

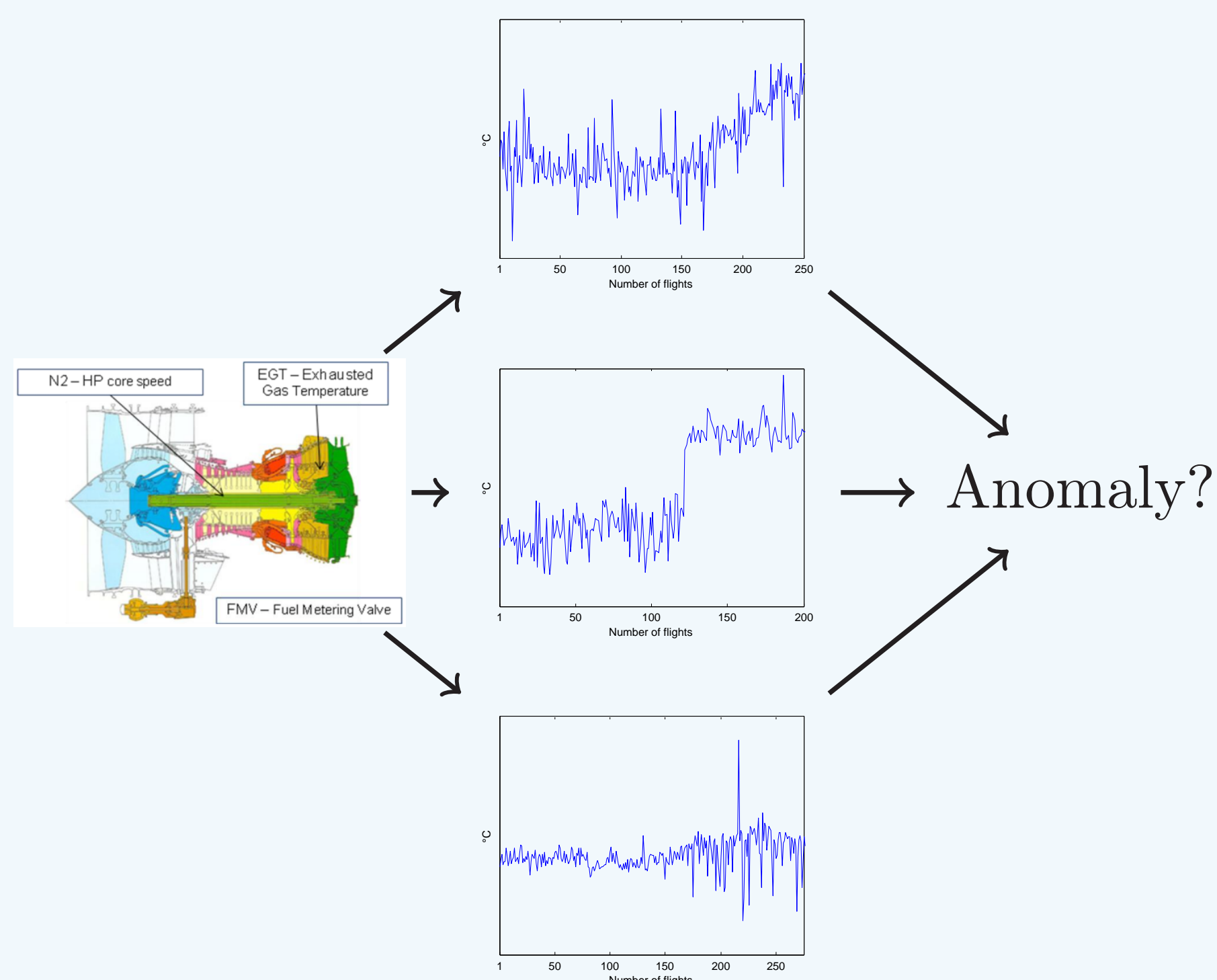
Search Strategies for Binary Feature Selection for a Naive Bayes Classifier

Tsirizo Rabenoro^{1,2}, Jérôme Lacaille¹, Marie Cottrell², and Fabrice Rossi²

1- Snecma, Groupe Safran, France

2- SAMM EA 4543, Université Paris 1 Panthéon-Sorbonne, France

Engine Monitoring



Industrial Constraints

We aim at providing **semi-automated interpretable** anomaly detection:

- simple signal processing techniques: low level anomaly detection based on statistical tests;
- simple classifiers (tree, linear model, naive bayes classifier);
- auditable models: why is that signal considered as non normal?

Very limited availability of:

1. labelled data;
2. signals with anomalies.

Industrial Strengths

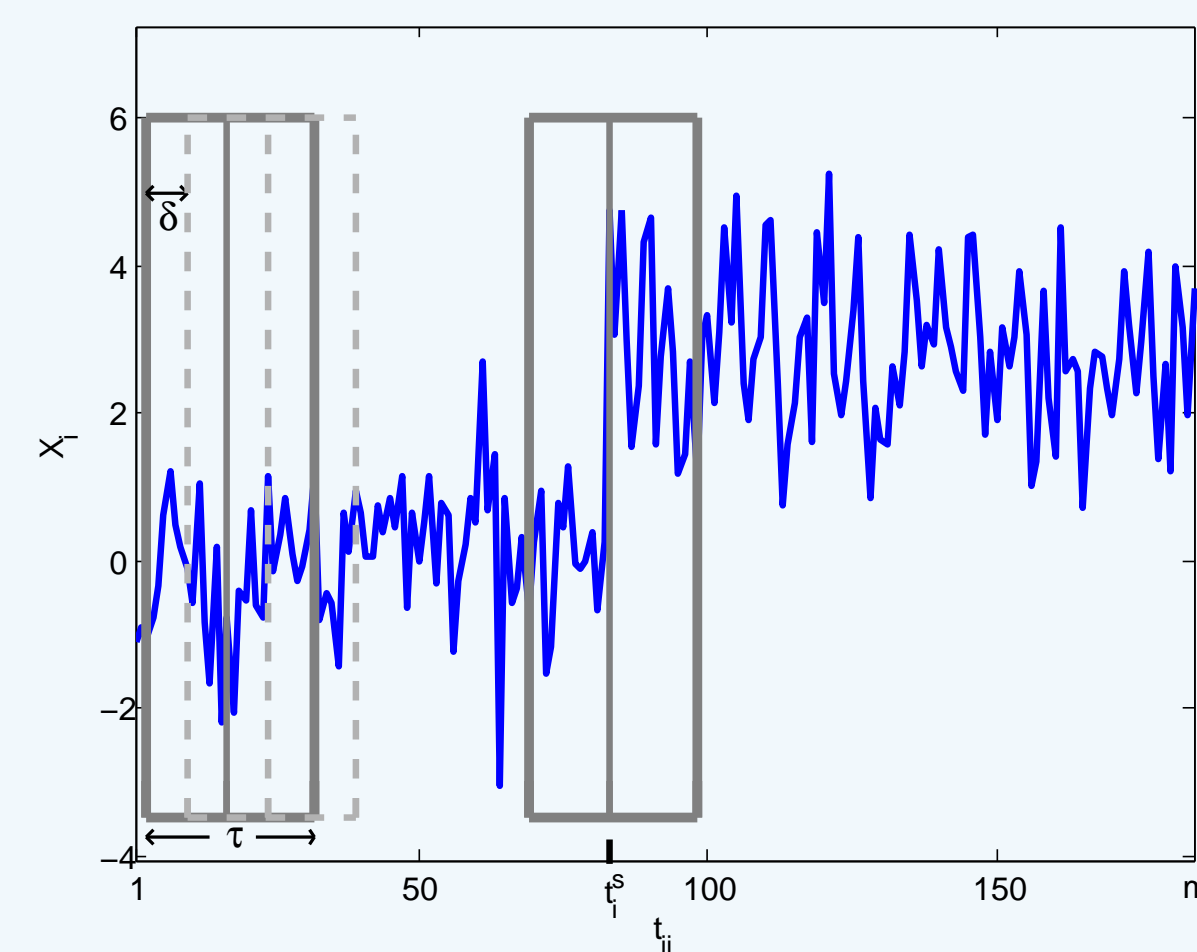
A large body of expertise:

1. signals are “normalized”: deviations from the stationary case are in most cases anomalies;
2. high level summaries of complex measures have been designed: e.g. one knows that averaging a certain quantity during a certain flight phase leads to a very informative numerical value;
3. a lot of low level tests have been designed.

But a fine tuning of the procedures has to be done.

Global strategy

1. chose a set of low level detectors, a grid of parameters and a set of significance levels;
2. transform each signal into a high dimensional binary vector by:
 - (a) applying each of the low level detector to the signal with all its possible parameter settings;
 - (b) thresholding the results into binary values based on the significance levels.
3. learn a Naive Bayes Classifier on the subset of the obtained binary features.



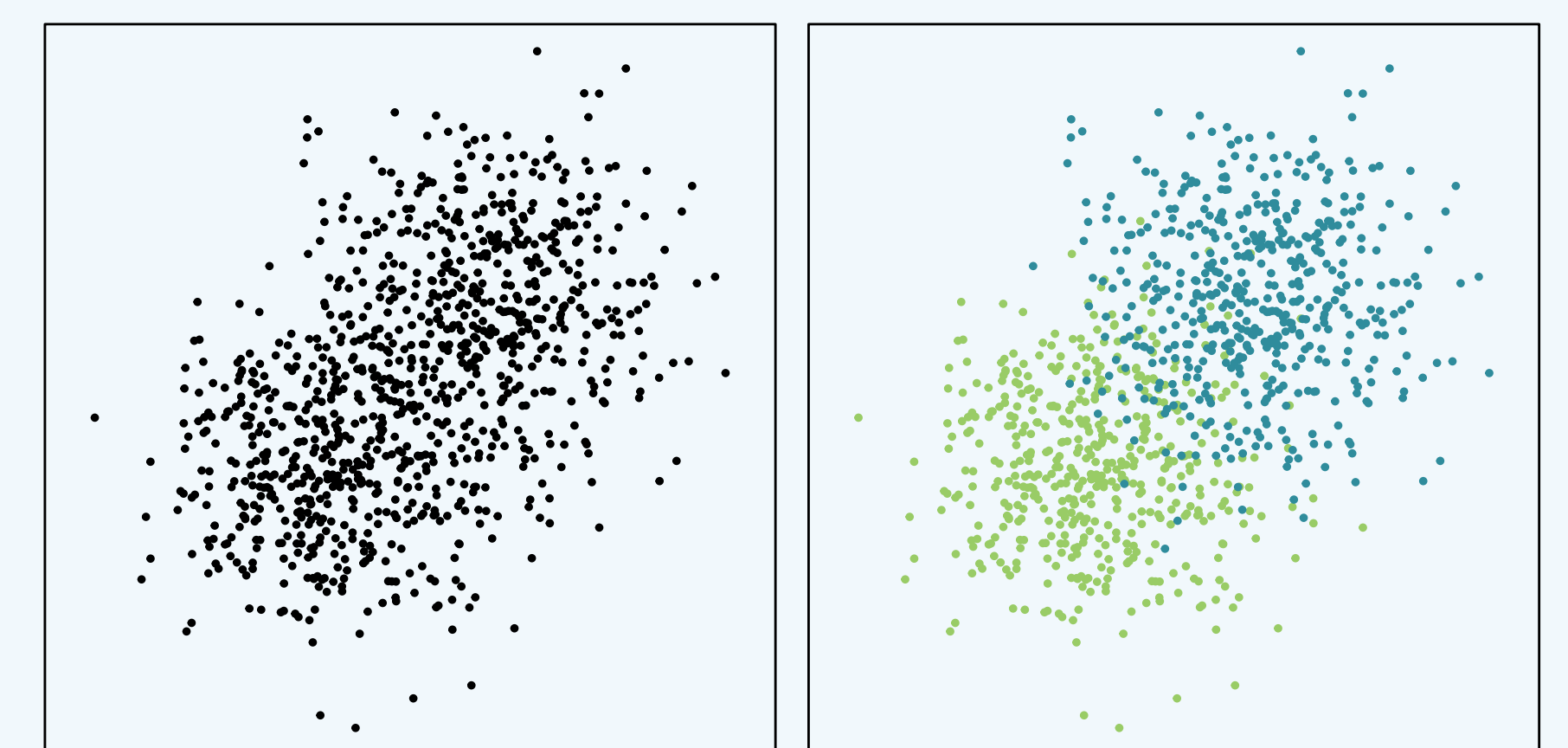
Two sample Mann-Whitney U test on δ separated τ width windows.

The Naive Bayes Classifier

The NBC is based on a conditional independence assumption

$$X_j \perp\!\!\!\perp X_k | Y.$$

This means that *features* (X_j and X_k) are independent once the class is known.



Optimal classification is easy under this model as

$$\mathbb{P}(Y = k | X) = \frac{\mathbb{P}(Y = k) \prod_{j=1}^d \mathbb{P}(X_j | Y = k)}{\mathbb{P}(X)}.$$

In particular, a d -dimensional density estimation problem $\mathbb{P}(X|Y)$ is replaced by d one-dimensional estimation problems $\mathbb{P}(X_j|Y)$.

Feature Selection for the NBC

Because of its simplicity, NBC gives satisfactory results only when used by good features.

Recommended standards:

- filter approaches (order the features according to a quality criterion that is not the NBC classification quality)
- Mutual information or slightly better: mRMR (minimum Redundancy Maximum Relevance)
- forward search

Our solution:

- wrapper approach
- “error probability” quality criterion
- forward backward search

Leverages specific aspects of the NBC.

Adding/removing features

Under the NBC independence assumption

$$\log \mathbb{P}(Y = k, X) =$$

$$\log \mathbb{P}(Y = k) + \sum_{j=1}^d \log \mathbb{P}(X_j | Y = k)$$

This allows incremental inclusion or removal of a feature at $O(1)$ cost per data point. Thus a full forward procedure for the NBC costs only $O(Nd^2)$ for N observations and d variables.

Evaluating a feature set S

Possible choices:

1. decrease in classification error
2. increase in log conditional likelihood
3. decrease in estimated classification error

$$\frac{1}{N} \sum_{i=1}^N \hat{\mathbb{P}}_S(Y \neq Y^i | X^i).$$

Experimental evaluation

We use simulated data. Two data sets with the same characteristics:

1. 3000 normal signals;
2. 3000 abnormal signals with 3 different types of anomalies;
3. sliding window based two-samples tests: we

have 814 binary indicators.

We split the first data set in two: the first half is used to estimate the probabilities needed by the NBC while the second half is used to evaluate the feature subsets during the search procedure. The second data set is used to compute performances.

Method	Evaluation	# of features	test error
MI filter	Error	422	0.1387
mRMR filter	Error	19	0.1435
Forward search	Error	136	0.1237
Forward search	Prob.	207	0.1225
Backward search	Error	27	0.1308
Backward search	Prob.	86	0.1283
Forward-Backward	Error	92	0.1238
Forward-Backward	Prob.	123	0.1237
Backward-Forward	Error	112	0.1267
Backward-Forward	Prob.	122	0.1168

Results with the log conditional likelihood are too bad to be worth reporting.