CENTRE DE RECHERCHE EN AUTOMATIQUE DE NANCY

IMPACT OF COST-SENSITIVITY AND OUTLIERS ROBUSTNESS IN PERCEPTRON LEARNING

P. THOMAS

AM2I

INS2I
INSIS

UMR 7039

UNIVERSITÉ DE LORRAINE

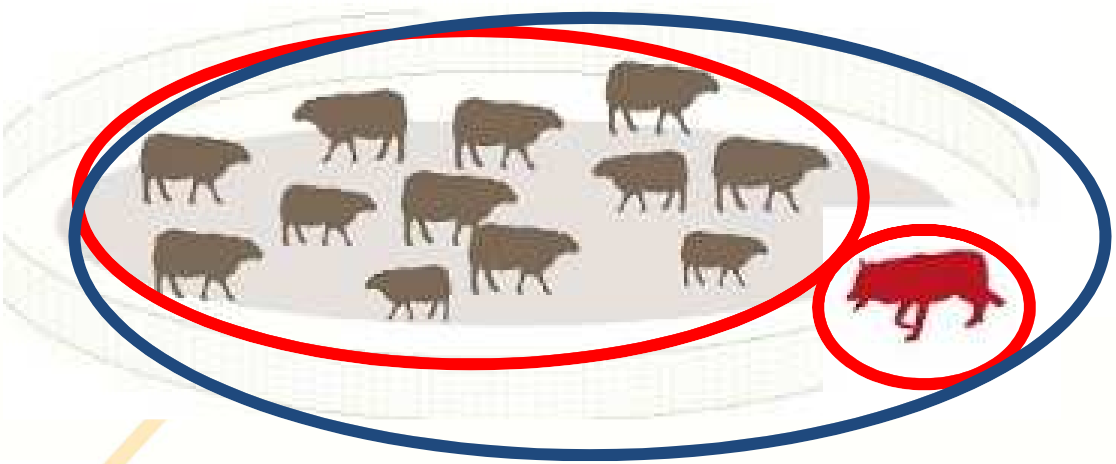
Institut de Cancérologie de Lorraine
Alexis Vautrin
Ensemble, construisons l'avenir

SUMMARY

- **Introduction**
  - Imbalanced data set
  - Industrial data and outliers
- **Multilayer perceptron**
  - Structure
  - Criterion to minimize
  - Robust-cost sensitive learning algorithm
- **Simulation example**
  - Experimental protocol
  - Results on outliers free dataset
  - Results on outliers polluted dataset
- **Conclusion**

IMBALANCED DATASET

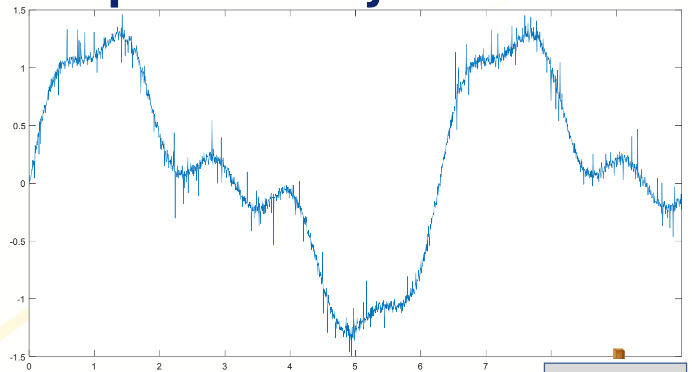
- Machine learning needs dataset !
- Classification goal: affect the good label to each pattern
- In many cases (quality monitoring, medical diagnosis, credit risk prediction…)
  - Classes are imbalanced
  - Some very bad model may have a good score
    - Leading to undesirable event !



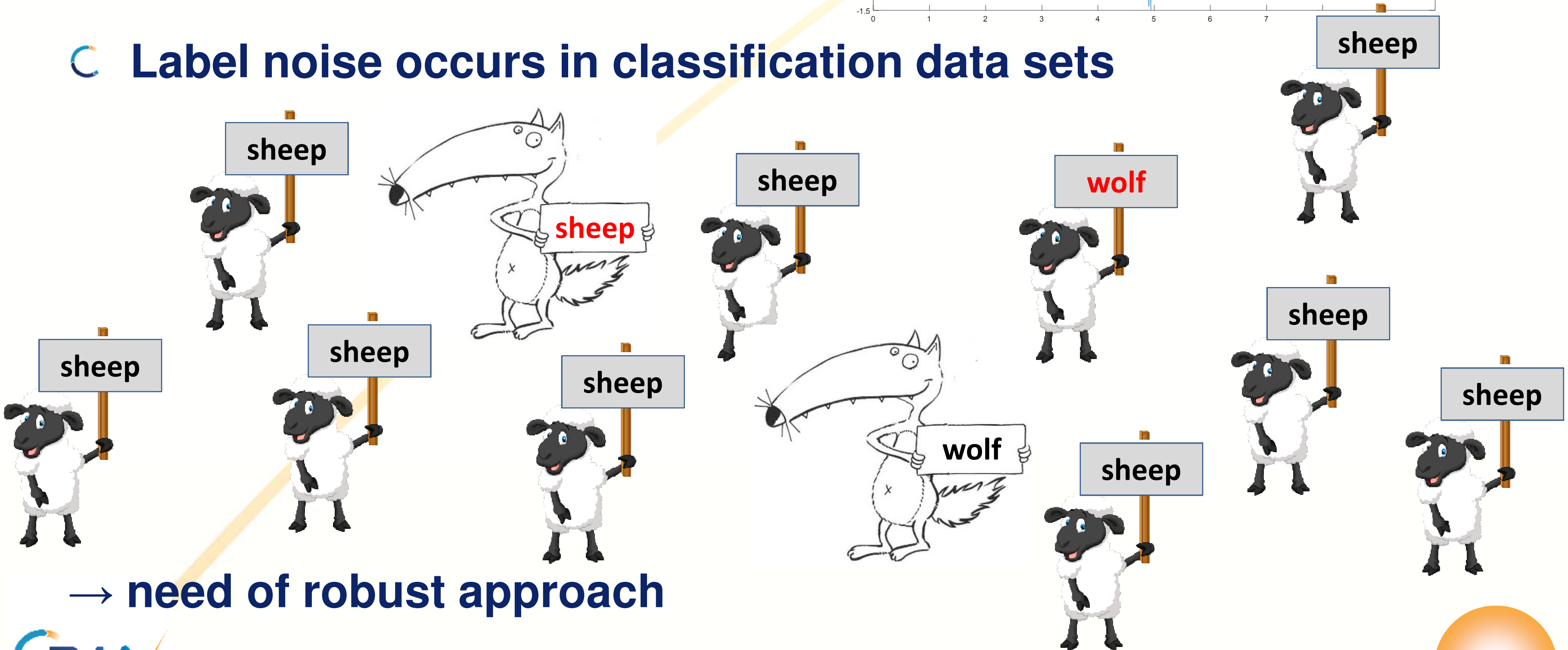
⟶ need of cost sensitive approach

ERROR SOURCES: NOISE, OUTLIERS AND LABEL NOISE

- **All industrial data sets are noisy and polluted by outliers**
  - Up to 10% of data are outliers
    - (Hampel 1971)



- **Label noise occurs in classification data sets**



→ need of robust approach

AN INDUSTRIAL EXAMPLE: ACTA MOBILIER

- **Lacquered panels manufactured for kitchen, stands, shops…**
- **Very high quality requirement for the surface**
- **Main defects are generated at the lacquering step**
- **Quality monitoring:**
  - 7 basics tools of quality
    - Detection of a process variation (after defects production)
  - Optimal Experimental Design
    - Setup robust to variation of some parameters (before defects production)
  - High quality requirement implies that process is often used at its technological limits
    - Robust setup may be insufficient

⟶ **necessity to be on-line**

Using of quality prediction model

Data mining approach

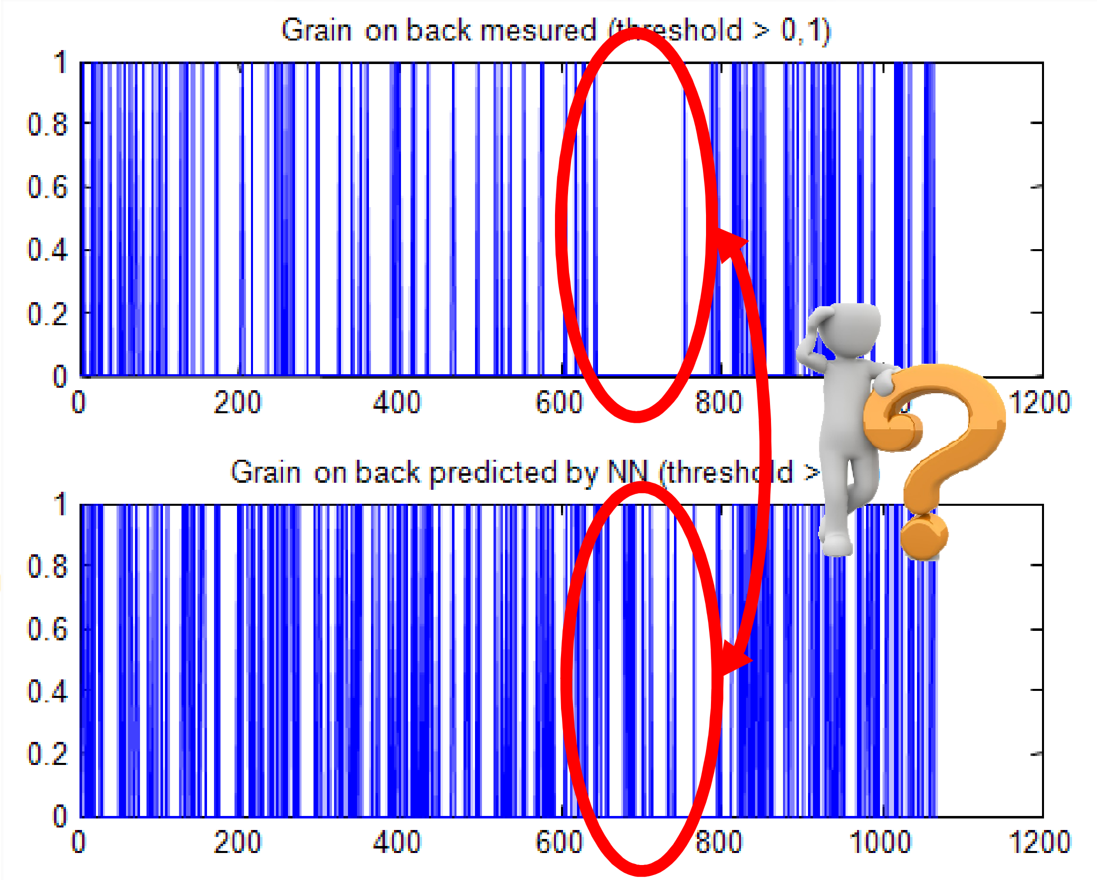
AN INDUSTRIAL EXAMPLE: DATA SET

- **Quality monitoring problem of a high quality lacquering robot**
  - Defects rate important and fluctuating (10% to 45%)
  - 25 different types of defects may be produced
  - Expert knowledge allows to identify impacting factors
    - Environmental factors
      - Temperature NOISE
      - Humidity NOISE
      - pressure NOISE
    - Setup parameters
      - load factor NOISE
      - basis weight NOISE
      - Product number
    - Routing parameters
      - number of passes
      - time per table NOISE
      - liter per table NOISE
      - number of layers
      - drying time NOISE

INTRO

AN INDUSTRIAL EXAMPLE: RESULTS

- For one type of defect



DEFECT OCCURRENCE IN DATASET

Non-detection rate : 11,8 %

False positive : 19,2 %

DEFECT PREDICTION BY MODEL

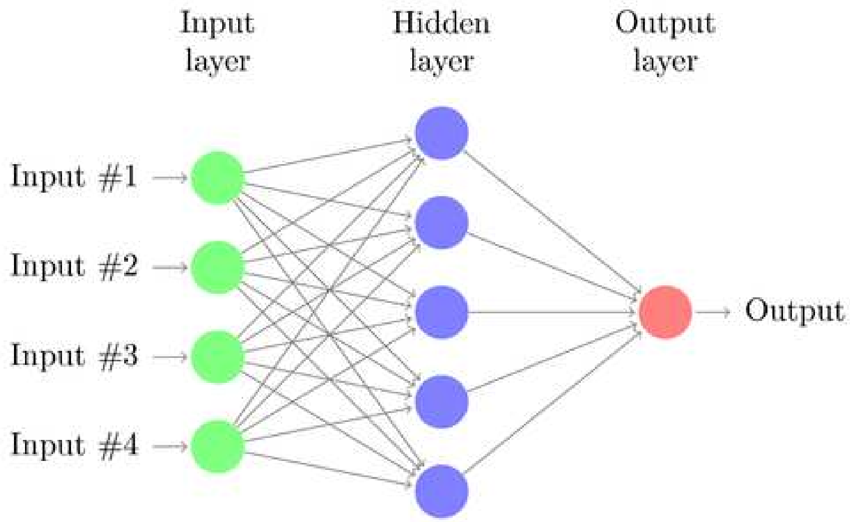
Interview of the manager:

- quality data manually collected
- absence replacement by temporary worker

INTRO

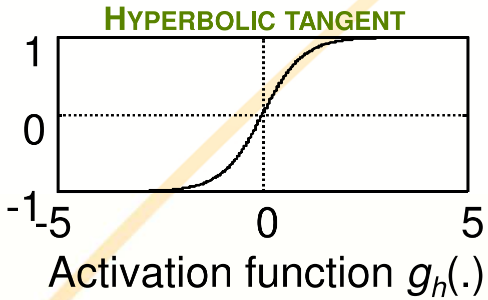
MULTILAYERS PERCEPTRON MLP: STRUCTURE

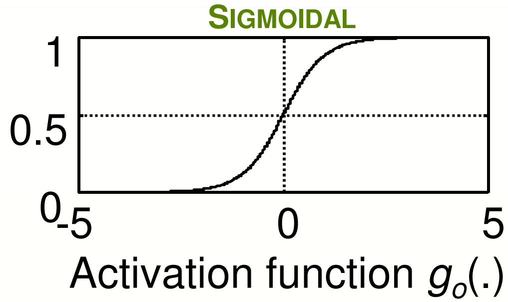
- **A neural network:**
  - Exploitation of a collected data
  - Simple implementation (neural model design partially automated)
  - Improving and adaptation on-line of the process



- **The multilayers perceptron:**
  - Universal approximator

$$z = g_2\left(\sum_{i=1}^{n_1} w_i^2 . g_1\left(\sum_{h=1}^{n_0} w_{ih}^1 . x_h^0 + b_i^1\right) + b\right)$$





- **Weights initialization (Nguyen and Widrow, 1990)**

MLP

MLP: Criterion to minimize

- **The classical criterion to minimize:** $V(\theta)=\frac{1}{2n}\sum_{k=1}^{n}\varepsilon^2(k,\theta)$
  - Hyp: Gaussian noise distribution
  - Where the prediction error: $\varepsilon(k,\theta)=y(k)-\hat{y}(k,\theta)$
    - Greater is the error → Greater is its influence on criterion value
    - Quid of the outliers and label noise?

- **Robust criterion (weighted by noise variance):** $V(\theta)=\frac{1}{2n}\sum_{k=1}^{n}\left(\frac{\varepsilon^2(k,\theta)}{\sigma^2(k)}\right)$
  - Hyp: mixture of Gaussian (Huber's Model):

$$e\sim(1-\mu)\,N(0,\sigma_1^2)+\mu\,N(0,\sigma_2^2)$$

  - Robust weight: $\sigma^2(k)=(1-\delta(k))\,\hat{\sigma}_1^2(i)+\delta(k)\,\hat{\sigma}_2^2(i)$ with: $\begin{cases}\hat{\sigma}_1(i)=\frac{MAD}{0.7}\\ \hat{\sigma}_2(i)=3.\hat{\sigma}_1(i)\end{cases}$

- **Robust cost sensitive criterion:** $V(\theta)=\frac{1}{2n}\sum_{k=1}^{n}\left(C_{ost}(k).\frac{\varepsilon^2(k,\theta)}{\sigma^2(k)}\right)$
  - Cost of misclassification:

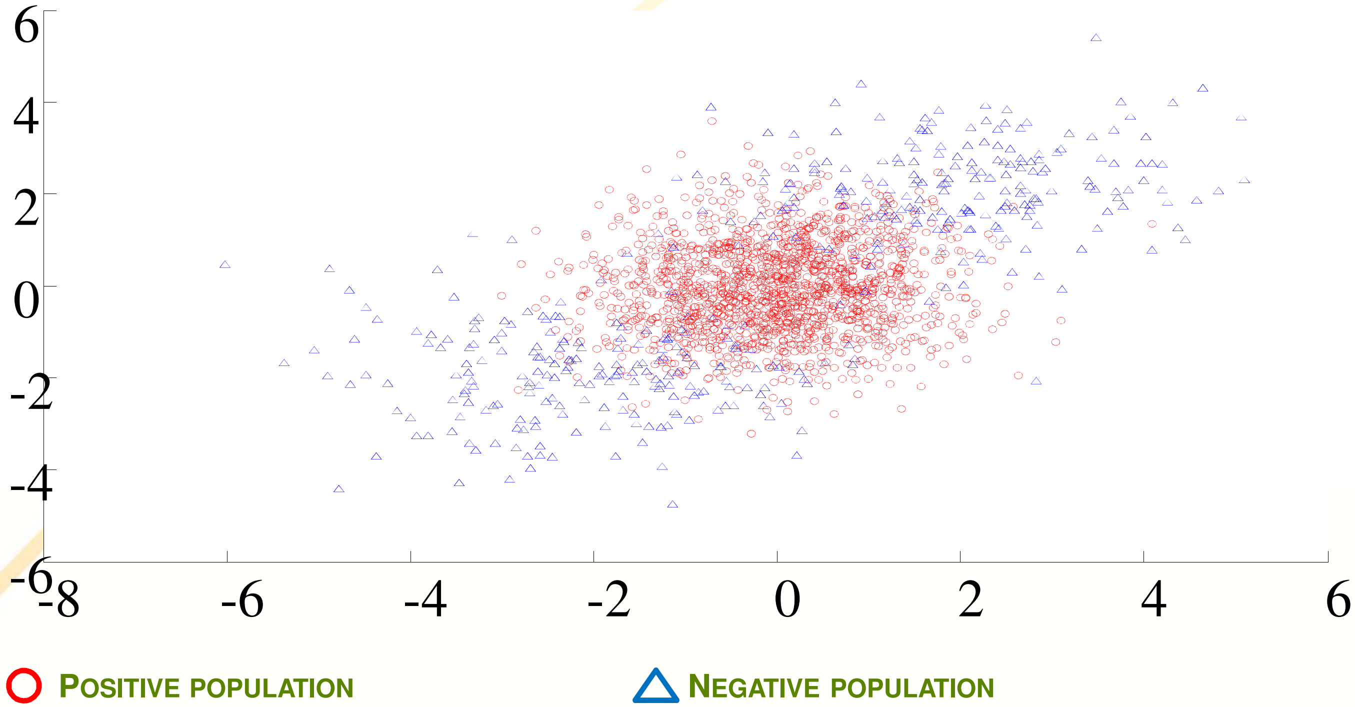
		predicted class	
		Class 0	Class1
real class	Class 0	C_{00}	C_{01}
	Class 1	C_{10}	C_{11}

MLP: ROBUST-COST SENSITIVE LEARNING ALGORITHM

- **The criterion to minimize:** $$V(\theta)=\frac{1}{2n}\sum_{k=1}^{n}\left(C_{ost}(k).\frac{\varepsilon^2(k,\theta)}{\sigma^2(k)}\right)$$
- **2nd order Taylor series expansion of *V*(*θ*):** $$\hat{\theta}^{i+1}=\hat{\theta}^{i}-(H(\hat{\theta}^{i}))^{-1}V^{'}(\hat{\theta}^{i})$$
  - Gradient of the criterion:
  $$V^{'}(\theta)=-\frac{1}{n}\sum_{k=1}^{n}\psi(k,\theta).C_{ost}(k).\frac{\varepsilon(k,\theta)}{\sigma^2(k)}$$
  - Estimation of the Hessian Matrix (Levenberg-Marquardt):
  $$H(\theta)=\frac{1}{n}\sum_{k=1}^{n}\psi(k,\theta)\frac{C_{ost}(k)}{\sigma^2(k)}\psi^{T}(k,\theta)+\beta I$$
  - Where $\Psi(k,\theta)$: the gradient of the network output $\hat{y}(k,\theta)$ with respect to θ.

SIMULATION EXAMPLE: DATASET

- **Population constituted with two subpopulations**
  - Positive subpopulation
    - Bivariate normal distribution with mean $(0, 0)^T$ and covariance matrix diag(1, 1)
  - Negative subpopulation
    - Bivariate normal distribution with mean $(2, 2)^T$ and covariance matrix diag(2, 1)
    - Bivariate normal distribution with mean $(-2, -2)^T$ and covariance matrix diag(2, 1)



SIMULATION EXAMPLE: PROTOCOL

- **Dataset comprising 2000 patterns**
  - 1000 for the learning
  - 1000 for the validation
- **Evaluation criterion: zero-one score function**

$$S_{01} = \frac{1}{n}\sum_{k=1}^{n} I\left(y(k), \hat{y}(k,\theta)\right)$$

- **Two other indicators:**
  - False Alarm rate (FA) $FA = \frac{FalsePos}{FalsePos + TrueNeg}$
  - Non-Detection rate (ND) $ND = \frac{FalseNeg}{FalseNeg + TruePos}$
- **Misclassification cost:**

$$Cost = C_{01}.FalsePos + C_{10}.FalseNeg$$

		predicted class	
		Class 0	Class1
real class	Class 0	1	2
	Class 1	5	1

		predicted class	
		Class 0	Class1
real class	Class 0	1	2
	Class 1	10	1

Results on outlier free dataset

- **Learning of MLP with 2 inputs and 10 hidden neurons**
- **Four different learning algorithms**
  - Classical Levenberg-marquardt (LM)
  - Robust Levenberg-marquardt (RLM)
  - Classical Levenberg-marquardt with cost (LMC)
  - Robust Levenberg-marquardt with cost (RLMC)

		Cost	S_{01}	FA rate	ND rate
$Cost_{01} = 2$, $Cost_{10} = 5$	Without Robust Without Cost	346	9.50%	5.40%	25.49%
	With Robust Without Cost	291	8.10%	4.77%	21.08%
	Without Robust With Cost	281	8.50%	6.03%	18.14%
	With Robust With Cost	290	8.80%	6.28%	18.63%
$Cost_{01} = 2$, $Cost_{10} = 10$	Without Robust Without Cost	606	9.50%	5.40%	25.49%
	With Robust Without Cost	506	8.10%	4.77%	21.08%
	Without Robust With Cost	446	9.90%	8.54%	15.20%
	With Robust With Cost	396	10.60%	10.43%	11.27%

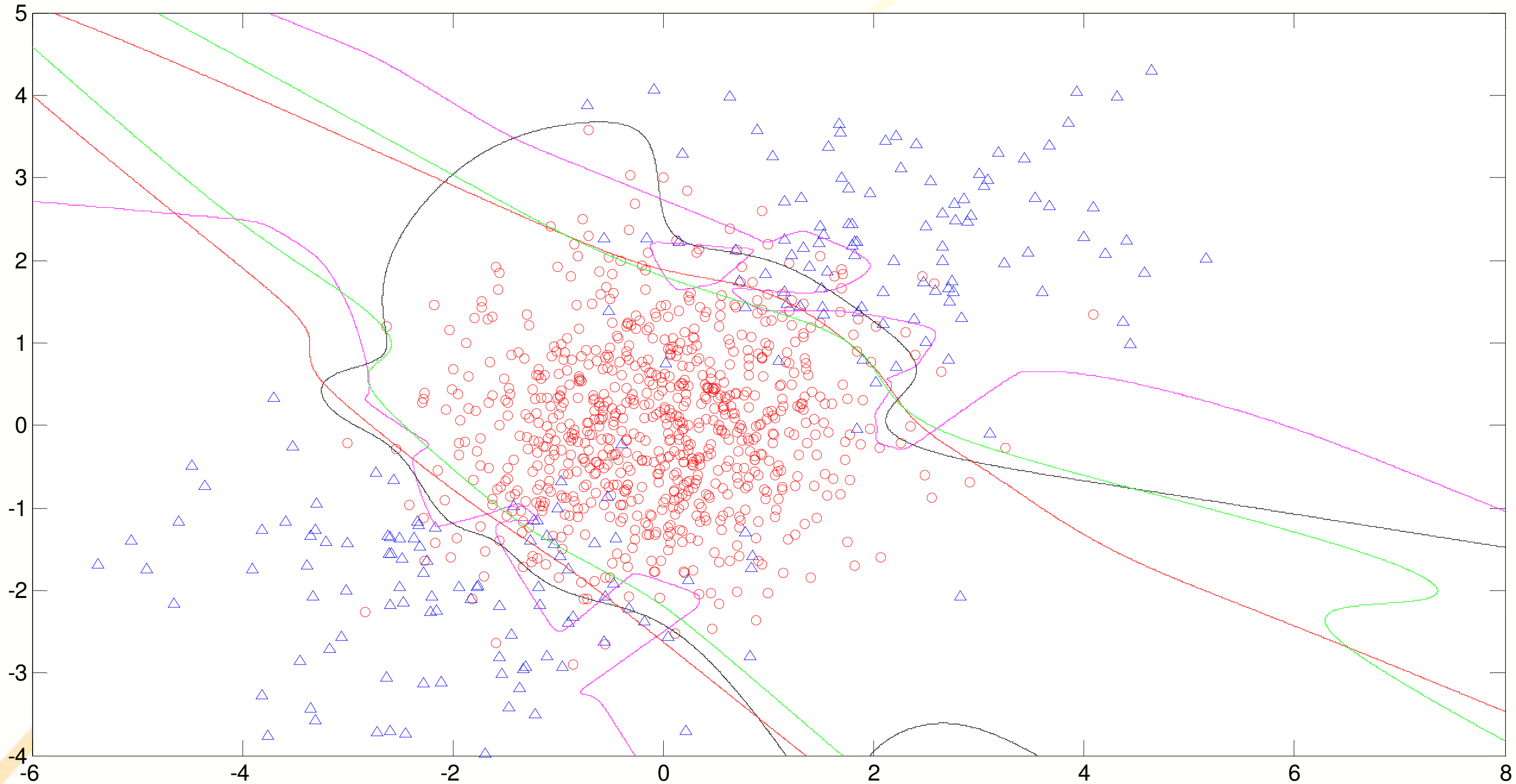
RESULTS ON OUTLIERS POLLUTED DATASET

- Learning dataset corrupted by 10% of noise label
- Same learning algorithms

		Cost	S_{01}	FA rate	ND rate
$Cost_{01} = 2$, $Cost_{10} = 5$	Without Robust Without Cost	381	9.90%	4.77%	29.90%
	With Robust Without Cost	305	8.20%	4.40%	23.40%
	Without Robust With Cost	333	8.40%	3.64%	26.96%
	With Robust With Cost	310	8.90%	5.65%	21.57%
$Cost_{01} = 2$, $Cost_{10} = 10$	Without Robust Without Cost	686	9.90%	4.77%	29.90%
	With Robust Without Cost	540	8.20%	4.40%	23.40%
	Without Robust With Cost	482	9.30%	7.04%	18.14%
	With Robust With Cost	384	10.40%	10.30%	10.78%

CLASSES BOUNDS (OUTLIERS POLLUTED DATASET)

- **LM** in magenta
- **RLM** in black
- **LMC** in red
- **RLMC** in green



SIMU

CONCLUSION

- **Classification model must take into account**
  - The risk of label noise
  - The cost of misclassification

- **Modification of the criterion to minimize**
  - Including of robust cost
  - Including of misclassification cost

- **The combination of robust cost and misclassification cost allows to:**
  - limit the impact of outliers and label noise
  - Favor the non-detection rate comparing to the false alarme rate

CENTRE DE RECHERCHE EN AUTOMATIQUE DE NANCY

IMPACT OF COST-SENSITIVITY AND OUTLIERS ROBUSTNESS IN PERCEPTRON LEARNING

THANK YOU FOR YOUR ATTENTION!

AM2I

INS2I
INSIS

UMR 7039

UNIVERSITÉ DE LORRAINE

Institut de Cancérologie de Lorraine
Alexis Vautrin
Ensemble, construisons l'avenir