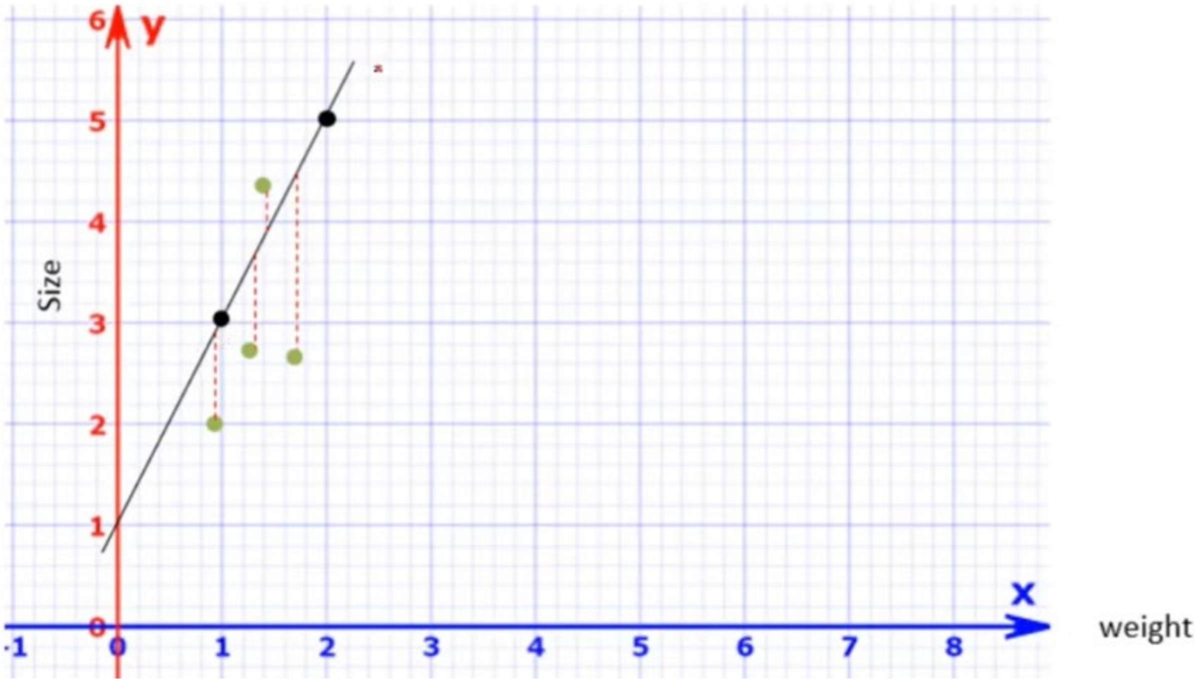
Ridge regularization (L2)

$$\text{Size} = \theta_0 + \theta_1 \times weight$$



$$\boxed{Size = 1 + 2 \times \ weight}$$

$$minimize\ j(\theta)$$

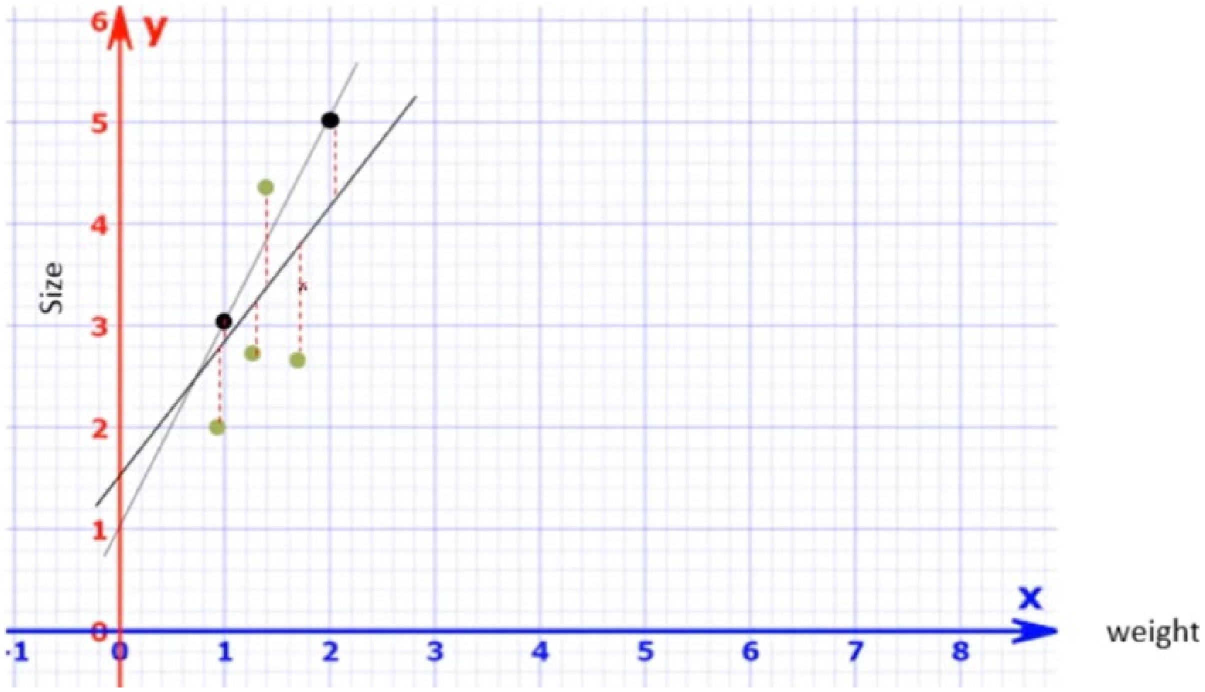
$$j(\theta) = (h(x) - y)^2$$

Least square error

$$(h(x) - y)^2 = 0$$

Dans cet exemple, il n'existe pas d'erreurs dans le modèle d'apprentissage

The main idea of regularization is to find a new line does not fit the training set very well



La descente du gradient va ôter pour le premier exemple, vu que l'erreur d'apprentissage est minimale.

Dans le deuxième exemple, intervient la régularisation pour imposer une pénalité sur le premier modèle.

$$j(\theta) = (h(x) - y)^2 + \lambda(\theta_1)^2$$

Théta 1 pour la pente, lambda*théta est le régularisateur

$$\boxed{Size = 1 + 2 \times \ weight}$$

Théta 1=2

Lambda=1

$$j(\theta) = (h(x) - y)^2 + \lambda(\theta_1)^2$$

Devient =0+1*$(2)^2$=4

Deuxième ligne est :

$$\boxed{Size = 1 + 1.5 \times \ weight}$$

Recalculons l'erreur,

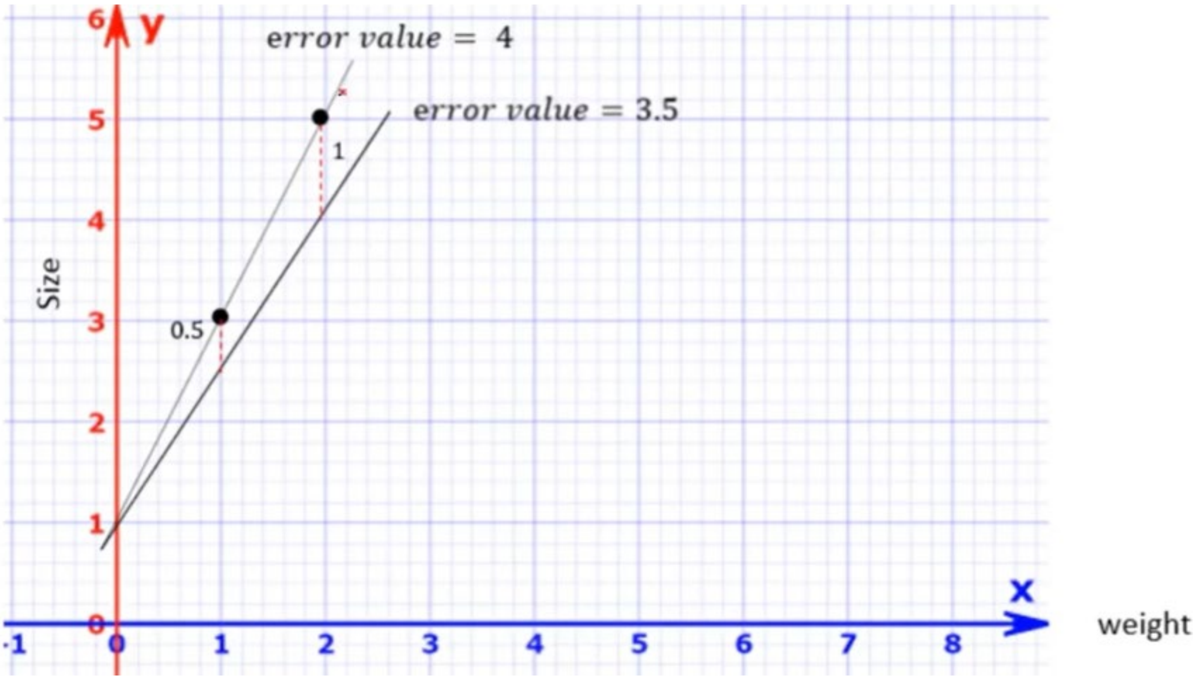
$$(h(x) - y)^2 = 0.25$$

$$(h(x) - y)^2 = 1$$

Théta=1.5

$$\lambda(\theta_1)^2 = 1 \times 2.25$$

La somme est : 3.5



Maintenant la descente du gradient va choisir la deuxième ligne

Le cas pour plusieurs caractéristiques

$$\boxed{Size = \theta_0 + \theta_1 \times \ weight + \theta_2 \times length + \cdots + \theta_n \times feature_n}$$

$$\lambda(\theta_1^2 + \theta_2^2 + \cdots + \theta_n^2)$$

Cost function

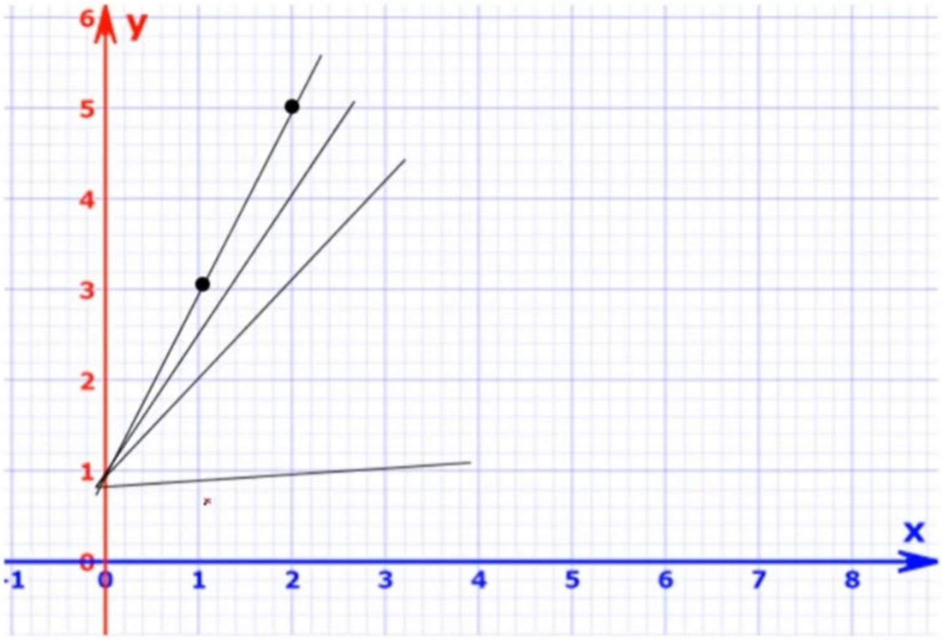
$$j(\theta) = \frac{1}{2}\sum_{i=1}^{m}(h(x^{(i)}) - y^{(i)})^2 + \lambda\sum_{j=1}^{n}\theta_j^2$$

Effet de lambda

$\lambda = 0$

$\lambda = 1$

$\lambda = 2$



Lambda plus grand, conduit théta à être proche de zéro,

Dans ce cas on aura un underfiting.

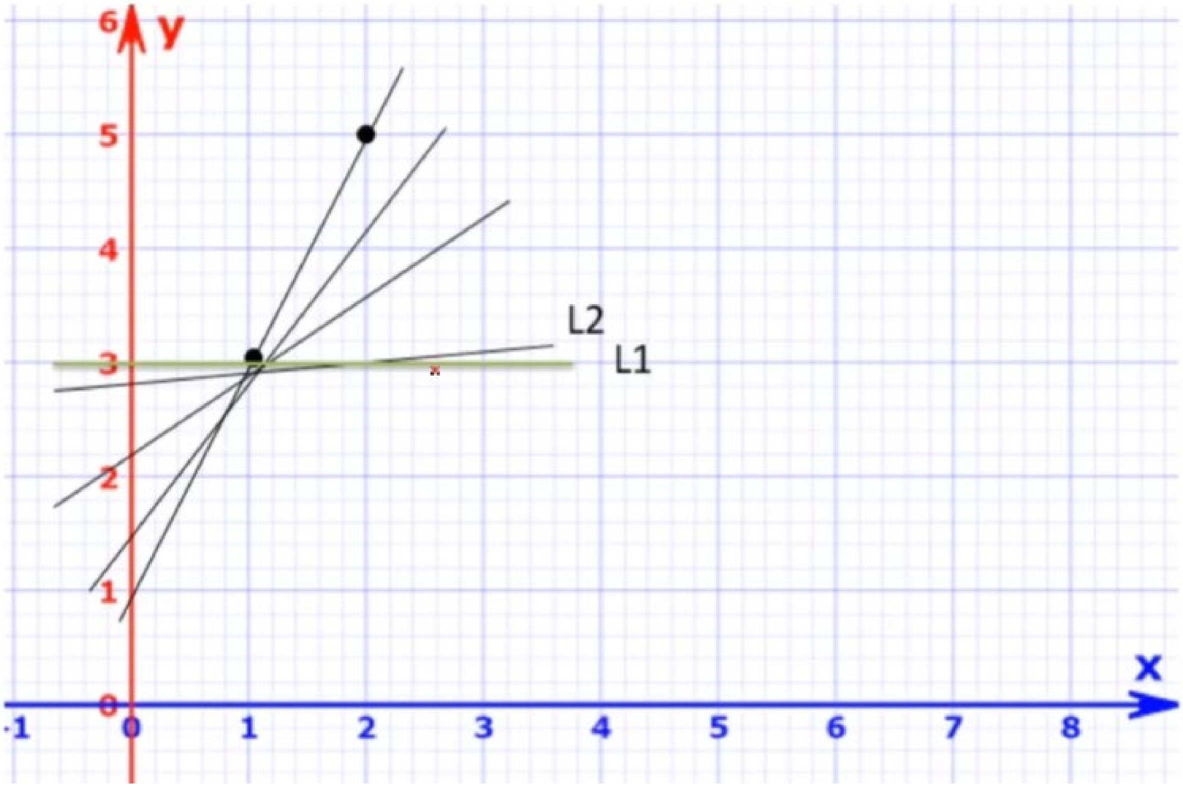
Lasso regularization (L1) Least absolute shrinkage and selection operator regression

La régression par l'opérateur de rétrécissement et de sélection des moindres carrés communément est une méthode de régression linéaire régularisée. Cette méthode utilise une régularisation de type L1 pour réduire la complexité du modèle et sélectionner les variables les plus pertinentes.

En ajoutant une contrainte sur la somme des valeurs absolues des coefficients du modèle, la régularisation L1 force certains coefficients à zéro, ce qui permet d'éliminer les variables les moins pertinentes. Cette propriété de sélection de variables est particulièrement utile lorsque l'on travaille avec des jeux de données contenant un grand nombre de variables, certaines étant potentiellement redondantes ou peu informatives.

$$j(\theta) = (h(x) - y)^2 + \lambda|\theta_j|$$

Opérateur de sélection

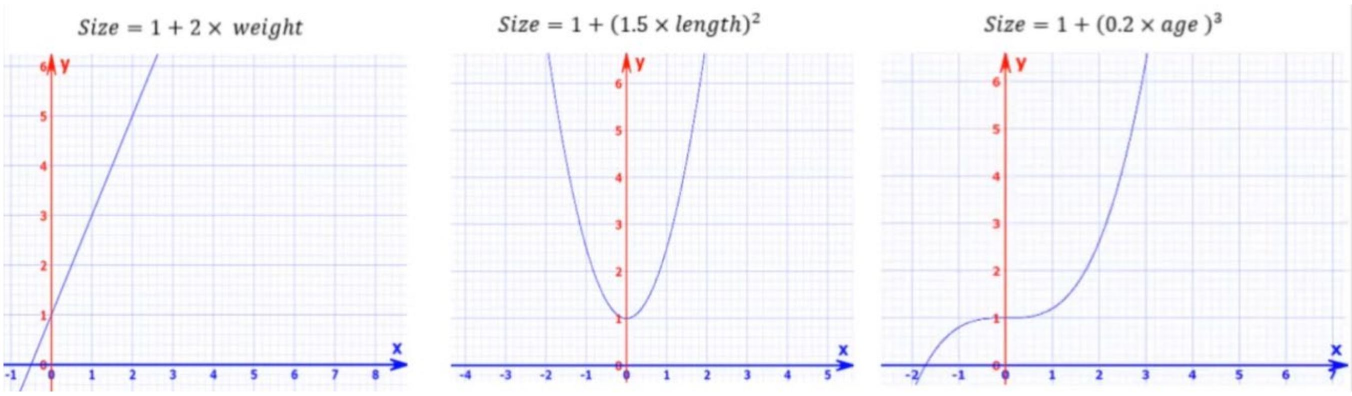


Lambda plus grand implique Théta= 0 ce qui permet d’éliminer une caractéristique.

$$Size = 1 + 2 \times\ weight + (1.5 \times length)^2 + (0.2 \times age\ )^3$$

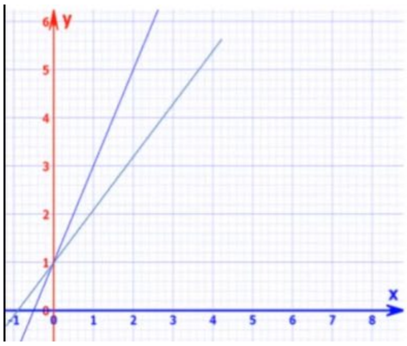
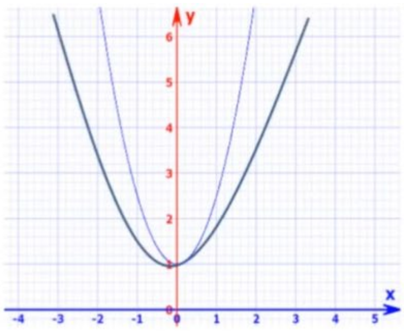
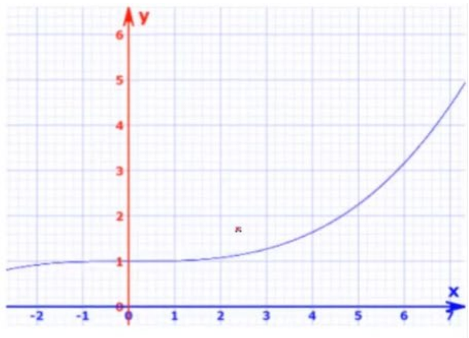
Prédire la taille en se basant sur les caractéristiques : Poids, Longueur et Age.

D’après la formule, la taille dépend fortement du poids (coefficient =2).

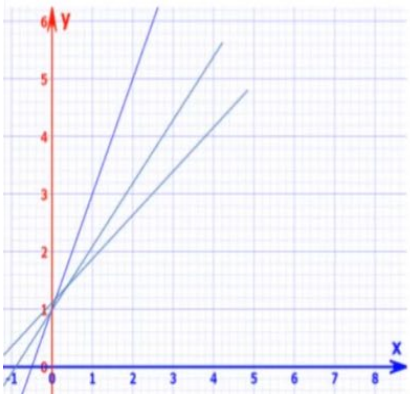
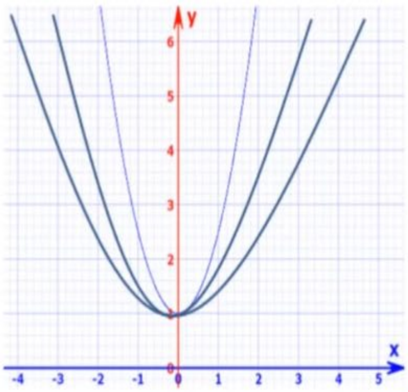
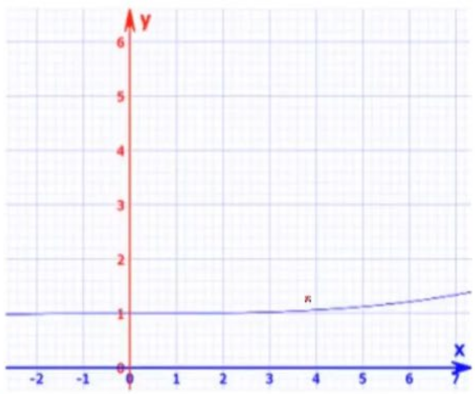


$\lambda = 0$ $\quad Size = 1 + 2 \times\ weight + (1.5 \times length)^2 + (0.2 \times age\ )^3$

$\lambda = 1$ $\quad Size = 1 + 1.3 \times\ weight + (0.9 \times length)^2 + (0.01 \times age\ )^3$

$\lambda = 2$ $$Size = 1 + 0.9 \times \ weight + (0.6 \times length)^2 + (0.002 \times age\ )^3$$

Dans le graphe 3, par la méthode L2, théta n'égale jamais 0

$\lambda = 3$ $$Size = 1 + 0.7 \times \ weight + (0.4 \times length)^2 + (0 \times age\ )^3$$