

Inria



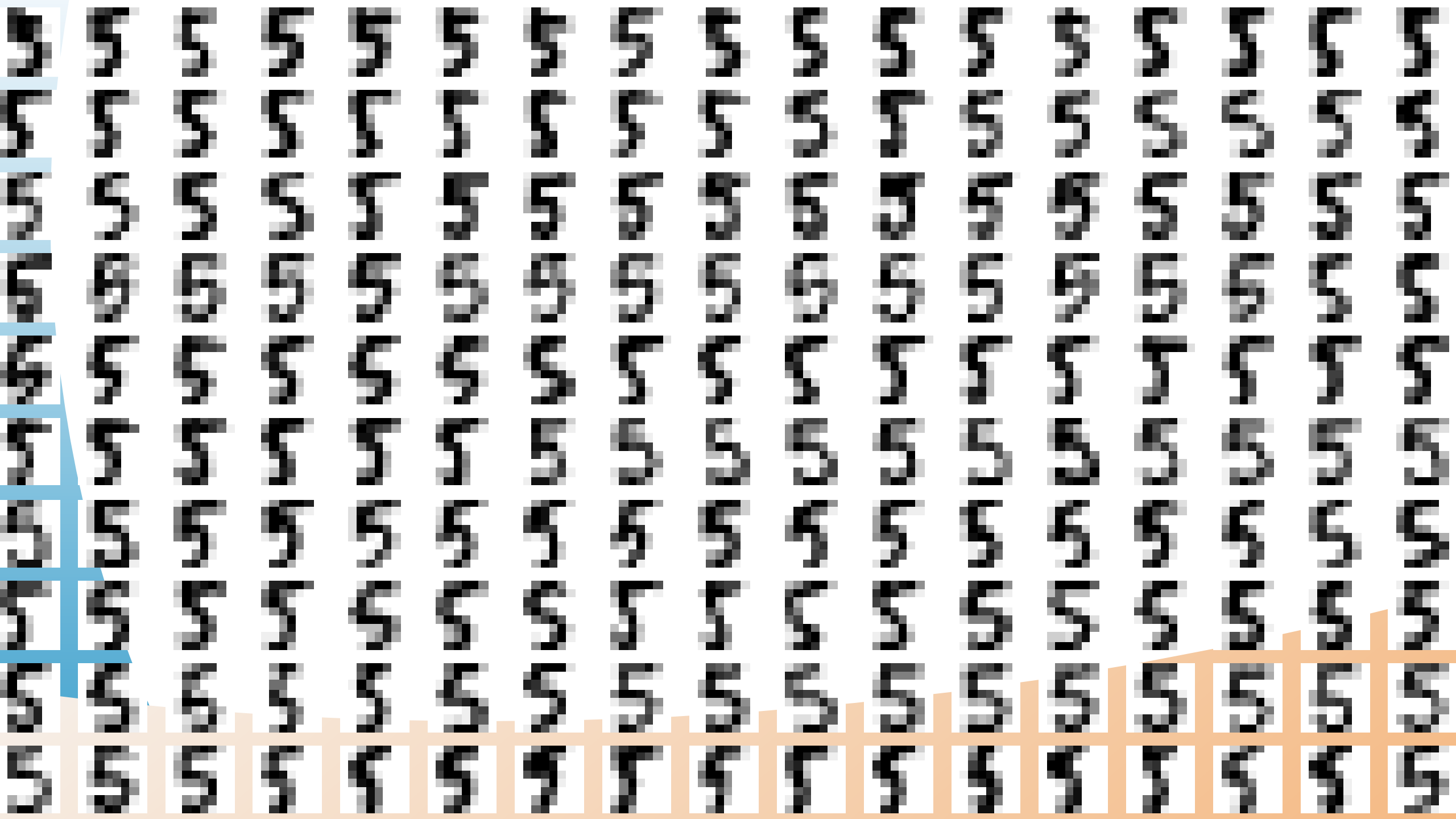
Alexandre Gramfort

Chercheur Inria
Membre du comité technique de scikit-
learn

Inria

L'apport de scikit-learn à la détection d'anomalies





A large field of green apples with one red apple in the center. The red apple is the only one of its color in the field, making it stand out as an anomaly.

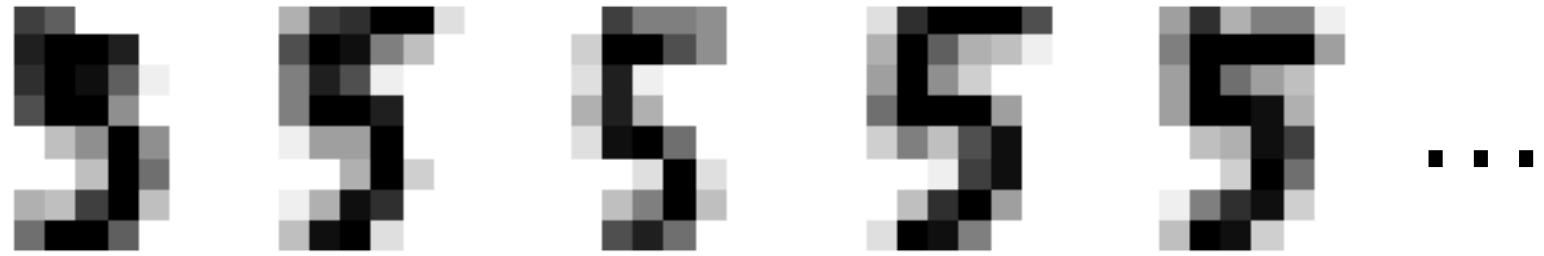
Objectif de la détection
d'anomalie:

Trouver la pomme
rouge

Trouver la pomme qui est
différente

Trouver celle qui n'a pas beaucoup d'autres
pommés similaires

Données



```
>>> from sklearn.ensemble import IsolationForest
>>> iforest = IsolationForest(contamination=0.05)
>>> iforest = iforest.fit(X)
```



permet de distinguer les:



Données normales



Anomalies

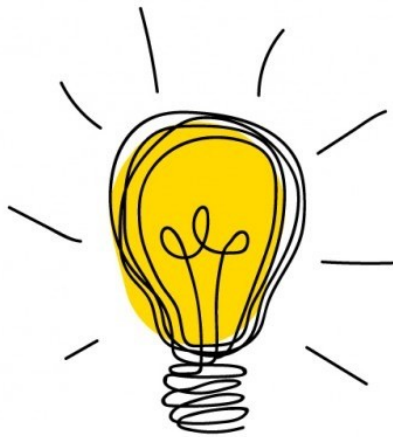
Applications:



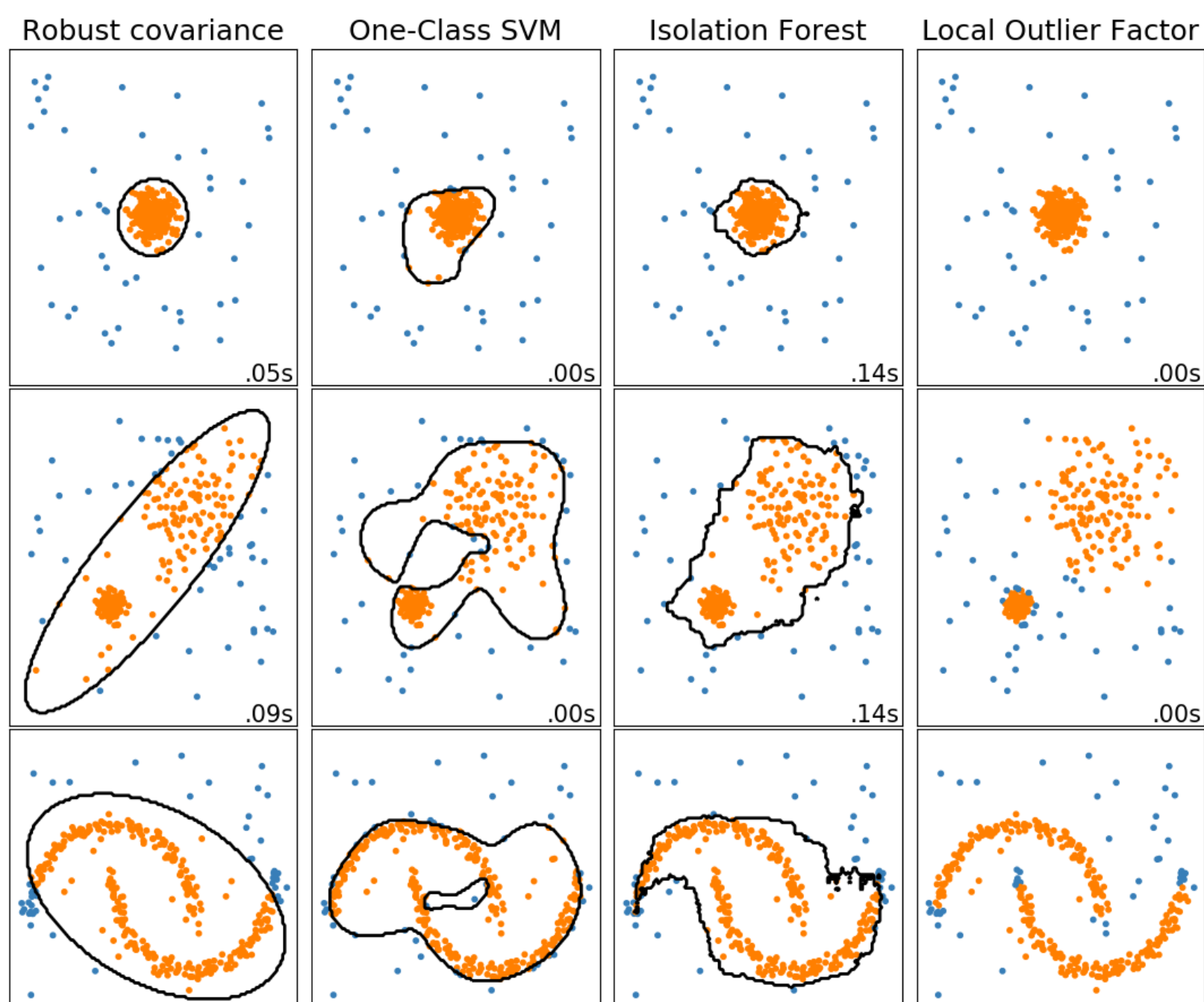
Electrocardiogram (ECG)



Suivi de processus industriel, détection de fraudes, maintenance pré-



Trouver les
événements
statistiquement
rares



Les évolutions de
scikit-learn:

- OCSVM v0.1
- Robust Cov. v0.10
- Isolation Forest v0.18
- Local Outlier Factor (LOF)

v0.19

Quelle suite?



Previous 2.6. Covarian...	Next 2.8. Density ...	Up 2. Unsupervis...
----------------------------------------------	------------------------------------------	----------------------------------------

scikit-learn v0.19.2
[Other versions](#)

Please **cite us** if you use the software.

2.7. Novelty and Outlier Detection

2.7.1. Novelty Detection

2.7.2. Outlier Detection

- 2.7.2.1. Fitting an elliptic envelope
- 2.7.2.2. Isolation Forest
- 2.7.2.3. Local Outlier Factor
- 2.7.2.4. One-class SVM versus Elliptic Envelope versus Isolation Forest versus LOF

2.7. Novelty and Outlier Detection

Many applications require being able to decide whether a new observation belongs to the same distribution as existing observations (it is an inlier), or should be considered as different (it is an outlier). Often, this ability is used to clean real data sets. Two important distinction must be made:

novelty detection:

The training data is not polluted by outliers, and we are interested in detecting anomalies in new observations.

outlier detection:

The training data contains outliers, and we need to fit the central mode of the training data, ignoring the deviant observations.

The scikit-learn project provides a set of machine learning tools that can be used both for novelty or outliers detection. This strategy is implemented with objects learning in an unsupervised way from the data:

```
estimator.fit(X_train)
```

new observations can then be sorted as inliers or outliers with a predict method:

```
estimator.predict(X_test)
```

Inliers are labeled 1, while outliers are labeled -1.

2.7.1. Novelty Detection