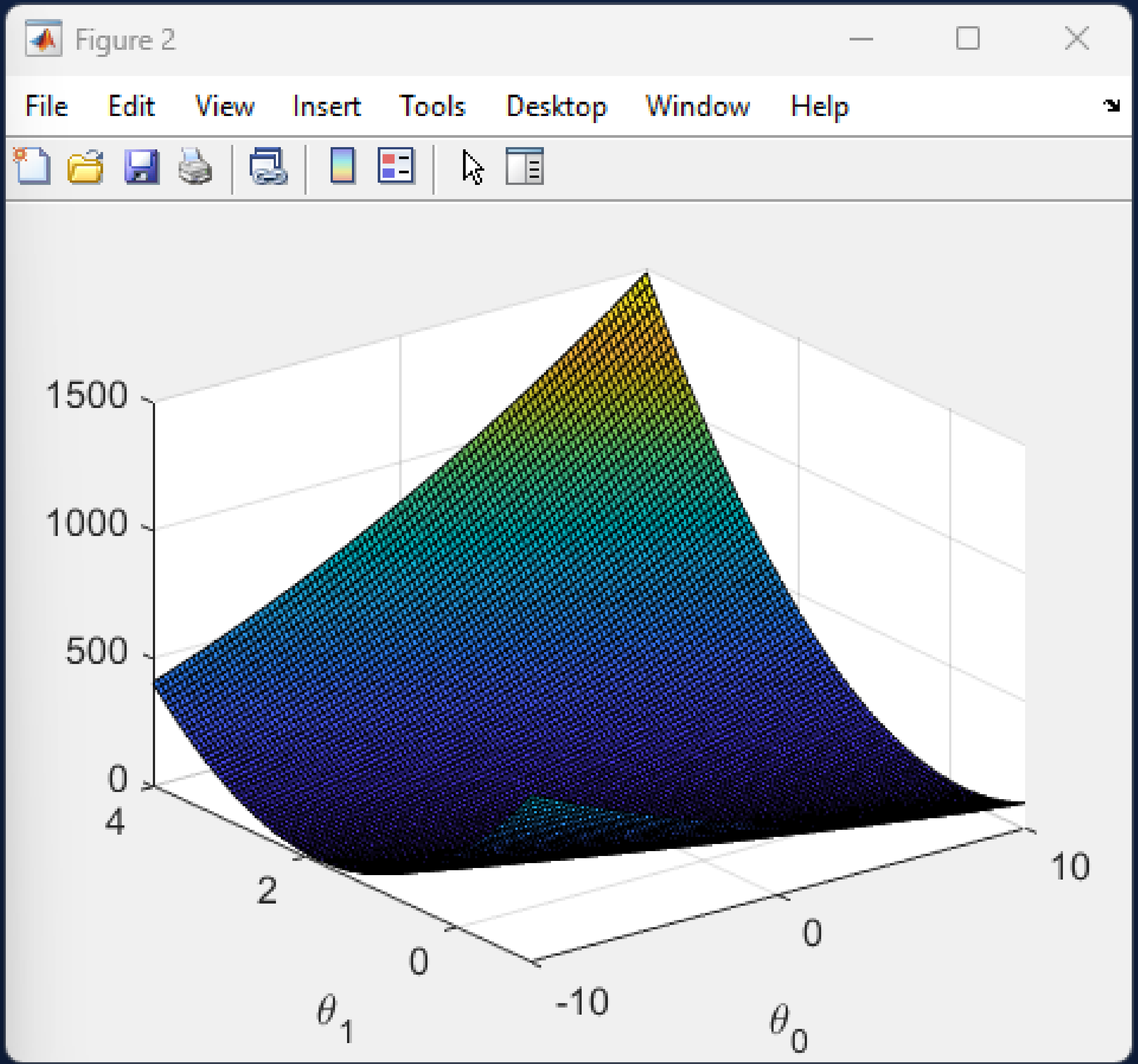
? qual è il profitto predetto per population = 35 ⟶ $\hat{Y}(\varphi_1 = 35) = +2912{,}76$ \$

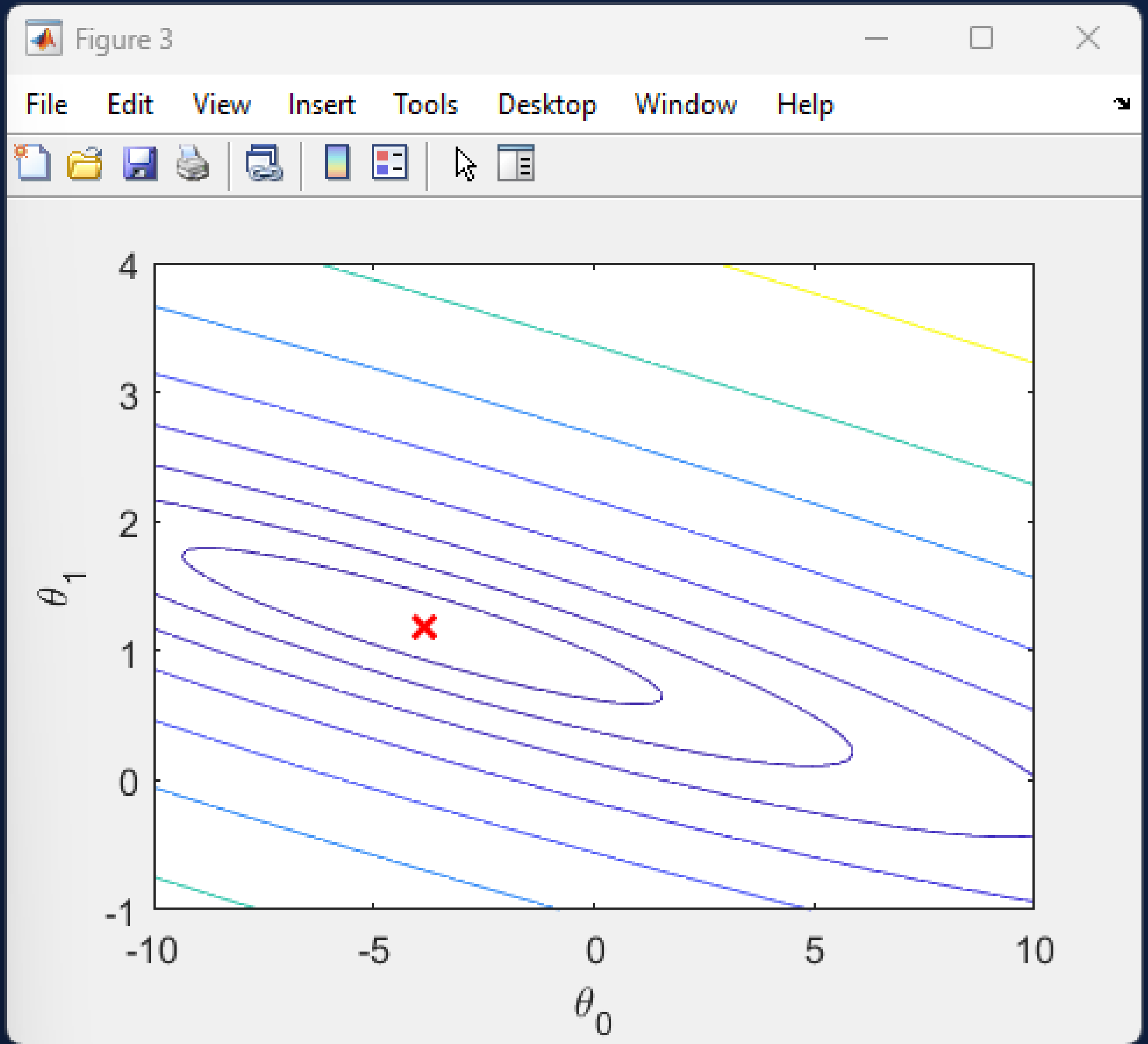
Compute cost (Funzione di costo)

$$J(\bar{\theta}) = \frac{1}{N} \Sigma \left(Y(i) - \theta_0 - \theta_1 \cdot \varphi(i) \right)^2 = \frac{1}{N} (Y - X\bar{\theta})^T (Y - X\bar{\theta})$$

```
Running Gradient Descent ...
Theta found by gradient descent: -3.878138 1.191261
For population = 35,000, we predict a profit of 2912.764904
Visualizing J(theta_0, theta_1) ...
>>
```







Esercizio 2: stima dei prezzi delle case

Vogliamo **stimare il prezzo** delle case a Portland, Oregon. L'output y è quindi il prezzo

- Ogni casa è descritta da:
  - ✓ φ_1: Area $[\text{feet}^2]$
  - ✓ φ_2: Numero di camere da letto

- Il dataset consiste di $N = 47$ case con $\varphi_1(i), \varphi_2(i)$ e $y(i)$, per $i = 1, \dots, N$

$$y(i) = \boldsymbol{\varphi}^\top(i)\boldsymbol{\theta} + \epsilon(i) \qquad \underset{3\times 1}{\boldsymbol{\varphi}(i)} = [1 \quad \varphi_1(i) \quad \varphi_2(i)]^\top$$

$$\underset{47\times 3}{X} = \begin{bmatrix} \boldsymbol{\varphi}^\top(1) \\ \boldsymbol{\varphi}^\top(2) \\ \vdots \\ \boldsymbol{\varphi}^\top(N) \end{bmatrix} \qquad \underset{3\times 1}{\boldsymbol{\theta}} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix} \qquad \underset{47\times 1}{Y} = \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(47) \end{bmatrix}$$

```
% Read data from file
data = csvread('ex2data.txt');
X = data(:, 1:2); % Features
y = data(:, 3); % Price
N = length(y); % Number of data

% Add intercept term to X
X = [ones(N, 1) X];

% Calculate the parameters from the
normal equation
theta_hat = pinv(X'*X)*X'*y;
```

```
% Estimate the price of a 1650 sq-
ft, 3 br house
price_hat = [1 3 1650]*theta_hat;
```

Punto non visto durante la stima di θ

Calcolare e implementare il gradiente

In generale, se abbiamo **più di un regressore** (ovvero, un vettore $\boldsymbol{\varphi} = [1\ \varphi_1,\ \varphi_2, \dots, \varphi_{d-1}]^\top \in \mathbb{R}^{d\times 1}$) possiamo implementare il gradient descent come di seguito:

```
For {
```

$$\theta_0 = \theta_0 - \alpha \cdot \frac{2}{N}\sum_{i=1}^{N}(y(i) - \boldsymbol{\varphi}^\top(i)\boldsymbol{\theta}) \cdot (-1)$$

$$\theta_1 = \theta_1 - \alpha \cdot \frac{2}{N}\sum_{i=1}^{N}(y(i) - \boldsymbol{\varphi}^\top(i)\boldsymbol{\theta}) \cdot \big(-\varphi_1(i)\big)$$

$$\vdots$$

$$\theta_{d-1} = \theta_{d-1} - \alpha \cdot \frac{2}{N}\sum_{i=1}^{N}(y(i) - \boldsymbol{\varphi}^\top(i)\boldsymbol{\theta}) \cdot \big(-\varphi_{d-1}(i)\big)$$

```
}
```

es 2 : φ_1 : Area [Feet^2]

φ_2: Numero di camere da letto

Y : prezzo [\$]

modello lineare

$Y(i) = \theta_0 + \theta_1\varphi_1 + \theta_2\varphi_2 + \varepsilon = \bar{\varphi}^T(i)\theta + \varepsilon(i)$

con $\bar{\varphi}(i) = [1 \quad \varphi_1(i) \quad \varphi_2(i)]^T$, $X = \begin{bmatrix} \bar{\varphi}^T(1) \\ \bar{\varphi}^T(2) \\ \vdots \\ \bar{\varphi}^T(N) \end{bmatrix}$, $\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix}$, $Y = \begin{bmatrix} Y(1) \\ Y(2) \\ \vdots \\ Y(47) \end{bmatrix}$ ε_N

Normal equation

pseudo inversa

$\hat{\theta} = (X^TX)^{-1}X^TY$ —MATLAB→ theta_hat = pinv(X'*X) * X'*Y

Gradient Descent :

Features normalize

$\varphi_j(i) = \dfrac{\varphi_j(i) - \hat{\mu}_j}{\sqrt{\hat{\sigma}_j^2}}$ $j = 1, \dots, d-1$

$\hat{\mu}_j = \frac{1}{N}\sum_{i=1}^{N} \varphi_j(i)$ $j = 1, \dots, d-1$

$\sigma_j^2 = \frac{1}{N}\sum_{i=1}^{N} (x_j(i) - \mu_j)^2$ $j = 1, \dots, d-1$

Gradiente

$\nabla J(\bar{\theta}) = -\frac{2}{N} \cdot X^T \cdot (Y - X\bar{\theta})$

stima

$\hat{\theta}^{(k+1)} = \theta^{(k)} - \alpha \nabla J(\bar{\theta})\big|_{\theta=\theta^{(k)}}$

Compute cost

$J(\theta) = \frac{1}{N}(Y - X\bar{\theta})^T(Y - X\bar{\theta})$

? Quanto vale la stima del vettore $\hat{\theta}$ → $\hat{\theta}_0 = 340398{,}69$ $\hat{\theta}_1 = 109855{,}30$ $\hat{\theta}_2 = -5873{,}74$

? Qual è il profitto predetto per una casa con area di 1650 sq-ft e 3 camer

ricordarsi di normalizzare i dati → ([1650 3] - μ) ./ sigma

↳ $\hat{Y}$ = \$ 293236,30

→ From normal equations : $\hat{\theta}_0 = 89597{,}91$ $\hat{\theta}_1 = 139{,}21$ $\hat{\theta}_2 = -8738{,}02$

→ From normal equations : $\hat{y}$ = \$ 293081,46 ✓

```
First 10 examples from the dataset:
 x = [2104 3], y = 399900
 x = [1600 3], y = 329900
 x = [2400 3], y = 369000
 x = [1416 2], y = 232000
 x = [3000 4], y = 539900
 x = [1985 4], y = 299900
 x = [1534 3], y = 314900
 x = [1427 3], y = 198999
 x = [1380 3], y = 212000
 x = [1494 3], y = 242500
Theta computed from the normal equations:
 89597.909543
 139.210674
 -8738.019112

Predicted price of a 1650 sq-ft, 3 br house (using normal equations):
 $293081.464335



Normalizing Features ...
Running gradient descent ...
Theta computed from gradient descent:
 340398.694491
 109855.300268
 -5873.743442

Predicted price of a 1650 sq-ft, 3 br house (using gradient descent):
 $293236.305203
>>
```