MACHINE LEARNING

03 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

02 REGRESSION WITH MULTIPLE INPUT VARIABLES

INTRODUCTION TO MULTIPLE LINEAR REGRESSION

Multiple features
Vectorization
Lab : Python, NumPy and vectorization
Gradient descent for multiple linear regression
Lab : Multiple linear regression

GRADIENT DESCENT

Feature scaling
Checking gradient descent for convergence
Choosing the learning rate
Feature engineering
Polynomial regression
Lab : Feature scaling and learning rate
Lab : Feature engineering and polynomial regression
Lab : Linear regression with scikit-learn
Lab : Linear regression

03 CLASSIFICATION

CLASSIFICATION WITH LOGISTIC REGRESSION

Motivations
Lab : Classification
Logistic regression
Lab : Sigmoid function and logistics
Decision boundary
Lab : Decision boundary

COST FUNCTION

Cost function for logistic regression
Lab : Logistic loss
Simplified cost function for logistic regression
Lab : cost function

GRADIENT DESCENT

Gradient Descent implementation
Lab : Gradient descent
Lab : Logistic regression with scikit-learn

THE PROBLEM OF OVERFITTING

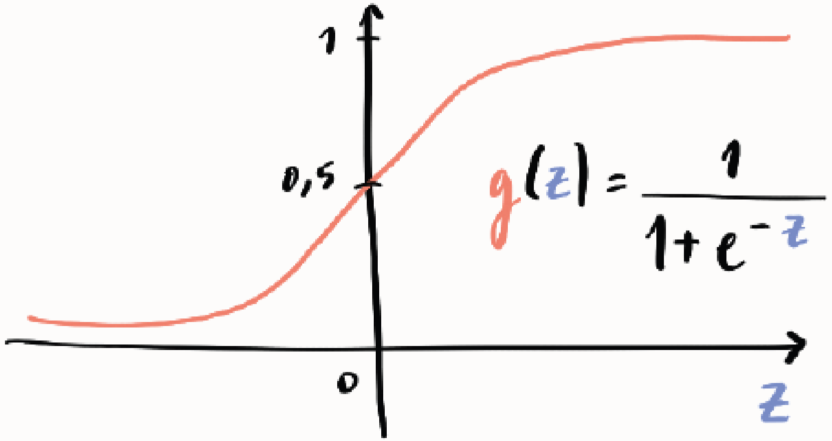
The problem of overfitting
Addressing overfitting
Lab : Overfitting
Cost function with regularization
Regularized linear regression
Regularized logistic regression
Lab : Regularization

01

CLASSIFICATION WITH LOGISTIC REGRESSION

THÉO GACHET

Logistic Regression



$$g(z) = \frac{1}{1+e^{-z}}$$

Sigmoid Function

binary classification

Logistic function with outputs in $[0, 1]$

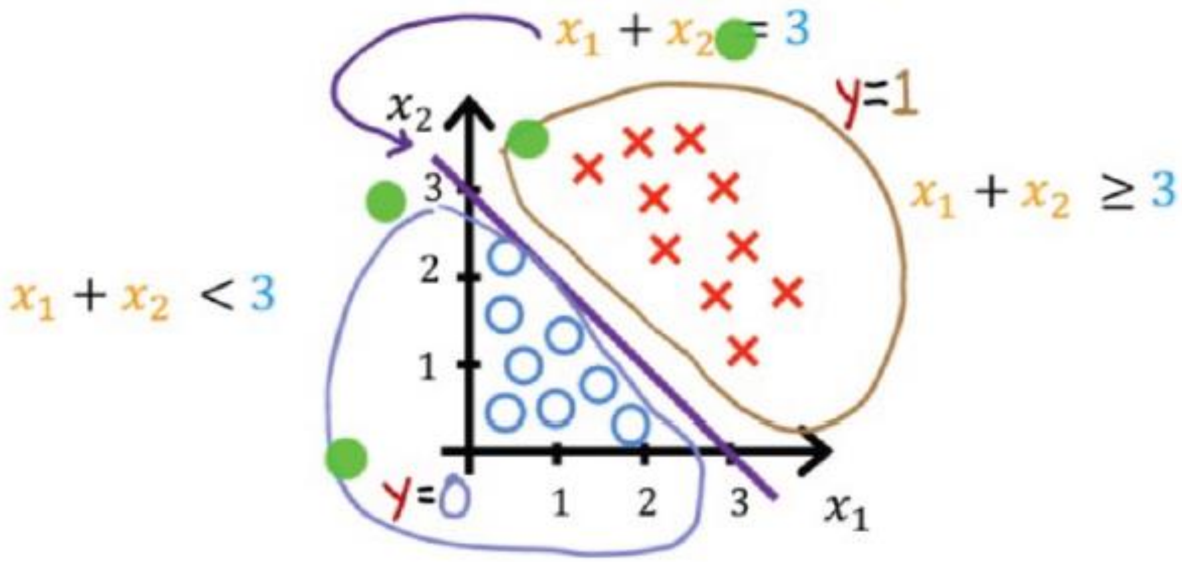
$$f_{\vec{w},b}(\vec{x}) = g(\underbrace{\vec{w}\cdot\vec{x}+b}_{z}) = \frac{1}{1+e^{-(\vec{w}\cdot\vec{x}+b)}} = P(y=1 \mid \vec{x};\vec{w},b)$$

Decision Boundary

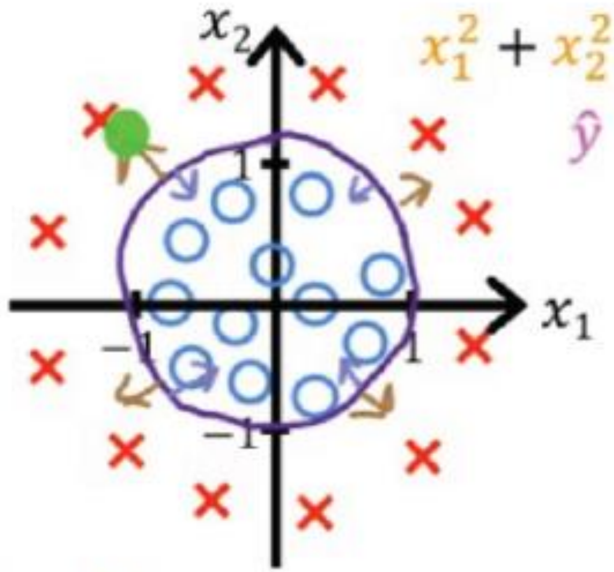
$$f_{\vec{w},b}(\vec{x}) = g(z) = g(\underset{1}{w_1}x_1 + \underset{1}{w_2}x_2 + \underset{-3}{b})$$

Decision boundary $z = \vec{w} \cdot \vec{x} + b = 0$

$z = x_1 + x_2 - 3 = 0$



Non-linear decision boundaries



$x_1^2 + x_2^2 \geq 1$

$\hat{y} = 1$

$x_1^2 + x_2^2 < 1$

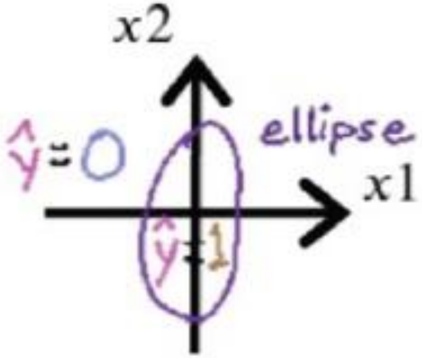
$\hat{y} = 0$

$$\overbrace{\underset{1}{w_1} x_1^2 + \underset{1}{w_2} x_2^2 + \underset{-1}{b}}^{z}$$

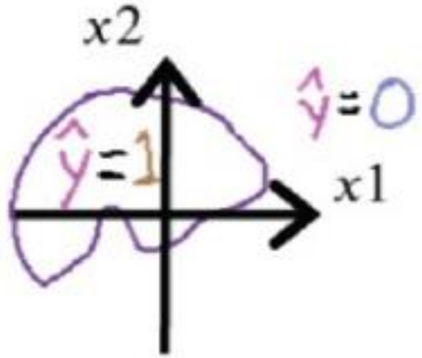
decision boundary $z = x_1^2 + x_2^2 - 1 = 0$

$x_1^2 + x_2^2 = 1$

Non-linear decision boundaries



$$f_{\vec{w},b}(\vec{x}) = g(z) = g(w_1x_1 + w_2x_2 + w_3x_1^2 + w_4x_1x_2 + w_5x_2^2 + w_6x_1^3 + \dots + b)$$



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

COST FUNCTION FOR LOGISTIC REGRESSION

THÉO GACHET

Loss Function

Loss = measure of the difference of a single example to its target value

Cost = measure of the losses over the training set

Squared Error Cost $$J(\vec{w},b) = \frac{1}{2m} \sum_{i=1}^{m} \left(f_{\vec{w},b}(x^{(i)}) - y^{(i)}\right)^2$$

<u>Problem</u>: $J(\vec{w},b)$ is convex for linear regression <u>but</u> non-convex for logistic regression

Logistic Loss Function

$$L\left(f_{\vec{w},b}(x^{(i)}), y^{(i)}\right) = \begin{cases} -\log\left(f_{\vec{w},b}(x^{(i)})\right) & \text{if } y^{(i)} = 1 \\ -\log\left(1 - f_{\vec{w},b}(x^{(i)})\right) & \text{if } y^{(i)} = 0 \end{cases}$$

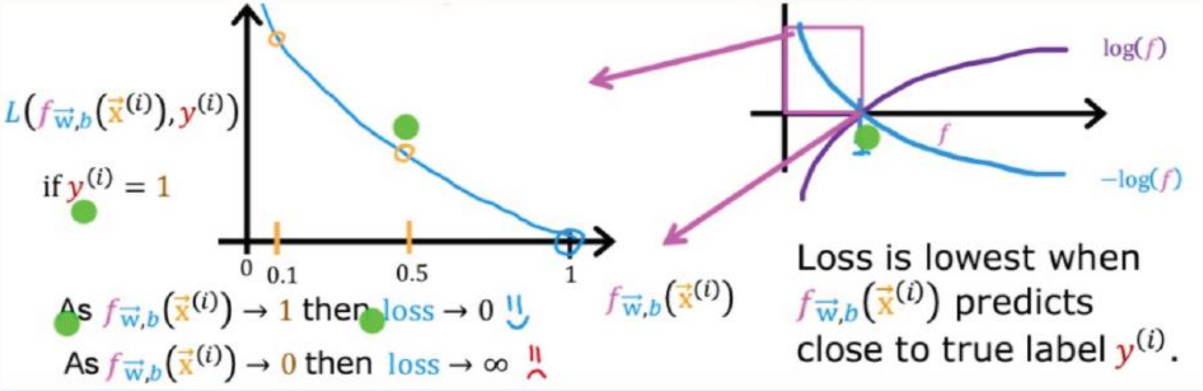
$$L\left(f_{\vec{w},b}(x^{(i)}), y^{(i)}\right) = -y^{(i)} \log\left(f_{\vec{w},b}(x^{(i)})\right) - \left(1 - y^{(i)}\right) \log\left(1 - f_{\vec{w},b}(x^{(i)})\right)$$

Logistic Loss Function

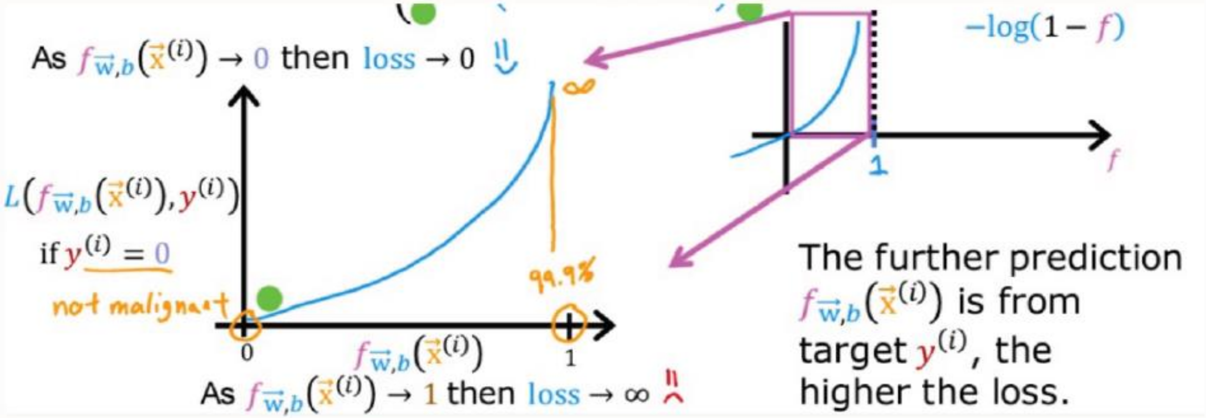
$$L\left(f_{\vec{w},b}(x^{(i)}), y^{(i)}\right) = \begin{cases} -\log\left(f_{\vec{w},b}(x^{(i)})\right) & \text{if } y^{(i)} = 1 \\ -\log\left(1 - f_{\vec{w},b}(x^{(i)})\right) & \text{if } y^{(i)} = 0 \end{cases}$$

$$L\left(f_{\vec{w},b}(x^{(i)}), y^{(i)}\right) = -y^{(i)} \log\left(f_{\vec{w},b}(x^{(i)})\right) - \left(1 - y^{(i)}\right) \log\left(1 - f_{\vec{w},b}(x^{(i)})\right)$$

$y = 1$



$y = 0$



Logistic Loss Function

$$L\left(f_{\vec{w},b}(x^{(i)}), y^{(i)}\right) = \begin{cases} -\log\left(f_{\vec{w},b}(x^{(i)})\right) & \text{if } y^{(i)} = 1 \\ -\log\left(1 - f_{\vec{w},b}(x^{(i)})\right) & \text{if } y^{(i)} = 0 \end{cases}$$

$$L\left(f_{\vec{w},b}(x^{(i)}), y^{(i)}\right) = -y^{(i)} \log\left(f_{\vec{w},b}(x^{(i)})\right) - \left(1 - y^{(i)}\right) \log\left(1 - f_{\vec{w},b}(x^{(i)})\right)$$

Cost Function

$$J(\vec{w}, b) = \frac{1}{m} \sum_{i=1}^{m} L\left(f_{\vec{w},b}(x^{(i)}), y^{(i)}\right)$$

$$= -\frac{1}{m} \sum_{i=1}^{m} \left[y^{(i)} \log\left(f_{\vec{w},b}(x^{(i)})\right) + \left(1 - y^{(i)}\right) \log\left(1 - f_{\vec{w},b}(x^{(i)})\right) \right]$$

LAB-05

COST FUNCTION FOR LOGISTIC REGRESSION

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

MINITEL

03

GRADIENT DESCENT FOR LOGISTIC REGRESSION

THÉO GACHET

Gradient Descent Implementation

cost

$$J(\vec{\mathrm{w}}, b) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log\left(f_{\vec{\mathrm{w}},b}(\vec{\mathrm{x}}^{(i)})\right) + (1-y^{(i)})\log\left(1-f_{\vec{\mathrm{w}},b}(\vec{\mathrm{x}}^{(i)})\right)\right]$$

repeat {

$j = 1 \ldots n$

$$w_j = w_j - \alpha \frac{\partial}{\partial w_j} J(\vec{\mathrm{w}}, b) \qquad \frac{\partial}{\partial w_j} J(\vec{\mathrm{w}}, b) = \frac{1}{m}\sum_{i=1}^{m}\left(f_{\vec{\mathrm{w}},b}(\vec{\mathrm{x}}^{(i)}) - y^{(i)}\right)x_j^{(i)}$$

$$b = b - \alpha \frac{\partial}{\partial b} J(\vec{\mathrm{w}}, b) \qquad \frac{\partial}{\partial b} J(\vec{\mathrm{w}}, b) = \frac{1}{m}\sum_{i=1}^{m}\left(f_{\vec{\mathrm{w}},b}(\vec{\mathrm{x}}^{(i)}) - y^{(i)}\right)$$

} simultaneous updates

repeat {

looks like linear regression!

$$w_j = w_j - \alpha \left[\frac{1}{m} \sum_{i=1}^{m} \left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)} \right) x_j^{(i)} \right]$$

$$b = b - \alpha \left[\frac{1}{m} \sum_{i=1}^{m} \left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)} \right) \right]$$

} simultaneous updates

(but $f_{\vec{w},b}$ is different due to the sigmoid)

Same concepts:

- Monitor gradient descent (learning curve)
- Vectorized implementation
- Feature scaling

Linear regression $\quad f_{\vec{w},b}(\vec{x}) = \vec{w} \cdot \vec{x} + b$

Logistic regression $\quad f_{\vec{w},b}(\vec{x}) = \dfrac{1}{1 + e^{(-\vec{w} \cdot \vec{x} + b)}}$

LAB-06

GRADIENT DESCENT FOR LOGISTIC REGRESSION

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

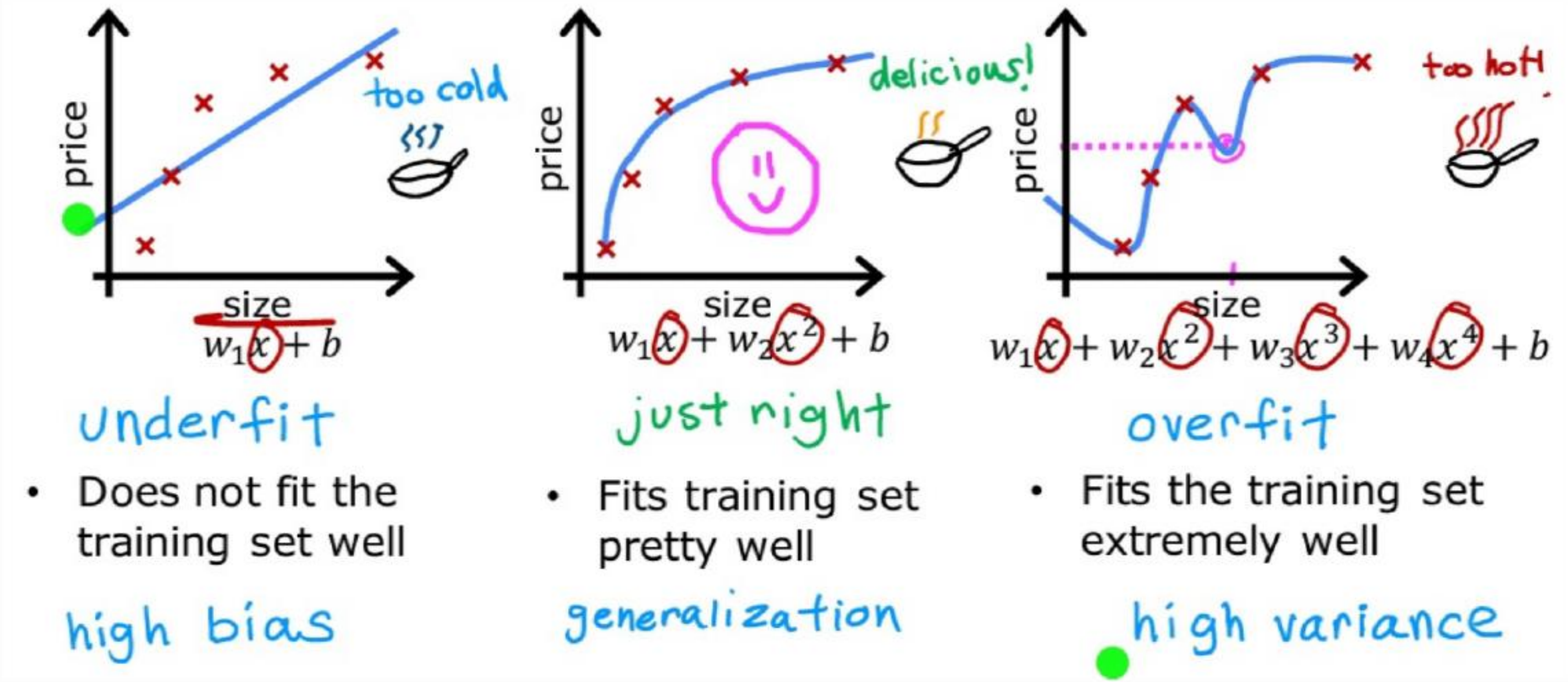
04

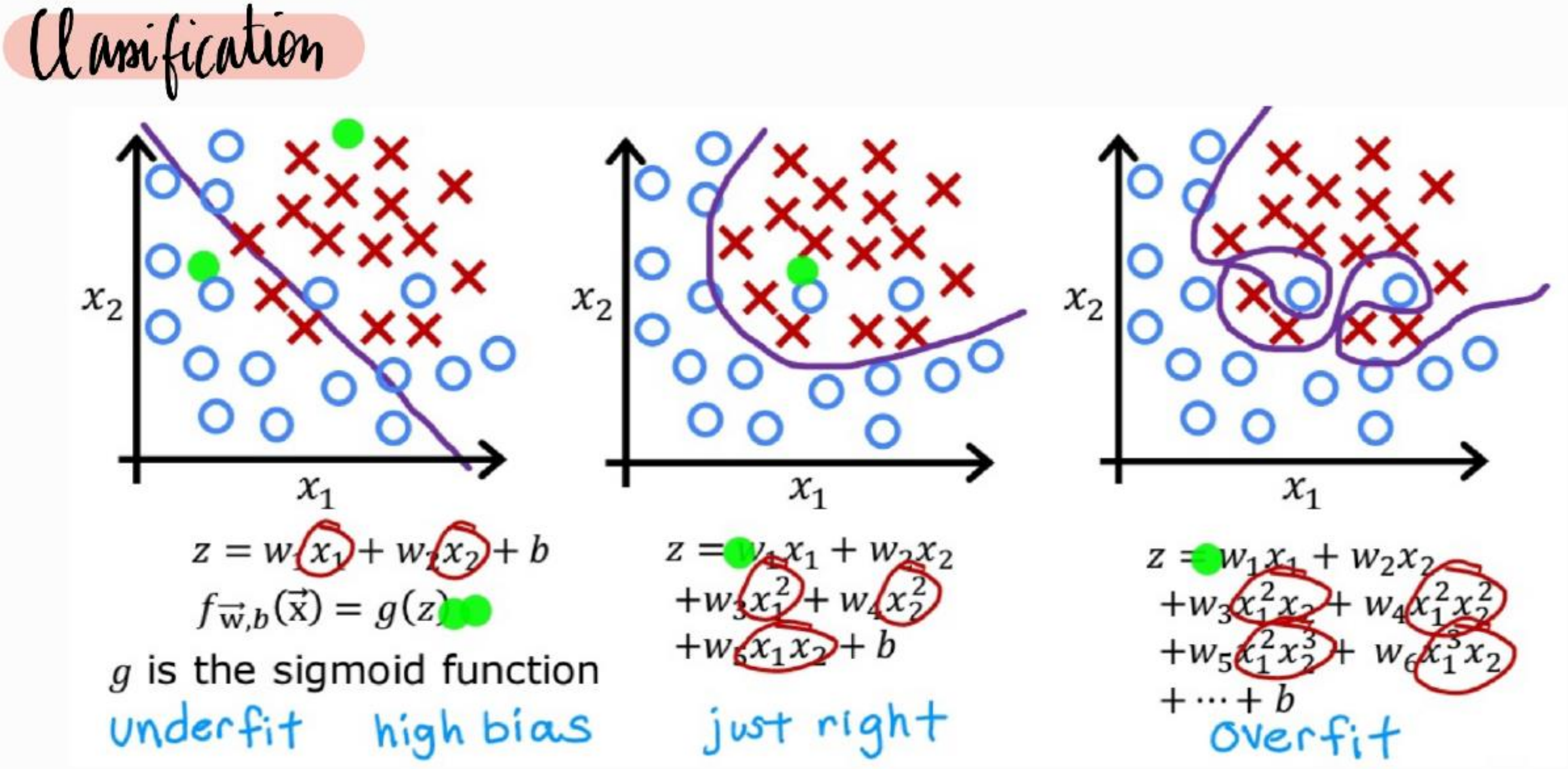
THE PROBLEM OF OVERFITTING

THÉO GACHET

Overfitting

Regression




Classification
x_2
x_1
$z = w_1 x_1 + w_2 x_2 + b$
$f_{\vec{w},b}(\vec{x}) = g(z)$
g is the sigmoid function
underfit high bias
x_2
x_1
$z = w_1 x_1 + w_2 x_2$
$+ w_3 x_1^2 + w_4 x_2^2$
$+ w_5 x_1 x_2 + b$
just right
x_2
x_1
$z = w_1 x_1 + w_2 x_2$
$+ w_3 x_1^2 x_2 + w_4 x_1^2 x_2^2$
$+ w_5 x_1^2 x_2^3 + w_6 x_1^3 x_2$
$+ \cdots + b$
overfit

Addressing Overfitting

- collect more training examples
- select features to include/exclude
- regularization : reduce the weights w_j of some features

Cost Function with Regularization

$$\min_{\vec{w},b} J(\vec{w},b) = \min_{\vec{w},b} \left[\overbrace{\underbrace{\frac{1}{2m}\sum_{i=1}^{m}\left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}\right)^2}_{\text{fit data}}}^{\text{mean square error}} + \overbrace{\underbrace{\frac{\lambda}{2m}\sum_{j=1}^{n} w_j^2}_{\text{keep } w_j \text{ small}}}^{\text{regularization term}} \right]$$

Regularized Linear Regression

$$\min_{\vec{w},b} J(\vec{w},b) = \min_{\vec{w},b}\left[\frac{1}{2m}\sum_{i=1}^{m}\left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}\right)^2 + \frac{\lambda}{2m}\sum_{j=1}^{n} w_j^2\right]$$

Gradient descent

repeat {

$$w_j = w_j - \alpha\frac{\partial}{\partial w_j}J(\vec{w},b) = \frac{1}{m}\sum_{i=1}^{m}\left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}\right)x_j^{(i)} + \frac{\lambda}{m}w_j$$

$j = 1, \dots, n$

$$b = b - \alpha\frac{\partial}{\partial b}J(\vec{w},b) = \frac{1}{m}\sum_{i=1}^{m}\left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}\right)$$

} simultaneous update

don't have to regularize b

$$w_j = \underbrace{1 w_j - \alpha\frac{\lambda}{m} w_j}_{w_j\left(1 - \alpha\frac{\lambda}{m}\right)} \underbrace{- \alpha\frac{1}{m}\sum_{i=1}^{m}\left(f_{w,b}(\vec{x}^{(i)}) - y^{(i)}\right)x_j^{(i)}}_{\text{usual update}}$$

shrink $w_j \longrightarrow \left(1 - \frac{\lambda}{m} < 1\right)$

Regularized Logistic Regression

$$J(\vec{w},b) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log\left(f_{\vec{w},b}(\vec{x}^{(i)})\right) + (1-y^{(i)})\log\left(1-f_{\vec{w},b}(\vec{x}^{(i)})\right)\right] + \frac{\lambda}{2m}\sum_{j=1}^{n} w_j^2$$

$\min_{\vec{w},b}$

Looks same as for linear regression!

Gradient descent

repeat {

$$w_j = w_j - \alpha \frac{\partial}{\partial w_j} J(\vec{w},b) \qquad = \frac{1}{m}\sum_{i=1}^{m}\left[\left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}\right)x_j^{(i)}\right] + \frac{\lambda}{m} w_j$$

$j=1\ldots n$

logistic regression

$$b = b - \alpha \frac{\partial}{\partial b} J(\vec{w},b) \qquad = \frac{1}{m}\sum_{i=1}^{m}\left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}\right)$$

}

don't have to regularize b

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

LAB-09

REGULARIZATION

03 CLASSIFICATION

CLASSIFICATION WITH LOGISTIC REGRESSION

Motivations
Lab : Classification
Logistic regression
Lab : Sigmoid function and logistics
Decision boundary
Lab : Decision boundary

COST FUNCTION

Cost function for logistic regression
Lab : Logistic loss
Simplified cost function for logistic regression
Lab : cost function

GRADIENT DESCENT

Gradient Descent implementation
Lab : Gradient descent
Lab : Logistic regression with scikit-learn

THE PROBLEM OF OVERFITTING

The problem of overfitting
Addressing overfitting
Lab : Overfitting
Cost function with regularization
Regularized linear regression
Regularized logistic regression
Lab : Regularization

MINITEL

MACHINE LEARNING

03 CLASSIFICATION

THÉO GACHET