# PERFORMANCE EVALUATION I METRICS FOR CLASSIFICATION

Dr. Lablanche Pierre-Yves
African Institute for Mathematical Sciences

# WHAT IS A GOOD CLASSIFIER?

The ultimate Classifier :

- Is very accurate !
- Is never confused !
- Is a 100% sure !
- Is never wrong !


**Need to evaluate performance.**

# ACCURACY

First straightforward metric : compute the **accuracy**

- **Accuracy** : Fraction of samples that have been correctly classified. For ***N*** samples :

$$Accuracy = \frac{N_{correct}}{N}$$

# ACCURACY ISSUE

Imagine an unbalanced data set with 92 dogs and 8 cats.

Once trained the classifier returns only dogs ⟶ the accuracy is 92% !!!
the accuracy is good but the classifier is apparently bad.

Need more different metrics :

- Precision
- Recall

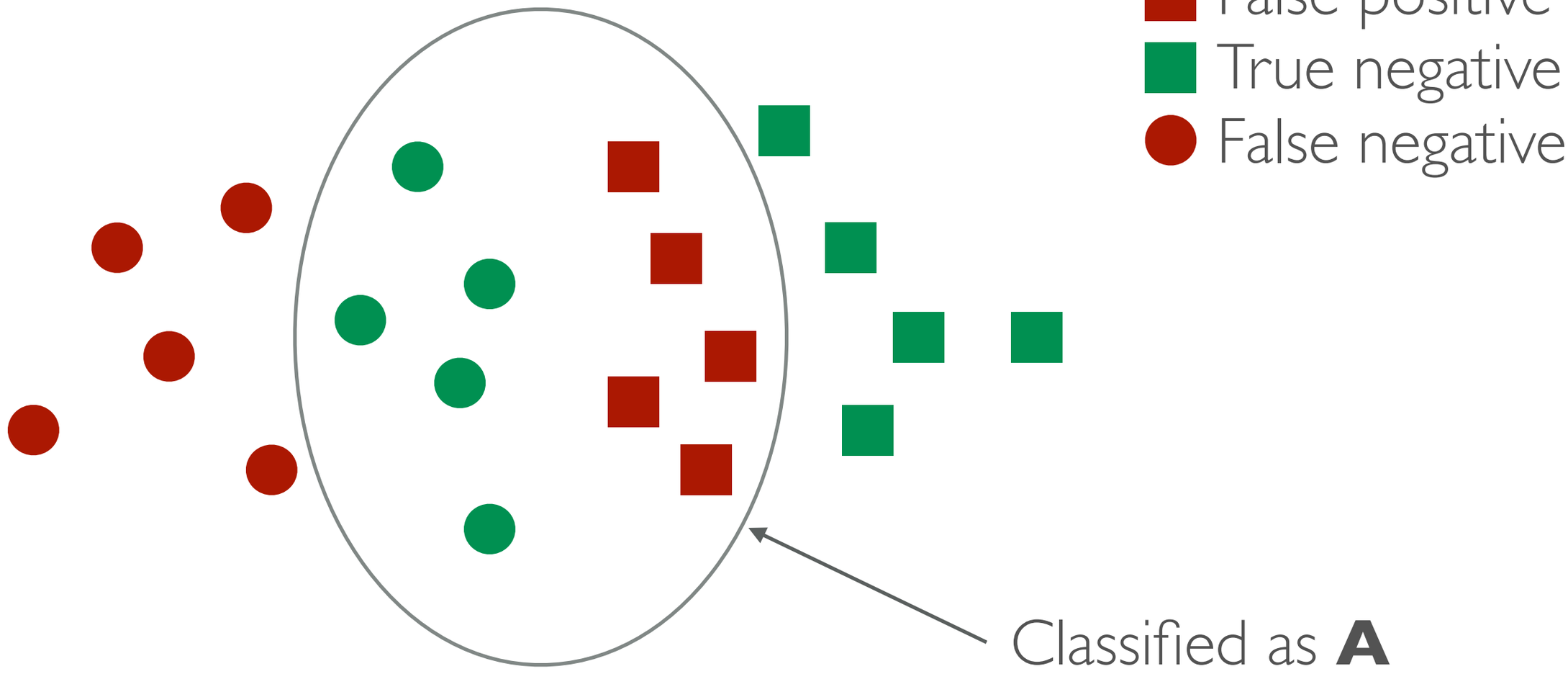
# TRUE AND FALSE PREDICTION

For a given class **A**

- True positive : sample of class **A** classified as **A**
- False positive : sample of a different class classified as **A**
- True negative : sample of a different class not classified as **A**
- False negative : sample of class **A** not classified as **A**

# TRUE AND FALSE PREDICTION

Circles : samples of class **A**
Squares : other class



# PRECISION AND RECALL

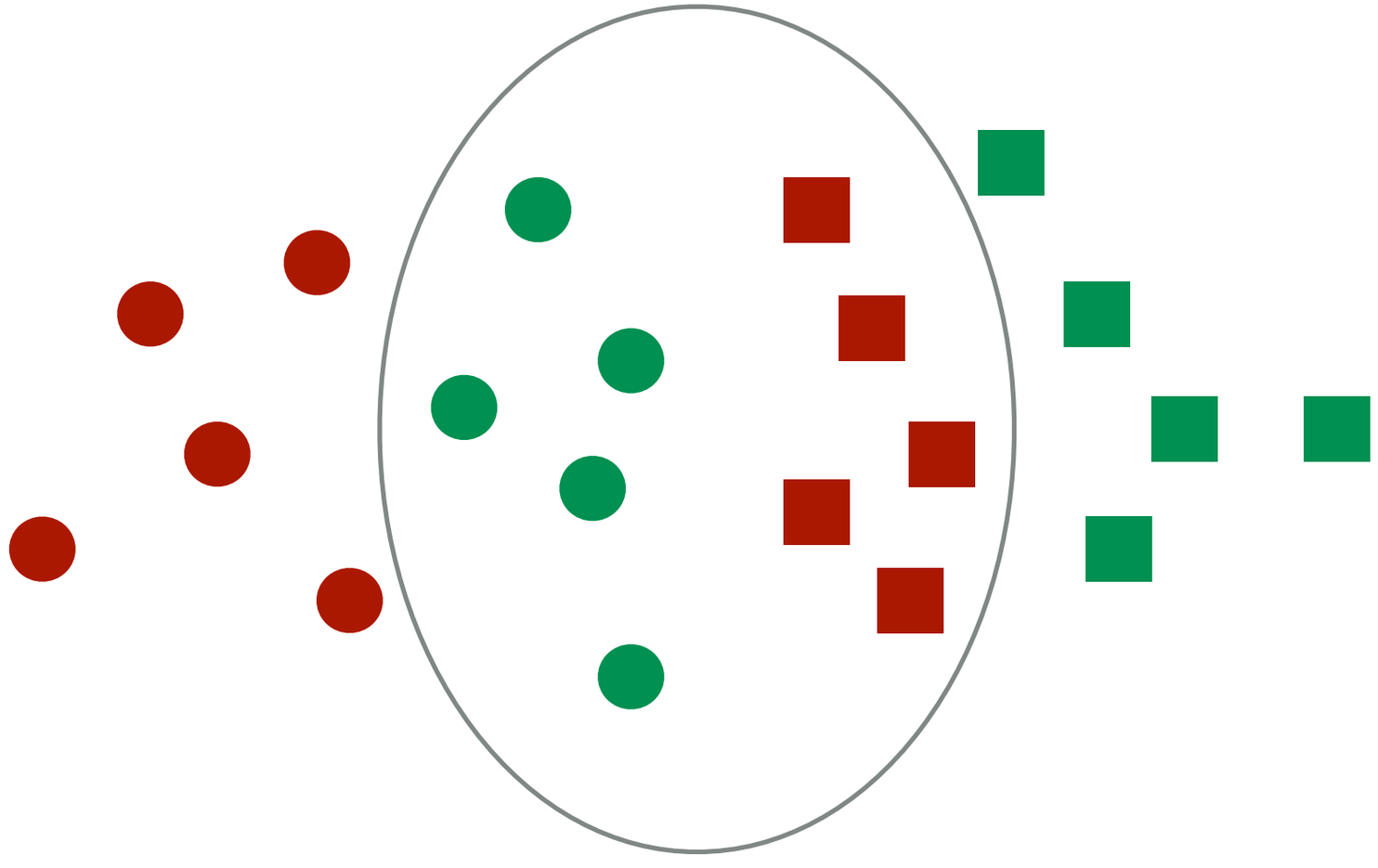
**Precision** : number of true positives divided by the sum of true positives and false positives

$$Precision = \frac{N_{TruePositive}}{N_{TruePositive} + N_{FalsePositive}}$$

**Recall** : number of true positives divided by the sum of true positives and false negatives

$$Recall = \frac{N_{TruePositive}}{N_{TruePositive} + N_{FalseNegative}}$$

# PRECISION AND RECALL



True positive
False positive
True negative
False negative

$$\text{Precision} = \frac{\bullet}{\bullet + \blacksquare}$$

$$\text{Recall} = \frac{\bullet}{\bullet + \bullet}$$

# PRECISION AND RECALL

In other words :

**Precision** :

fraction of samples classified as **A** that are truly from class **A**

**Recall** :

fraction of samples from class **A** that are correctly classified as **A**

Choosing to favour precision over recall or recall over precision depends on the problem !

Example : detection of rocks vs landmines… think about it

# F1 SCORE

F1-Score evaluates balance between precision and recall :

$$F1 = \frac{2(precision \times recall)}{precision + recall}$$

$$= \frac{2 \times N_{TruePositive}}{2 \times N_{TruePositive} + N_{FalsePositive} + N_{FalseNegative}}$$

Perfect classifier get F1 = 1.0 !

# CONFUSION

When the number of class>2, useful to look in more details where a classifier went wrong. How ? Answer : **Confusion Matrix** !

Confusion Matrix = $[C_{ij}]$ where $C_{ij}$ is the number (or fraction) of samples from class $i$ classified as being from class $j$.

Hence :

- diagonal elements are number of correctly classified samples.
- non-diagonal elements are number of misclassified samples.

# CONFUSION

Illustration : 3 classes **A**, **B** and **C**

| | | | |
|---|---|---|---|
| **A** | **A** correctly classified | **A** mistaken as **B** | **A** mistaken as **C** |
| **B** | **B** mistaken as **A** | **B** correctly classified | **B** mistaken as **C** |
| **C** | **C** mistaken as **A** | **C** mistaken as **B** | **C** correctly classified |
| | **A** | **B** | **C** |

Confusion Matrix elements *Cij* are numbers or fractions

# CODING TIME

- Use two classifiers of your choice on the iris dataset.
- Use the different metrics on the trained classifiers;
- Do results make sense ?