

# Représentation par Modèles

---

Régression Multivariée

# Définition du problème

Age (x)	Taille en m (y)
2.0658746e+00	7.7918926e-01
2.3684087e+00	9.1596757e-01
2.5399929e+00	9.0538354e-01
2.5420804e+00	9.0566138e-01
2.5490790e+00	9.3898890e-01
2.7866882e+00	9.6684740e-01
2.9116825e+00	9.6436824e-01
3.0356270e+00	9.1445939e-01
3.1146696e+00	9.3933944e-01
3.1582389e+00	9.6074971e-01
3.3275944e+00	8.9837094e-01
3.3793165e+00	9.1209739e-01
3.4122006e+00	9.4238499e-01
3.4215823e+00	9.6624578e-01
3.5315732e+00	1.0526500e+00
3.6393002e+00	1.0143791e+00
3.6732537e+00	9.5969426e-01
3.9256462e+00	9.6853716e-01
4.0498646e+00	1.0766065e+00

$$h_{\theta}(X) = \theta_0 + \theta_1 \times x$$

# Définition du problème

Superficie (m <sup>2</sup> )	Nombre de chambres	Prix (USD)
2104	3	3.999e+005
1600	3	3.299e+005
2400	3	3.69e+005
1416	2	2.32e+005
3000	4	5.399e+005
1985	4	2.999e+005
1534	3	3.149e+005
1427	3	1.99e+005
1380	3	2.12e+005
1494	3	2.425e+005
1940	4	2.4e+005
2000	3	3.47e+005
1890	3	3.3e+005
4478	5	6.999e+005
1268	3	2.599e+005
2300	4	4.499e+005
1320	2	2.999e+005
1236	3	1.999e+005
2609	4	500 000,00
3031	4	5.99e+005

# Définition du problème

x	$x^2$	Y
22.589	510.26	134.57
26.232	688.1	230.17
-4.9205	24.212	28.406
26.535	704.11	212.24
15.294	233.92	51.353
-6.0984	37.19	42.956
1.1399	1.2994	17.676
11.875	141.02	33.836
28.3	800.91	277.85
28.596	817.7	261.87
-3.6955	13.657	41.69
28.824	830.81	282.03
28.287	800.14	287.55
9.415	88.643	13.206
22.011	484.49	154.2
-4.3245	18.702	39.125
6.8705	47.203	-3.455
26.629	709.13	244.43
21.688	470.38	120.09
28.38	805.41	256.13

# Base d'entraînement

X	$X^2$	Y
22.589	510.26	134.57
26.232	688.1	230.17
-4.9205	24.212	28.406
26.535	704.11	212.24
15.294	233.92	51.353
-6.0984	37.19	42.956
1.1399	1.2994	17.676
11.875	141.02	33.836
28.3	800.91	277.85
28.596	817.7	261.87
-3.6955	13.657	41.69
28.824	830.81	282.03
28.287	800.14	287.55
9.415	88.643	13.206
22.011	484.49	154.2
-4.3245	18.702	39.125
6.8705	47.203	-3.455
26.629	709.13	244.43
21.688	470.38	120.09
28.38	805.41	256.13

- Notations :
  - $m$  : nombre d'observations,
  - $X$  : variable d'entrée ou attribut,
  - $Y$  : variable de sortie ou objectif,
  - $(X,Y)$  : une observation,
  - $(X_i, Y_i)$  :  $i^{\text{ème}}$  observation,
  - $n$  : nombre d'attributs,
  - $X_{i,j}$  :  $j^{\text{ème}}$  attribut de la  $i^{\text{ème}}$  observation.
- Exemple
  - $X_{1,2} = 510,26$ .
  - $X_{3,1} = -4,9205$ .

# Hypothèse

$$\text{---} h_{\theta}(X) = \theta_0 + \theta_1 \times x \text{---}$$

$$h_{\theta}(X) = \theta_0 + \theta_1 \times x_1 + \theta_2 \times x_2 + \dots + \theta_n \times x_n$$

# Hypothèse

$$h_{\theta}(X) = \theta_0 + \theta_1 \times x_1 + \theta_2 \times x_2 + \dots + \theta_n \times x_n$$

$$x_0 = 1$$

$$h_{\theta}(X) = \theta_0 \times x_0 + \theta_1 \times x_1 + \theta_2 \times x_2 + \dots + \theta_n \times x_n$$

$$h_{\theta}(X) = X^T \theta$$

# Hypothèse

$$h_{\theta}(X) = X^T \theta$$

$$\theta = [\theta_0 \ \theta_1 \ \dots \ \theta_n]^T \in \mathbb{R}^{n+1}$$

$$X = [x_0 \ x_1 \ \dots \ x_n]^T \in \mathbb{R}^{n+1}$$



# Contexte

- Hypothèse :  $h_{\theta}(X) = X^T \theta$

$$h_{\theta}(X) = \theta_0 \times x_0 + \theta_1 \times x_1 + \theta_2 \times x_2 + \dots + \theta_n \times x_n$$

- Paramètres :  $\theta$
- Fonction de coût (fonctionnelle) :

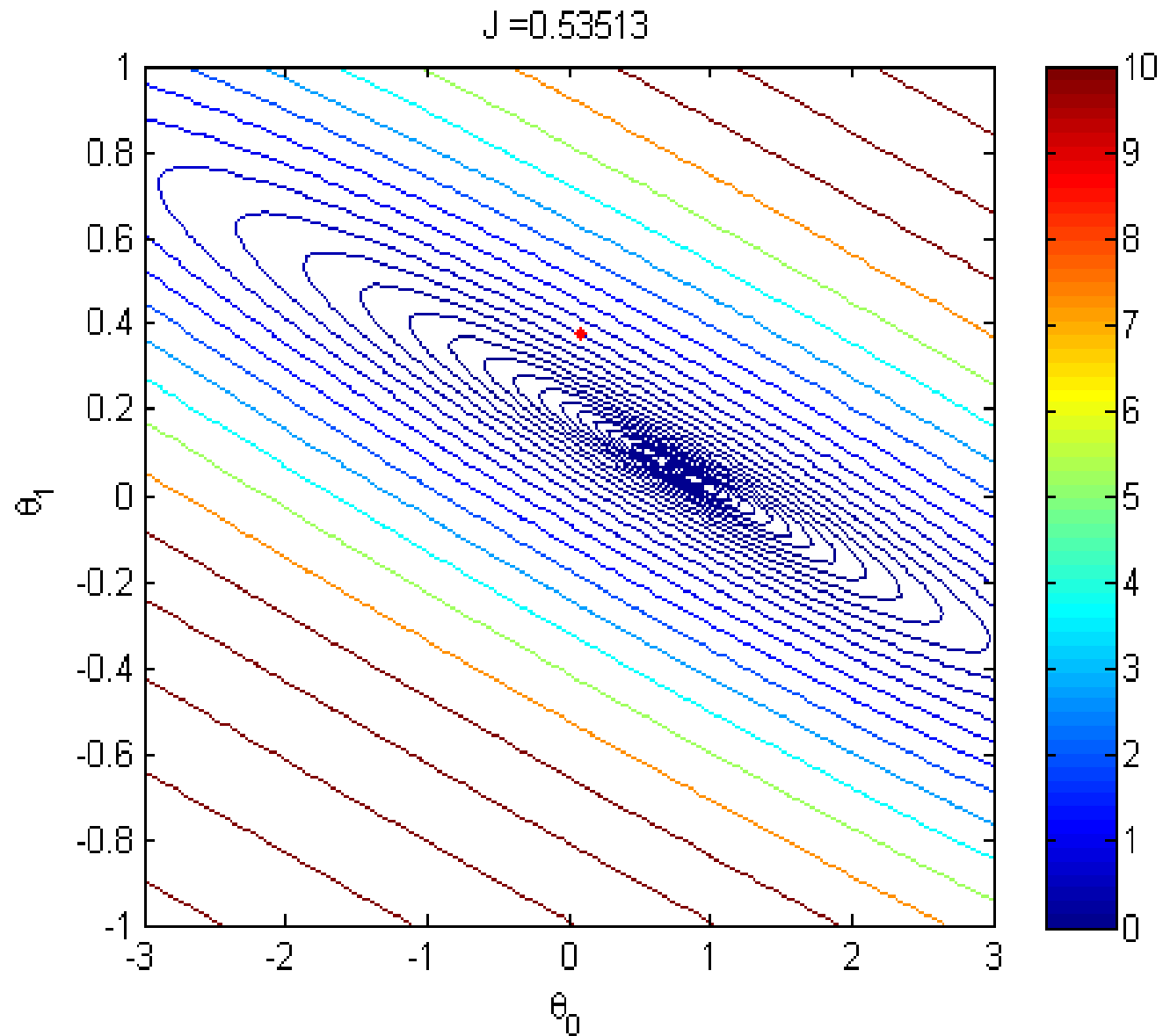
$$J(\theta) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(X^{(i)}) - Y^{(i)})^2$$

- Objectif :  $\min_{\theta} J(\theta)$

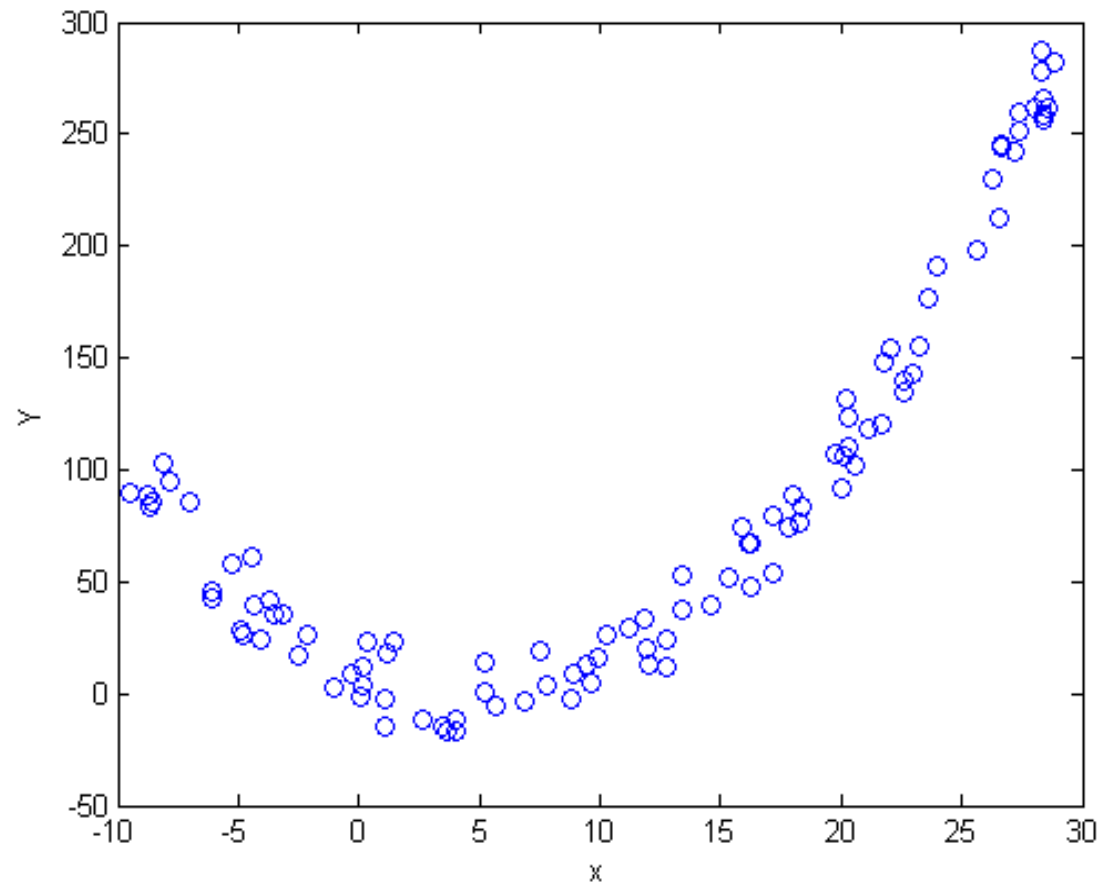
# Estimation des paramètres

- Méthodes des moindres carrés
- Descente de gradient

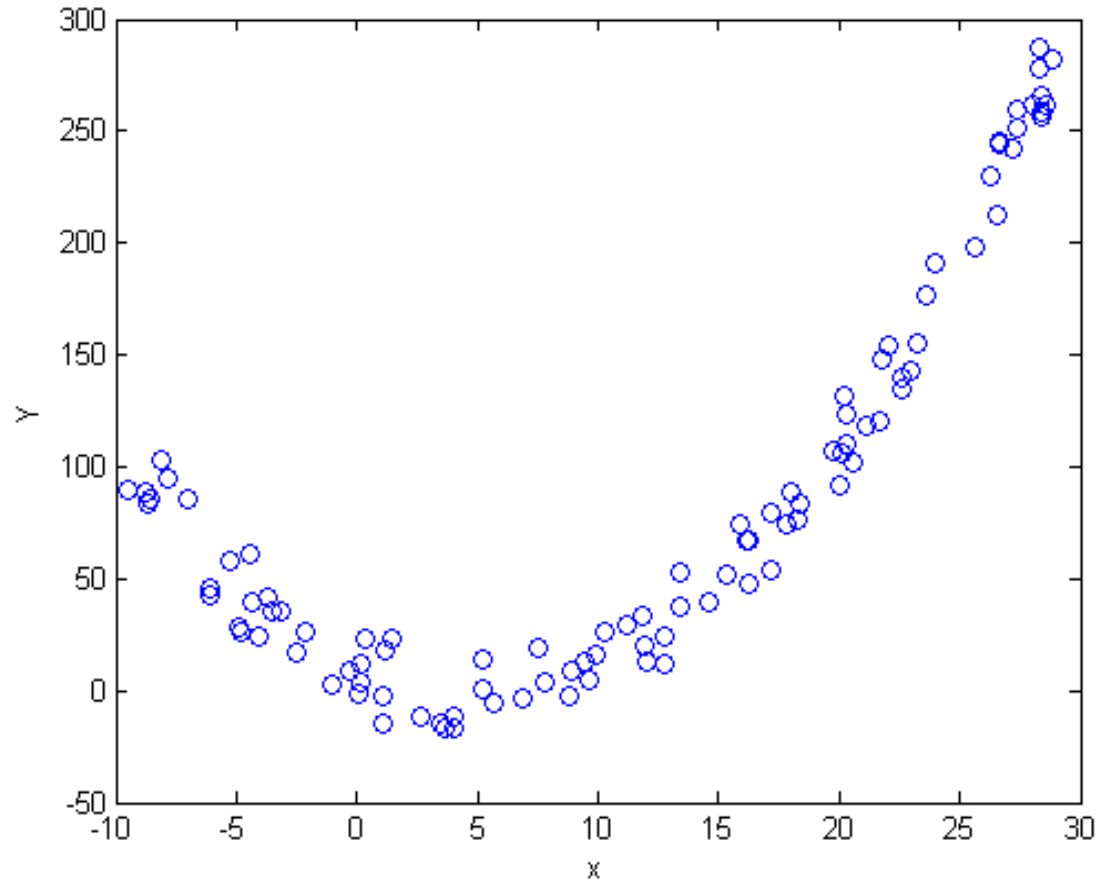
# Pré-traitement des données



# Pré-traitement des données

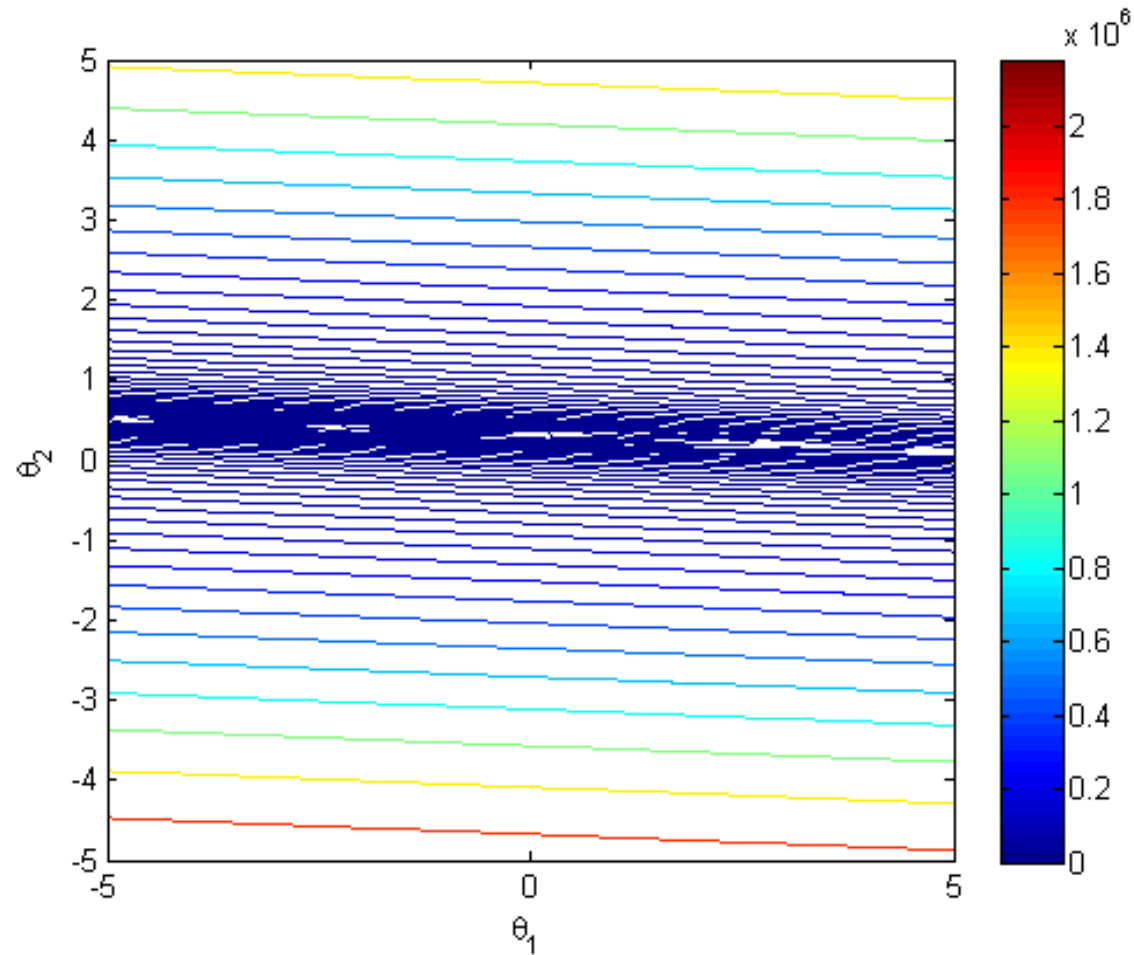


# Pré-traitement des données



$$h_{\theta}(X) = \theta_0 \times x_0 + \theta_1 \times x_1 + \theta_2 \times x_2$$

# Pré-traitement des données



# Pré-traitement des données

$$0 \leq x^j \leq 1$$

$$-1 \leq x^j \leq 1$$

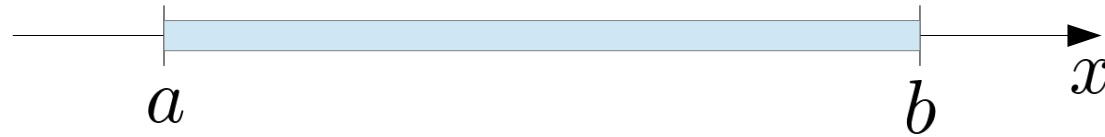
$$x^j = \frac{x^j - \mu_j}{\sigma_j}$$

$\mu_j$ : moyenne de l'attribut  $j$

$\sigma_j$ : écart type de l'attribut  $j$

# Normalisation

- $a \leq x \leq b$



- $0 \leq x' \leq c$

$$x' = x - \min(x) = x - a \text{ et } c = b - a$$



- $0 \leq x'' \leq 1$

$$x'' = \frac{x'}{\max(x')} = \frac{x'}{c}$$





# Régression Polynomiale

$$n \text{ ?}$$

$$x_j = x \text{ ?}$$

$$x_j = x^2 \text{ ?}$$

$$x_j = \sqrt{x} \text{ ?}$$

**Merci de votre Attention**

# Questions