MINITEL

MACHINE LEARNING

01 SUPERVISED MACHINE LEARNING : REGRESSION AND CLASSIFICATION

THÉO GACHET

UNE SÉRIE DE FORMATIONS

PARTIE 1 : SUPERVISED MACHINE LEARNING : REGRESSION AND CLASSIFICATION

Introduction to Machine Learning
Regression with multiple input
Classification

PARTIE 2 : ADVANCED LEARNING ALGORITHMS

Neural Networks
Neural Networks training
Advice for applying Machine Learning
Decision Trees

PARTIE 3 : UNSUPERVISED LEARNING, RECOMMENDERS, REINFORCEMENT LEARNING

Unsupervised Learning
Recommander Systems
Reinforcement Learning

01 SUPERVISED MACHINE LEARNING : REGRESSION AND CLASSIFICATION

SUPERVISED VS. UNSUPERVISED ML

What is Machine Learning ?
Supervised Learning
Unsupervised Learning
Jupyter Notebooks
Lab : Python and Jupyter Notebooks

REGRESSION MODEL

Linear regression model
Lab : Model representation
Cost function formula
Cost function intuition
Visualizing the cost function
Visualization examples
Lab : Cost function

TRAIN THE MODEL WITH GRADIENT DESCENT

Gradient descent
Implementing gradient descent
Gradient descent intuition
Learning rate
Gradient descent for linear regression
Running gradient descent
Lab : Gradient descent

01

SUPERVISED VS. UNSUPERVISED MACHINE LEARNING

THÉO GACHET

Supervised learning

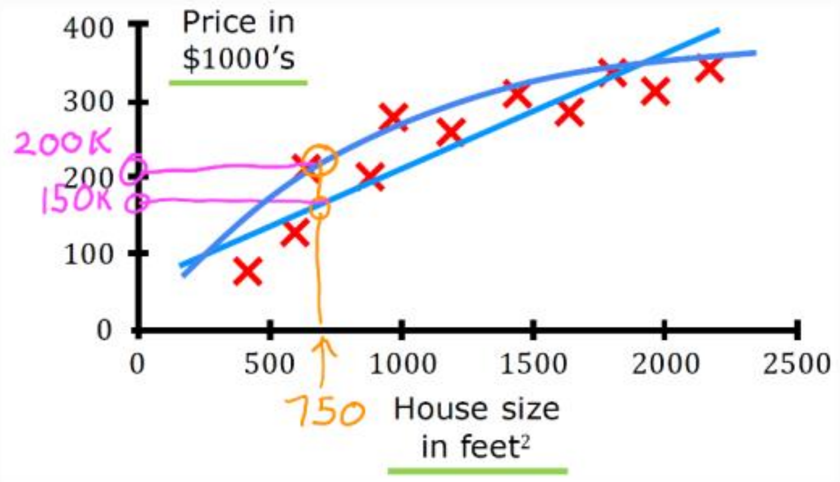
x input → y output label

learns from being given data labeled with the "right answers"

Regression

Predict a number

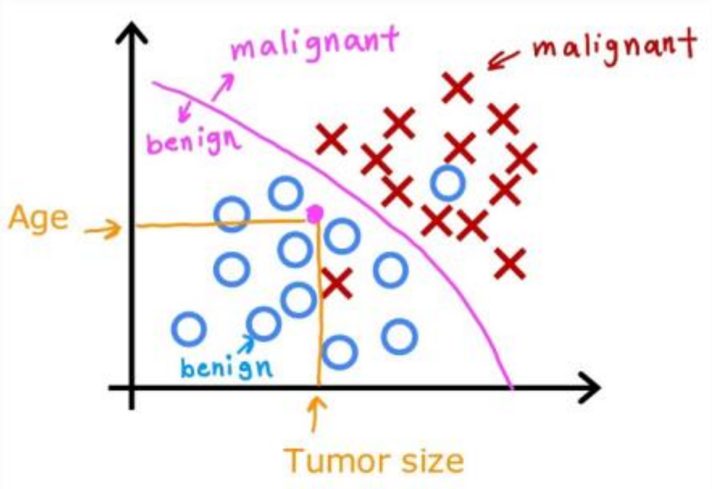
Infinitely many possible outcomes



Classification

Predict categories

Small number of possible outputs



Unsupervised Learning

Data only comes with inputs x, but not output labels y

Algorithm has to find structure (= something interesting in unlabeled data)

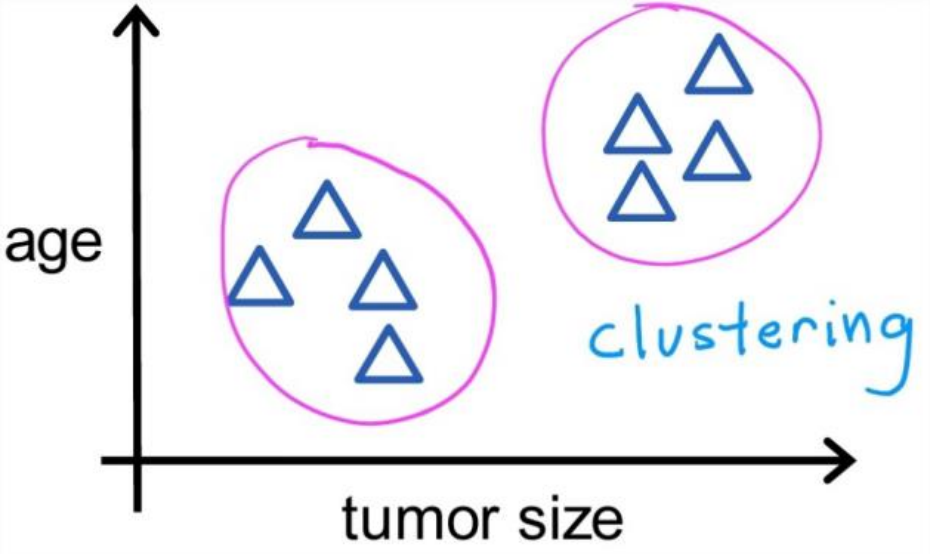
Clustering
Group similar data points together.

Anomaly detection
Find unusual data points.

Dimensionality reduction
Compress data using fewer numbers.

Clustering

Example : Google News



MINITEL

LAB-01

INTRODUCTION TO PYTHON AND JUPYTER NOTEBOOKS

THÉO GACHET

QUESTIONS ?

SUR UN CONCEPT ? UNE IDÉE ?
SUR UN DÉTAIL DU CODE ?
(ENVIE D'UNE PAUSE ?)

N'HÉSITEZ PAS !

IL N'Y A PAS DE QUESTION BÊTE, SI VOUS AVEZ UN DOUTE, D'AUTRES ONT SÛREMENT LE MÊME

02

REGRESSION MODEL

Linear Regression Model

Terminology

x	y
$x^{(1)}$	$y^{(1)}$
$\vdots$	$\vdots$
$x^{(m)}$	$y^{(m)}$

(training set)

x = "input" variable / feature

y = "output" variable / "target" variable

m = number of variables in the training set

w = weight / b = bias

(x, y) = single training example

$(x^{(i)}, y^{(i)})$ = i^{th} training example training set

↓

learning algorithm

↓

$x \rightarrow \boxed{f} \rightarrow \hat{y}$

feature — model — prediction (estimated y)

target → y

$f_{w,b}(x) = wx + b$ (written $f(x)$)

Linear regression with one variable.

Univariate linear regression.

Uni: one; variate: variable

LAB-02

MODEL REPRESENTATION

THÉO GACHET

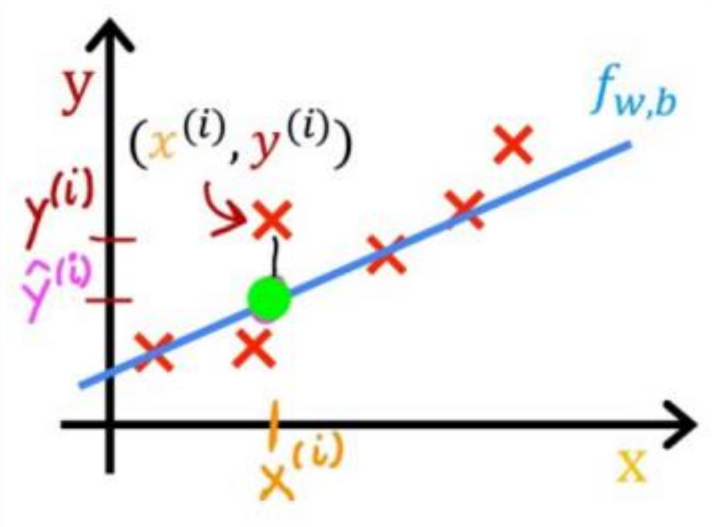
QUESTIONS ?

SUR UN CONCEPT ? UNE IDÉE ?
SUR UN DÉTAIL DU CODE ?
(ENVIE D'UNE PAUSE ?)

Cost Function

$$\hat{y}^{(i)} = f_{w,b}(x^{(i)}) = wx^{(i)} + b$$

Find (w,b) such as $\hat{y}^{(i)} \approx y^{(i)} \ \forall (x,y)$

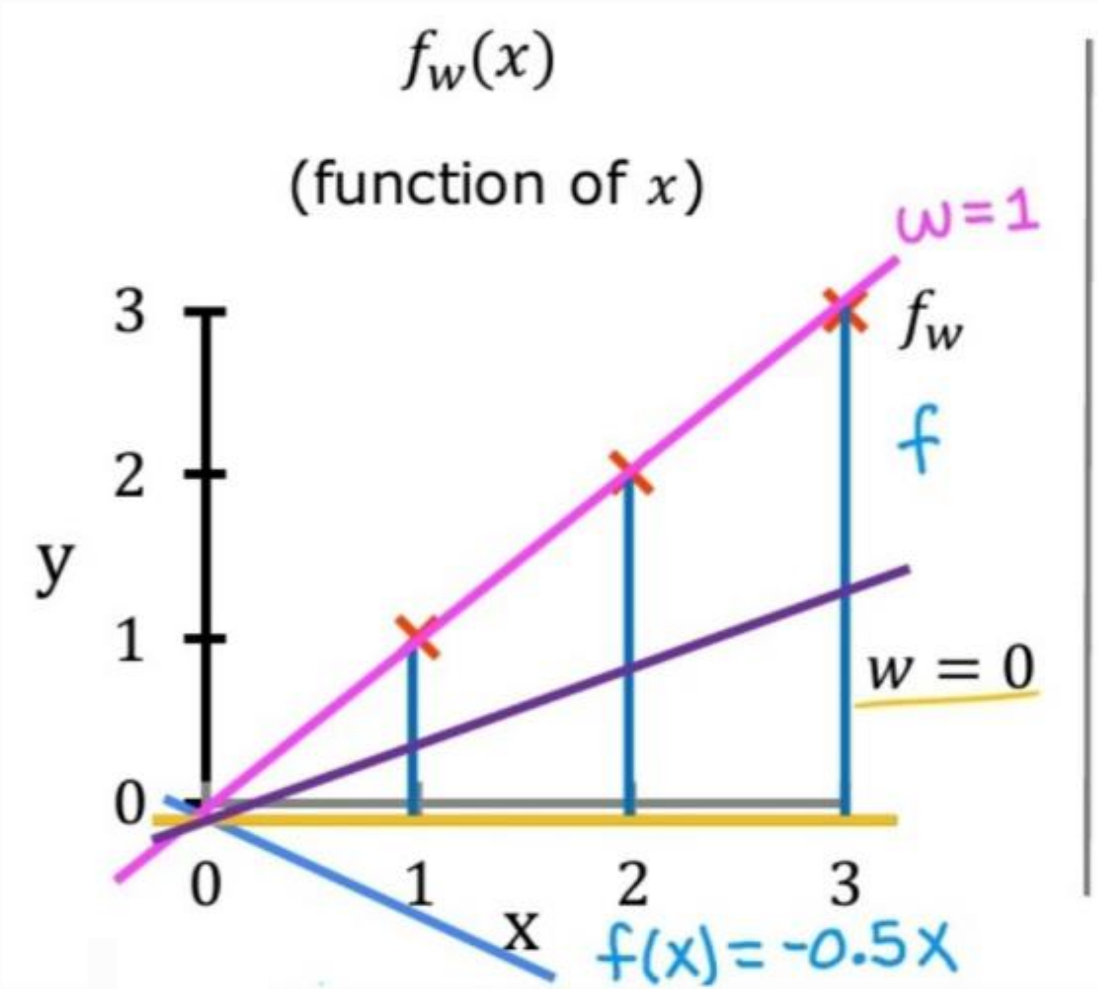


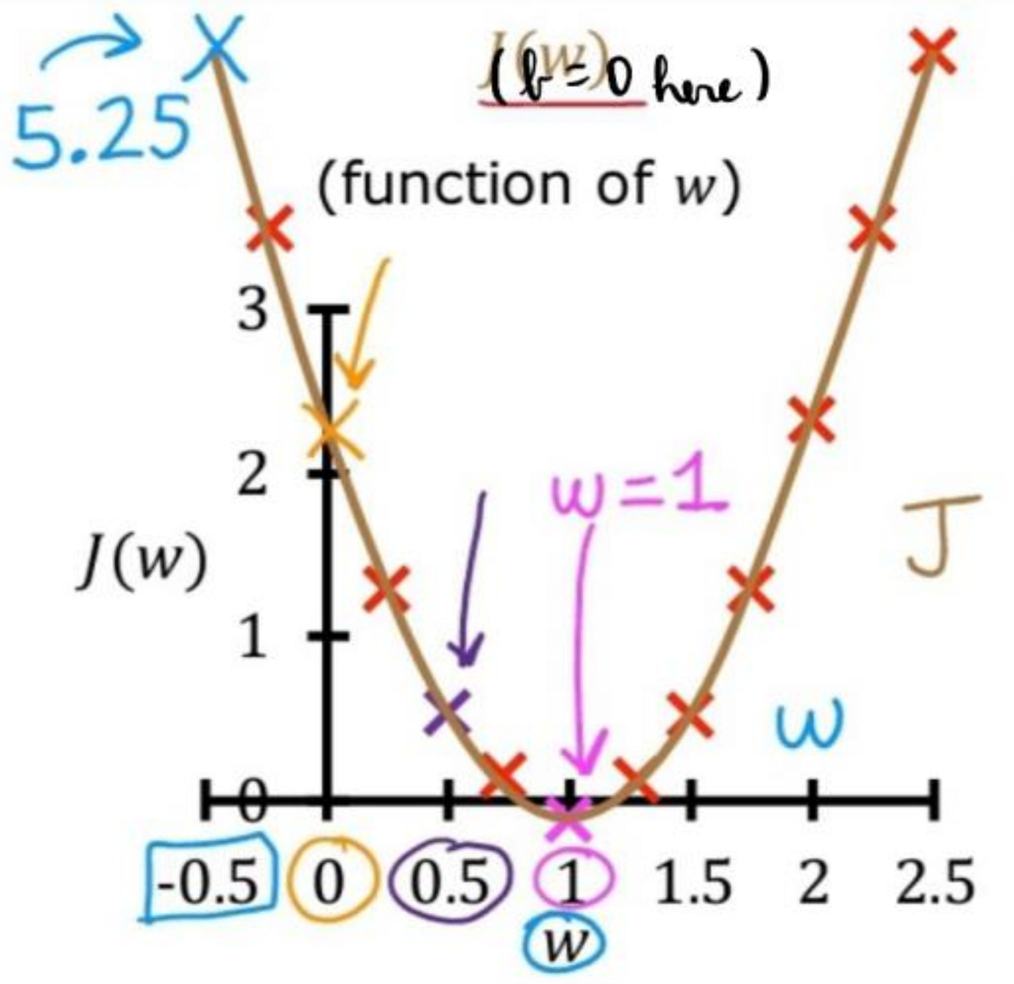
Squared error cost function

$$J(w,b) = \frac{1}{2m} \sum_{i=1}^{m} \Big(\underbrace{f_{w,b}(x^{(i)})}_{\hat{y}^{(i)}} - y^{(i)} \Big)^2$$

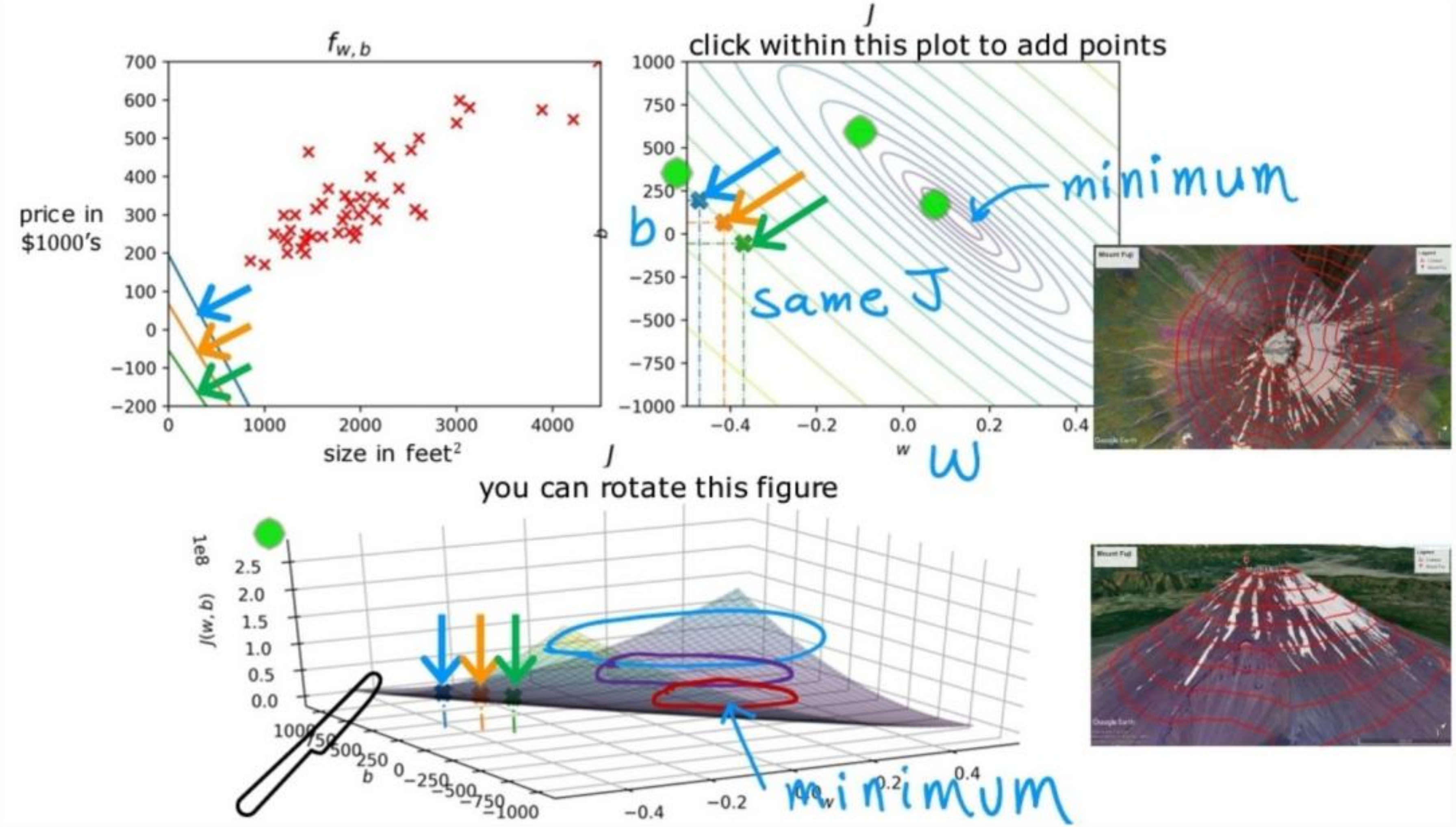
goal : to minimize $J(w,b)$

Cost Function Intuition





Visualizing the Cost Function → 3D Surface Plot vs. Contour Plot



LAB-03

COST FUNCTION

THÉO GACHET

QUESTIONS ?

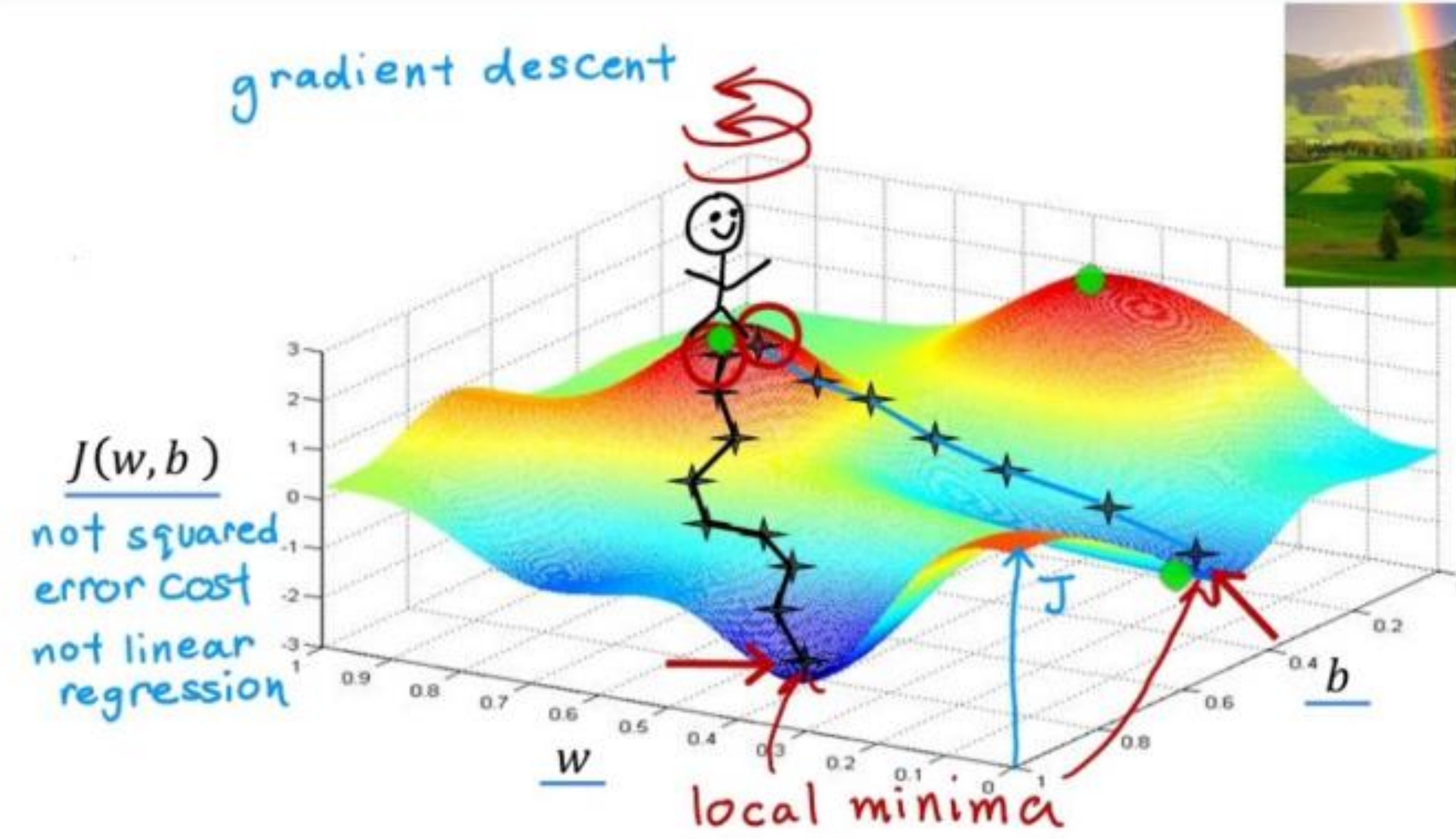
SUR UN CONCEPT ? UNE IDÉE ?
SUR UN DÉTAIL DU CODE ?
(ENVIE D'UNE PAUSE ?)

N'HÉSITEZ PAS !
IL N'Y A PAS DE QUESTION BÊTE, SI VOUS AVEZ UN DOUTE, D'AUTRES ONT SÛREMENT LE MÊME

03

TRAIN THE MODEL WITH GRADIENT DESCENT

Gradient Descent



Gradient Descent Algorithm

Repeat until convergence :

$$w = w - \alpha \frac{\partial}{\partial w} J(w, b)$$

$$b = b - \alpha \frac{\partial}{\partial b} J(w, b)$$

learning rate

Simultaneous update : ⚠

$$tmp\_w = w - \alpha \frac{\partial}{\partial w} J(w, b)$$

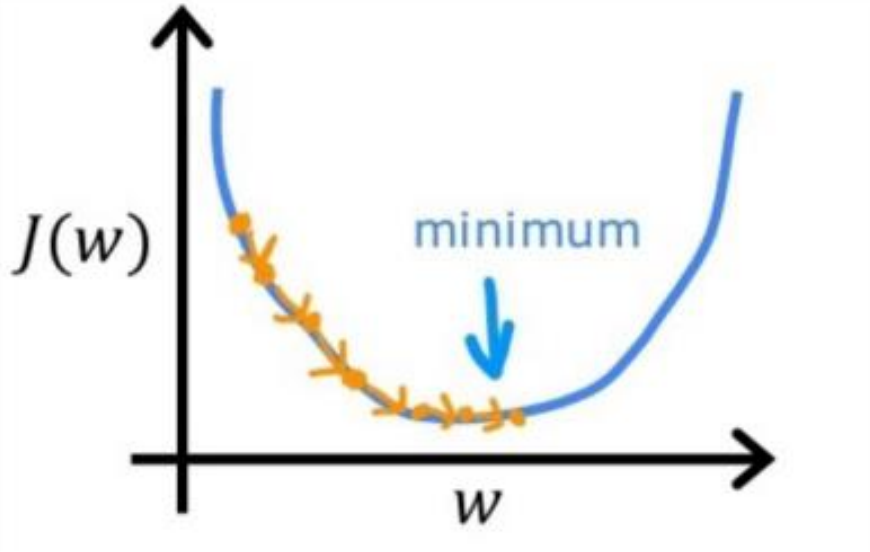
$$tmp\_b = b - \alpha \frac{\partial}{\partial b} J(w, b)$$

$$w = tmp\_w$$

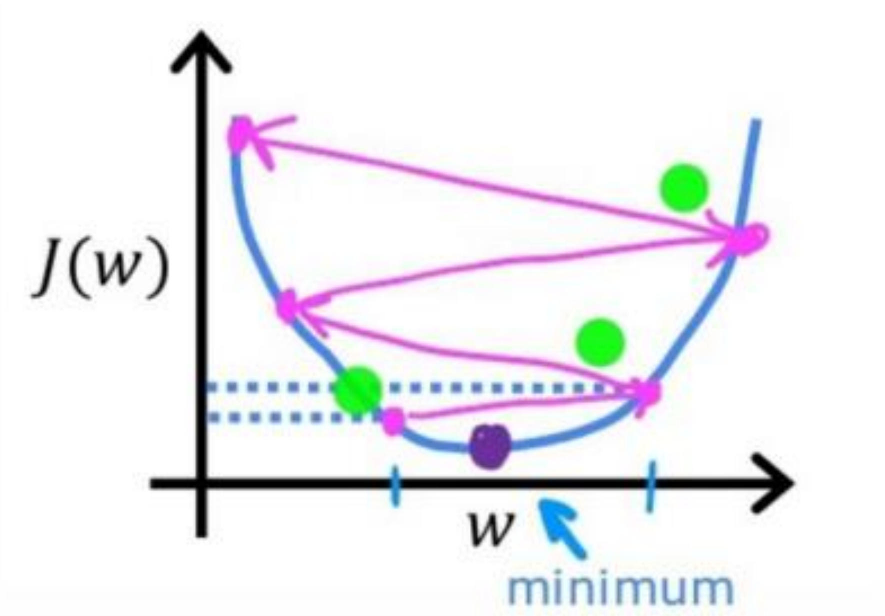
$$b = tmp\_b$$

Batch = each step of gradient descent uses all the training examples

Learning Rate



If α is too small...
Gradient descent may be slow.



If α is too large...

Gradient descent may:
- Overshoot, never reach minimum
- Fail to converge, diverge

MINITEL

LAB-04

GRADIENT DESCENT FOR LINEAR REGRESSION

THÉO GACHET

QUESTIONS ?

SUR UN CONCEPT ? UNE IDÉE ?
SUR UN DÉTAIL DU CODE ?
(ENVIE D'UNE PAUSE ?)

MINITEL

MACHINE LEARNING

01 SUPERVISED MACHINE LEARNING : REGRESSION AND CLASSIFICATION

THÉO GACHET