MACHINE LEARNING (IMC-4302C)

LAB 3: LOGISTIC REGRESSION

SLIM BEN AMOR: SLIM.BEN-AMOR@INRIA.FR

School Year
2017-2018

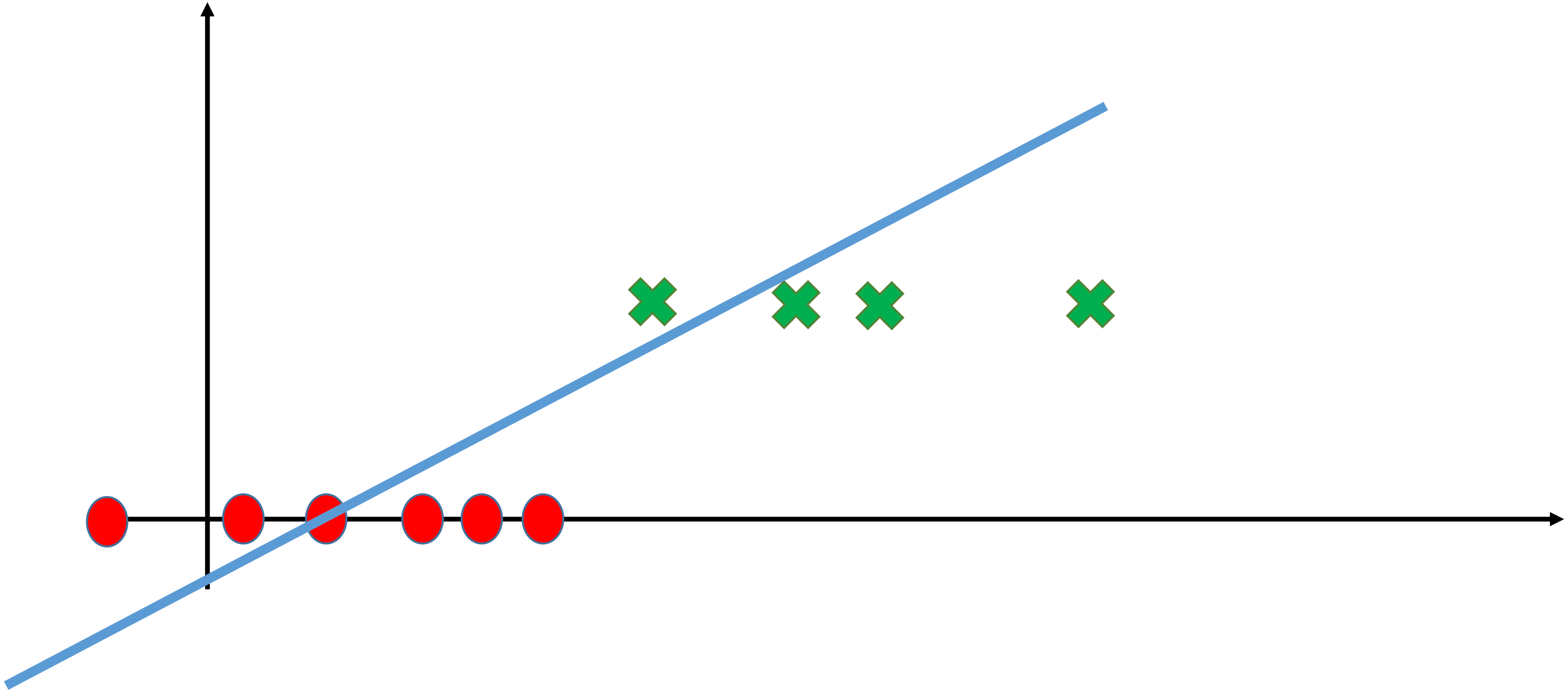
13/2/2018

About this session

- **OUTLINE:**
  - Correction of last lab
  - Working on lab 3

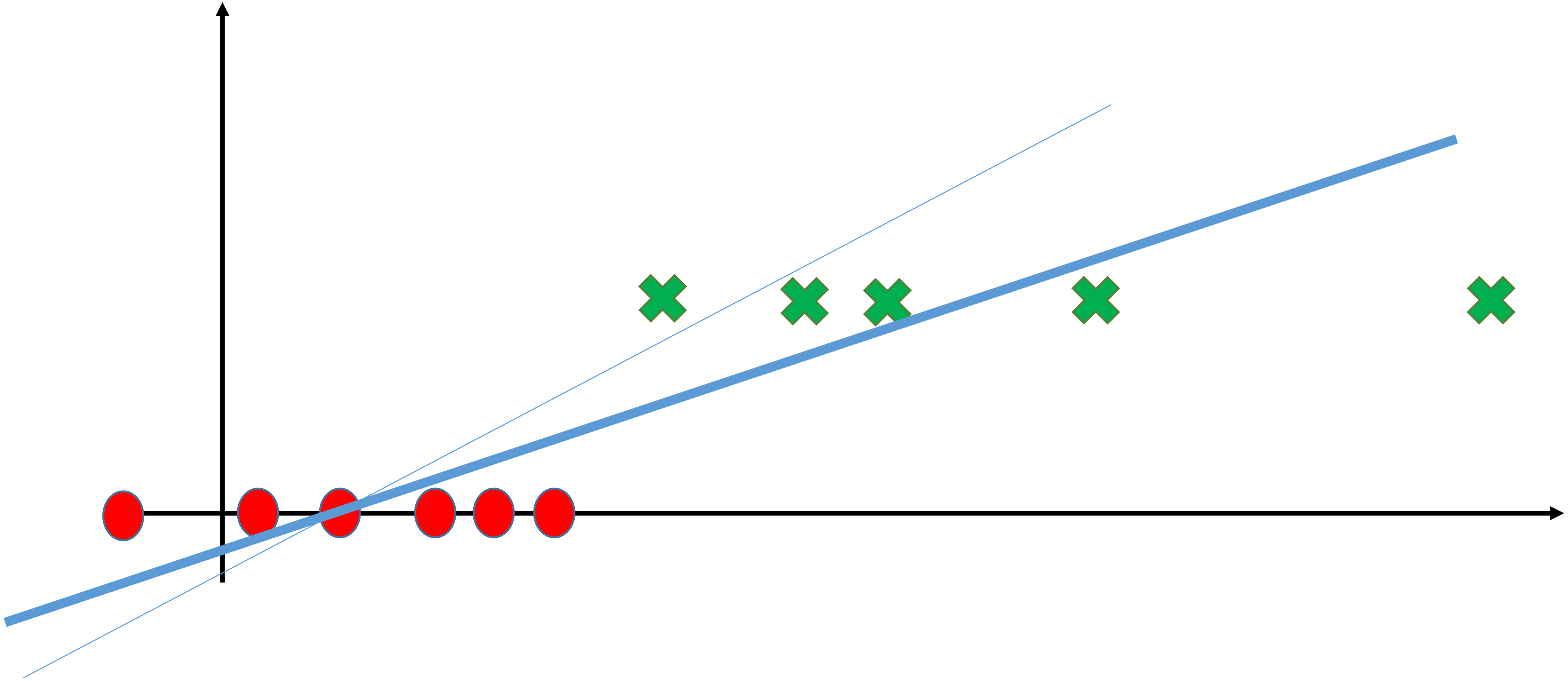
- Make groups of 2 persons to work on the coming labs and on the final project.
- For the report, you may submit one per group.

Why not linear regression



Why not linear regression

If additional positive point:



Why not linear regression

- The linear model is very sensitive to additional data.

- We could get: $h_\theta(x) > 1 \; or \; h_\theta(x) < 0$ which is wrong because $h_\theta(x)$ represent a probability distribution.

Logistic Cost Function

- Cost Function $J(\theta) = \frac{-1}{m} \sum_{i=1}^{m} [y \times log(h_\theta(x)) + (1-y) \times log(1-h_\theta(x))]$

  Where: $h_\theta(x_i) = P(x_i) = \frac{e^{\theta^\top x_i}}{1+e^{\theta^\top x_i}} = \frac{1}{1+e^{-\theta^\top x_i}} = sigmoid(\theta^\top x_i)$

- Partial derivative $$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x_i) - y_i)x_j$$

$$j = 0 \ldots n-1$$

Vectorized Implementation

- Cost function

$$J(\theta) = \frac{-1}{m}\ sum[y \times log(sigmoid(X\theta)) + (1 - y) \times log(1 - sigmoid(X\theta))]$$

- Cost function gradient

$$\nabla J(\theta) = \begin{bmatrix} \frac{\partial J(\theta)}{\partial \theta_0} \\ \vdots \\ \frac{\partial J(\theta)}{\partial \theta_{n-1}} \end{bmatrix} = \frac{1}{m} X^\top (h_\theta(X) - y) = \frac{1}{m} X^\top (sigmoid\ (X\theta) - y)$$

- Gradient descent update $\theta = \theta - \alpha \nabla J(\theta)$