Application of SsVGMM to medical data - classification with novelty detection

Presentation · November 2017

CITATIONS
0

READS
32

4 authors, including:

Fan Yang
Nara Institute of Science and Technology
5 PUBLICATIONS **1** CITATION
SEE PROFILE

J. Soriano
8 PUBLICATIONS **5** CITATIONS
SEE PROFILE

Application of SsVGMM to Medical Data - Classification with Novelty Detection

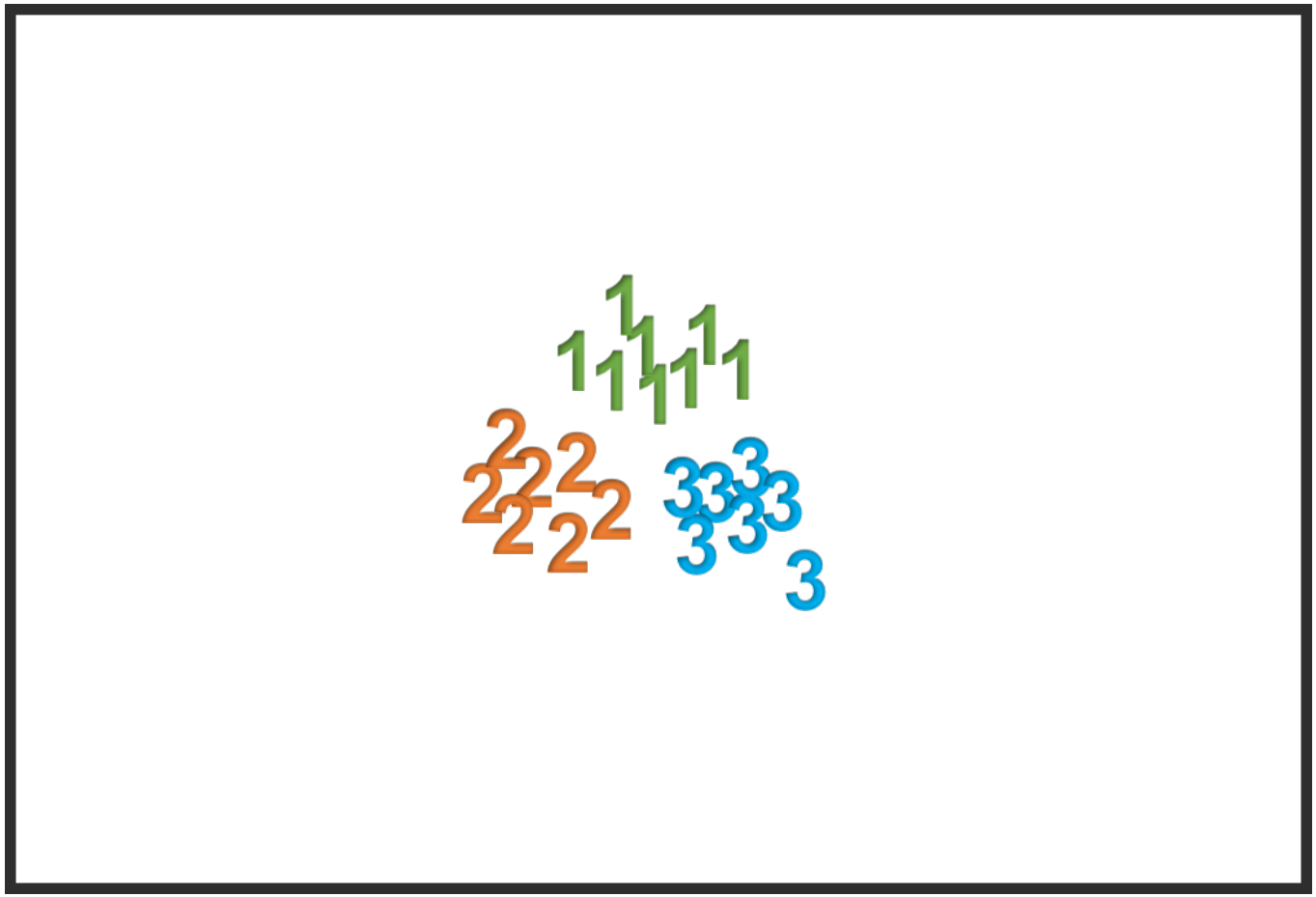
Fan Yang, Jaymar Soriano, Takatomi Kubo and Kazushi Ikeda

Nara Institute of Science and Technology, Nara, Japan

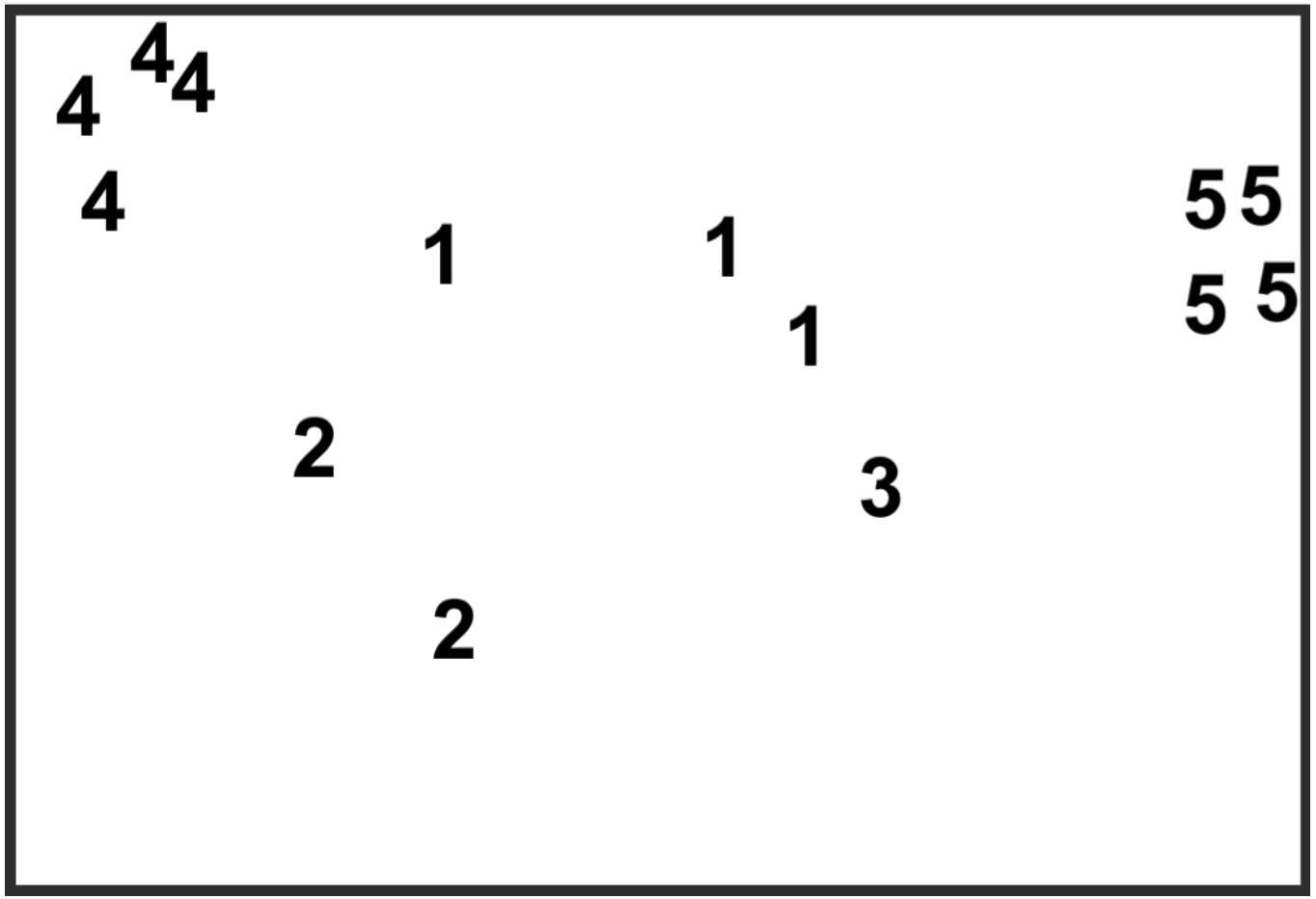
数理情報学研究室
Mathematical Informatics Lab.

What is classification with novelty detection?

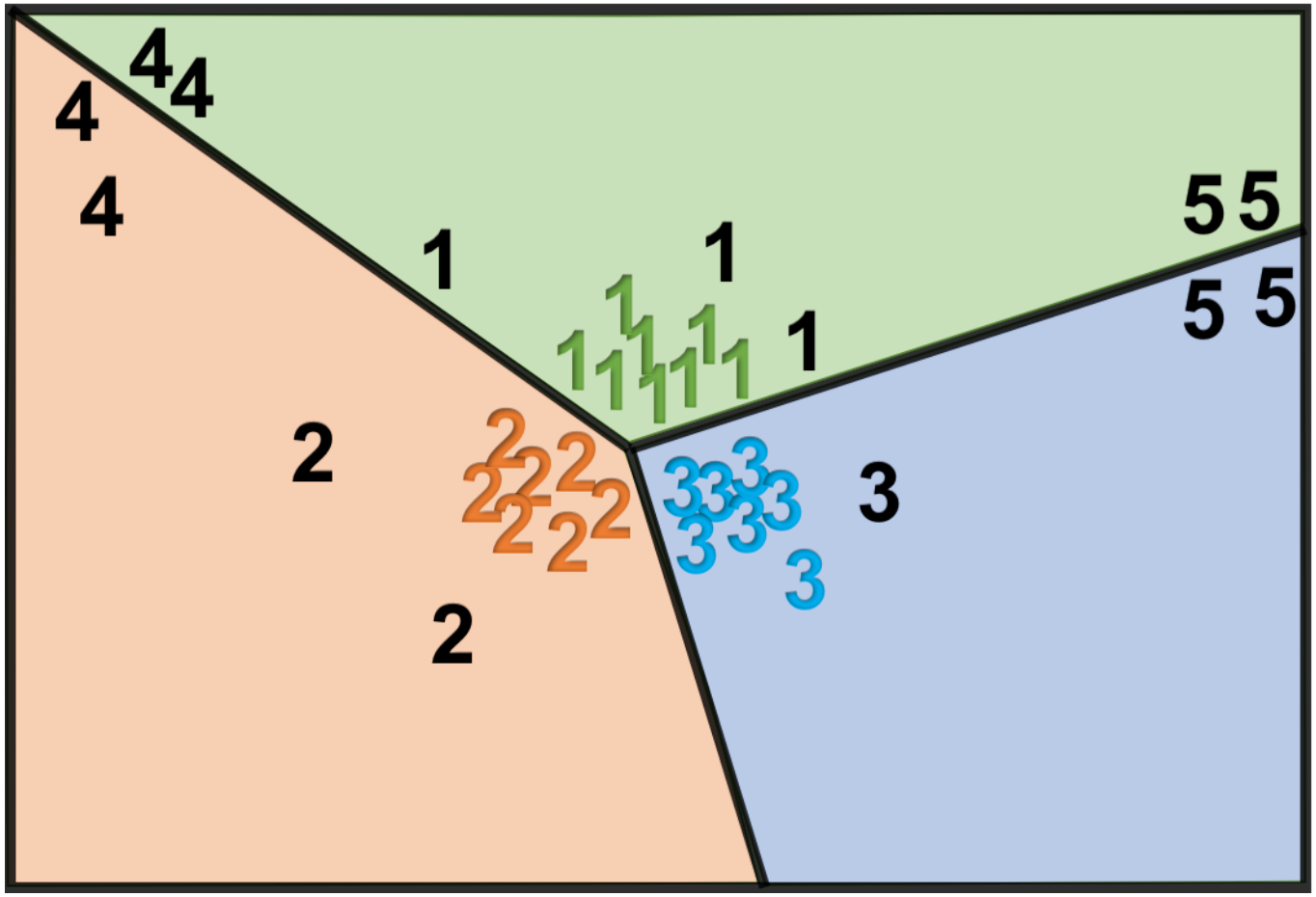
Example: we only have labeled samples from three predefined disease classes



But unlabeled samples include another two undefined (novel) disease classes

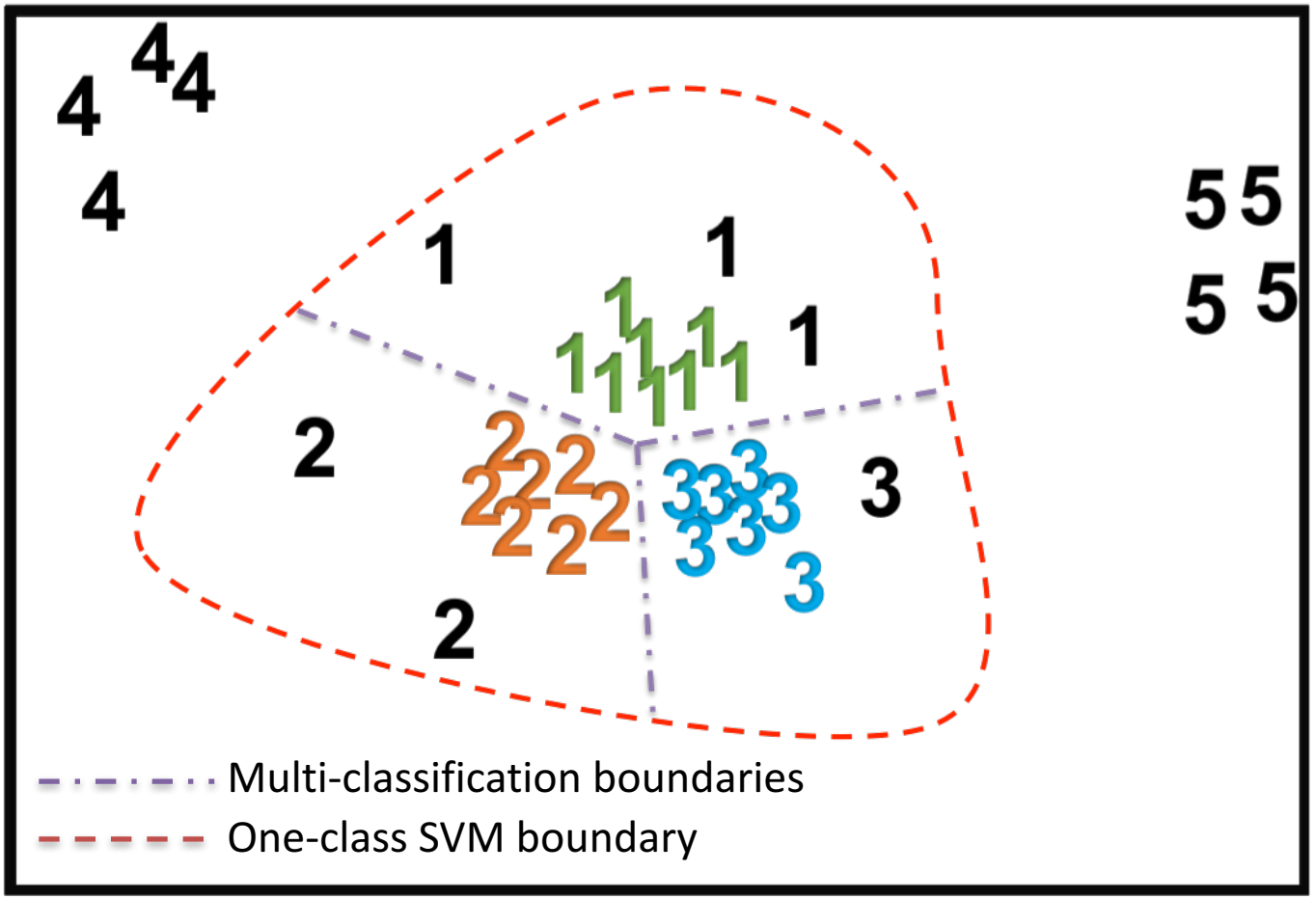


In a traditional classifier:



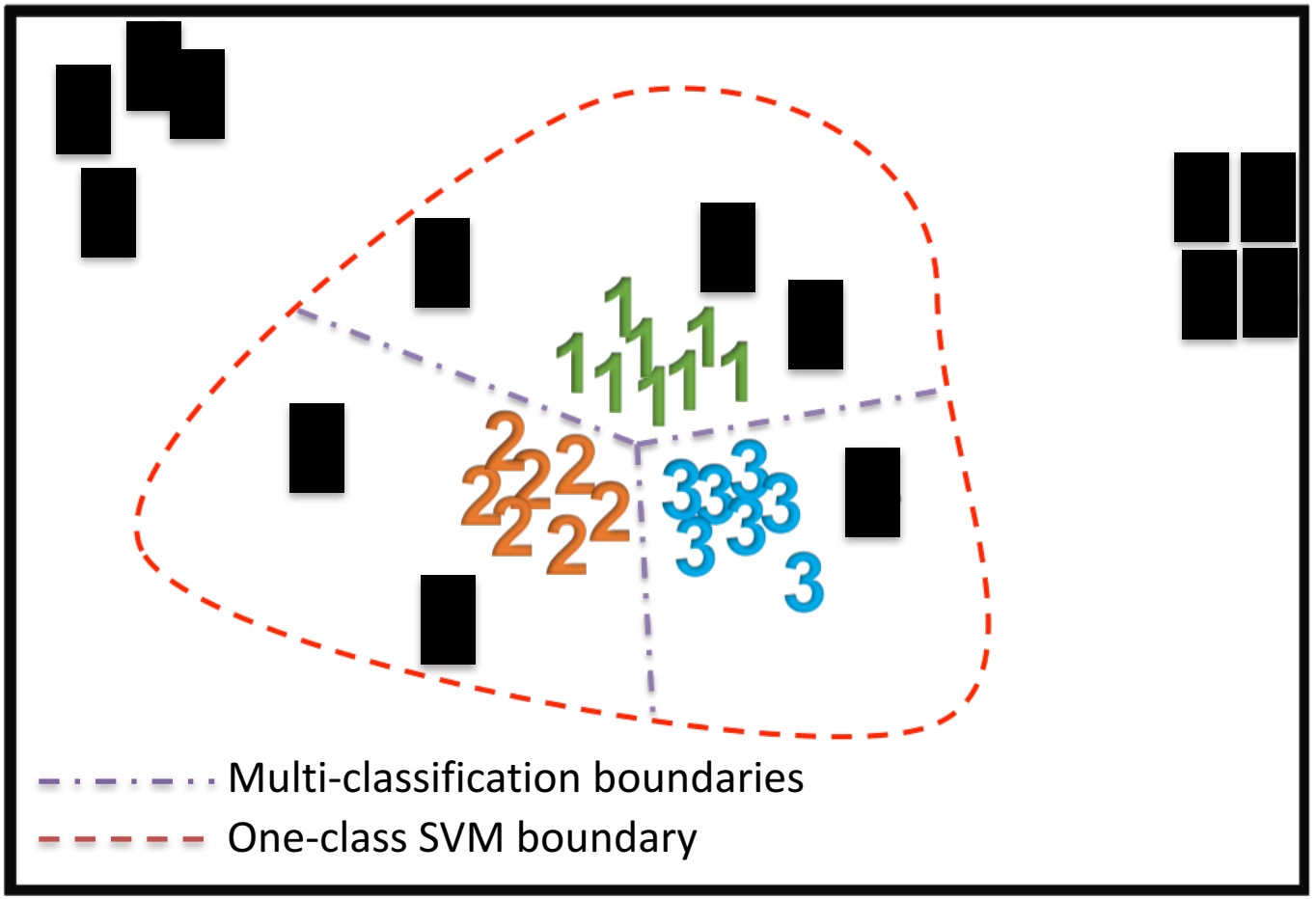
A novelty detection before the classification will help

One-class SVM & multi-classification



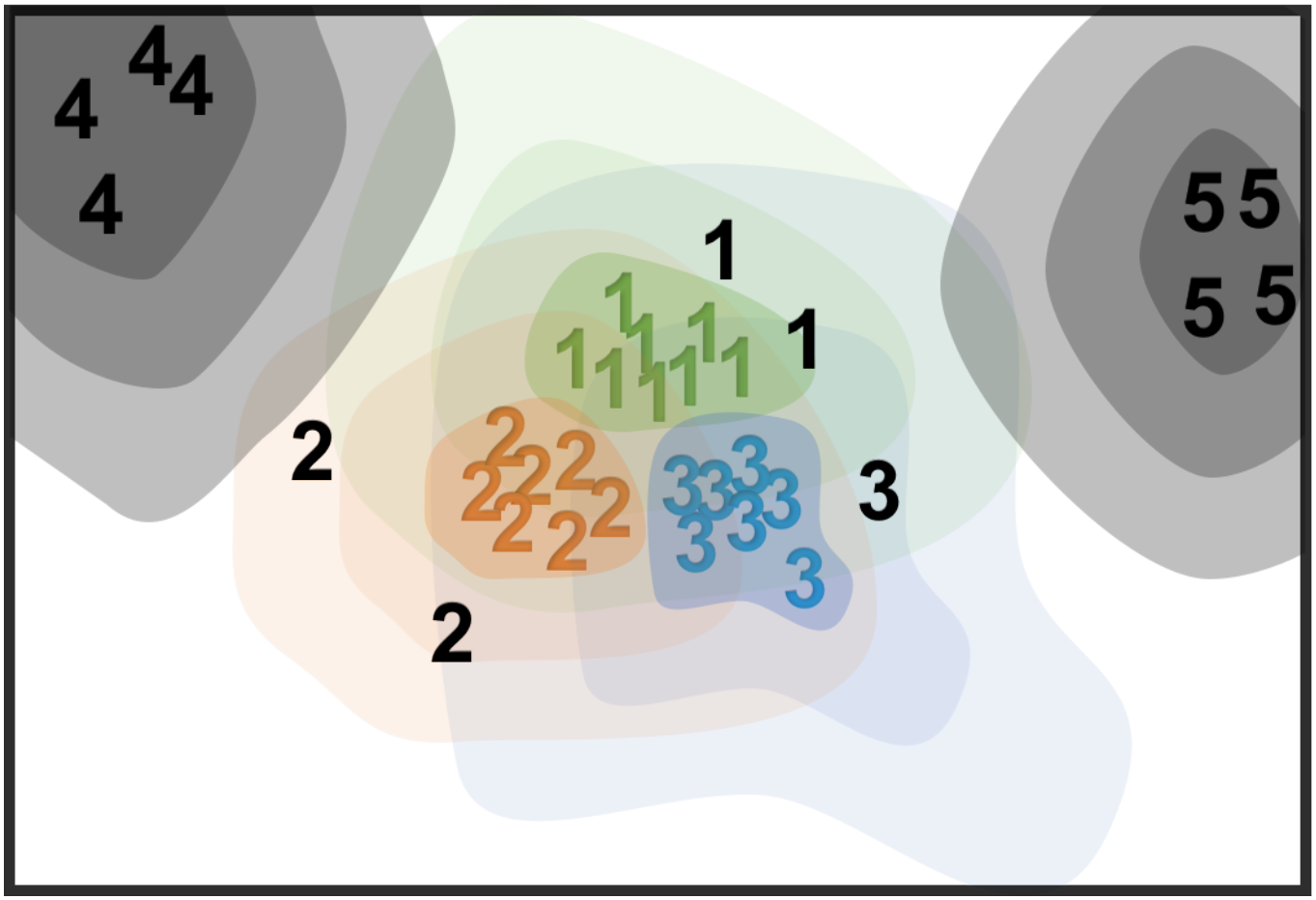
Python code: http://scikit-learn.org/stable/auto_examples/svm/plot_oneclass.html

One-class SVM & multi-classification



Python code: http://scikit-learn.org/stable/auto_examples/svm/plot_oneclass.html

Semi-supervised variational Gaussian mixture model (SsVGMM)



The code can be found at https://github.com/fandulu/SsVGMM

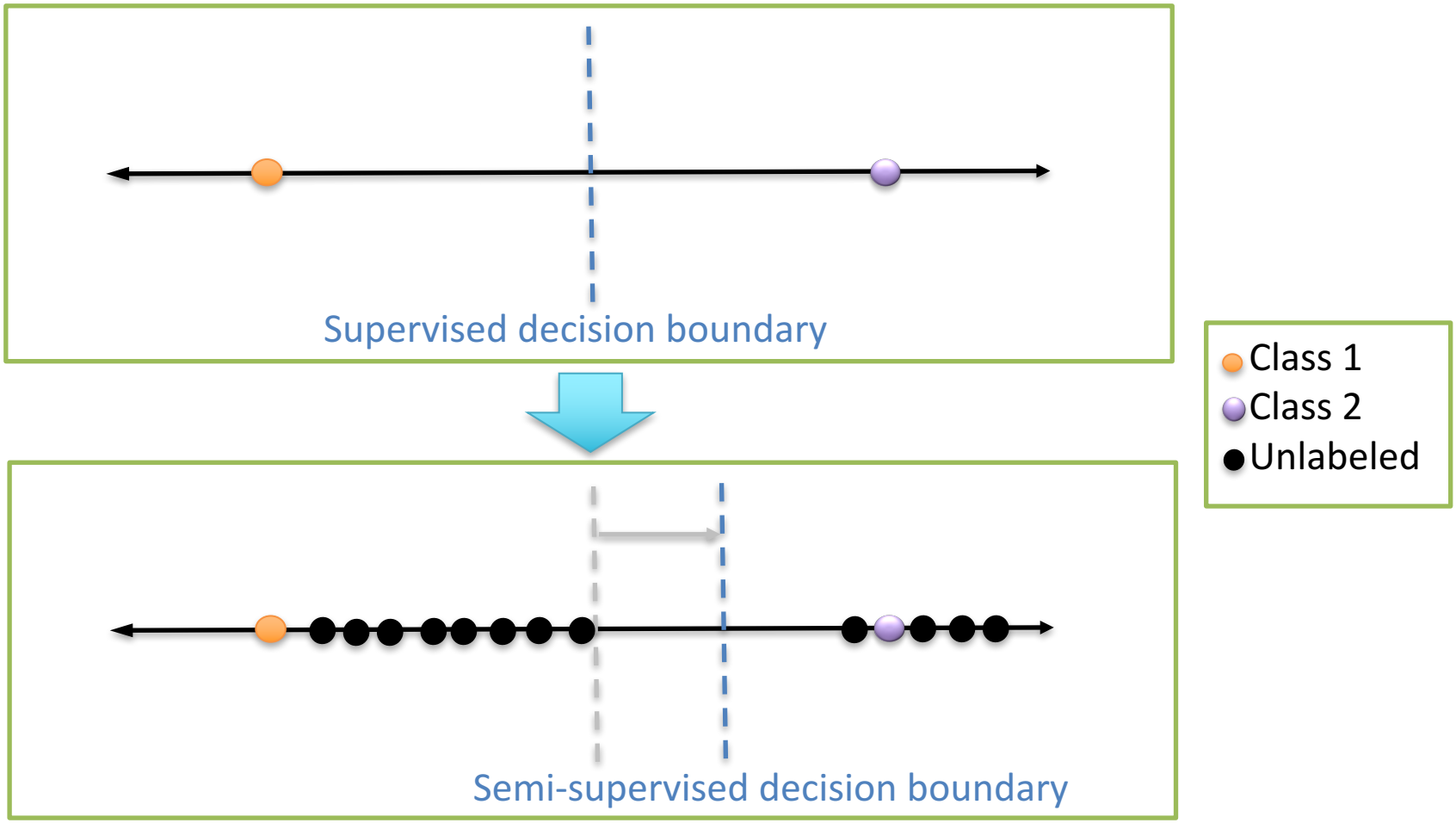
What is SsVGMM ?

Semi-supervised

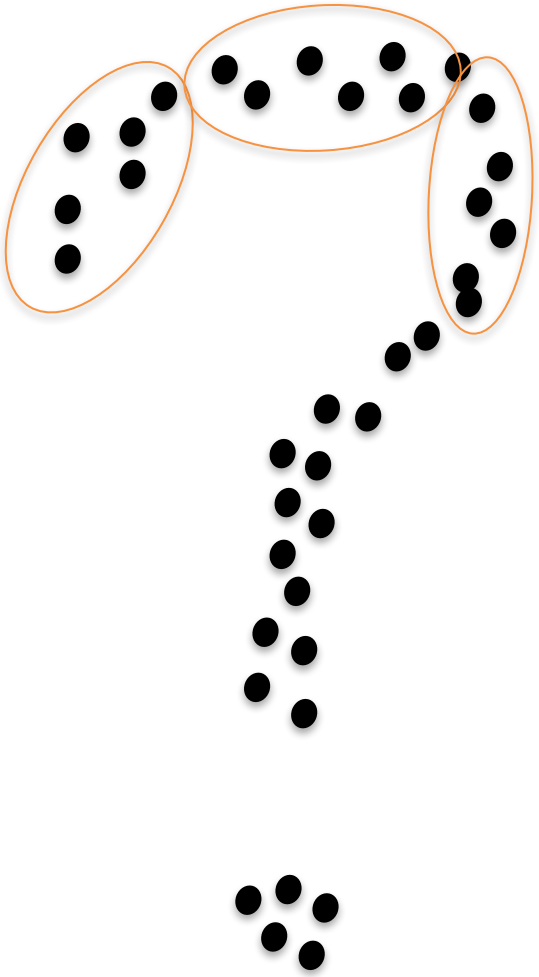
variational

Gaussian mixture model

Semi-supervised learning



Gaussian mixture model

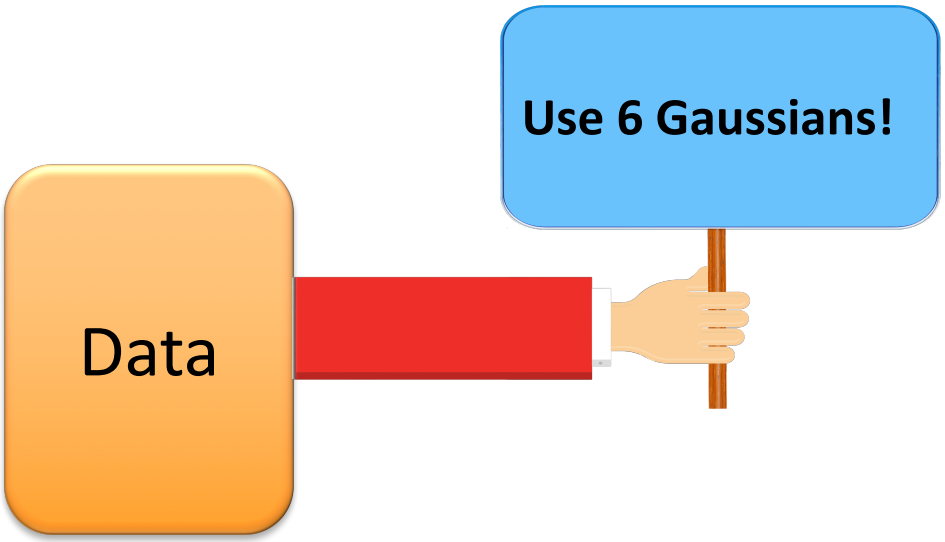


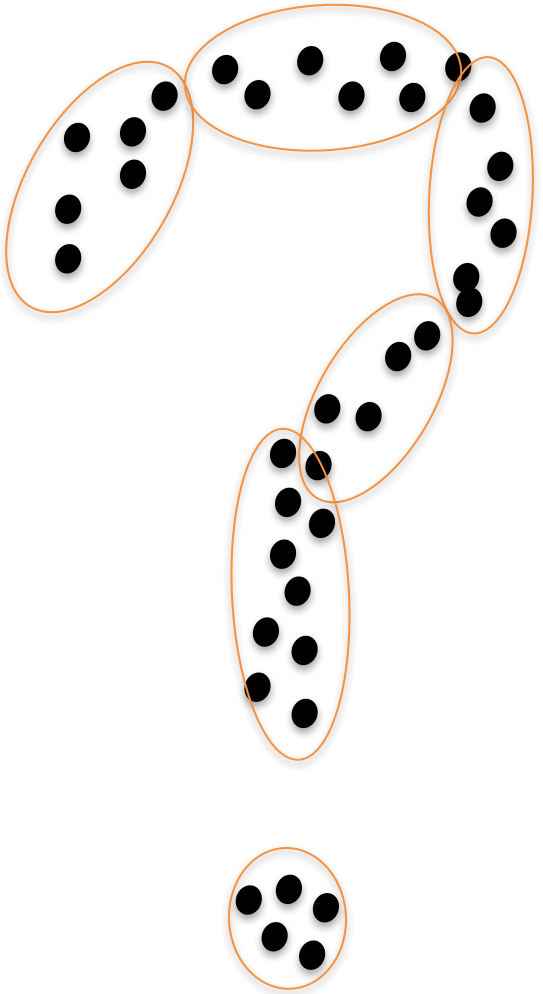
Use of multiple Gaussians to approximate the data distribution

Variational Gaussian mixture model

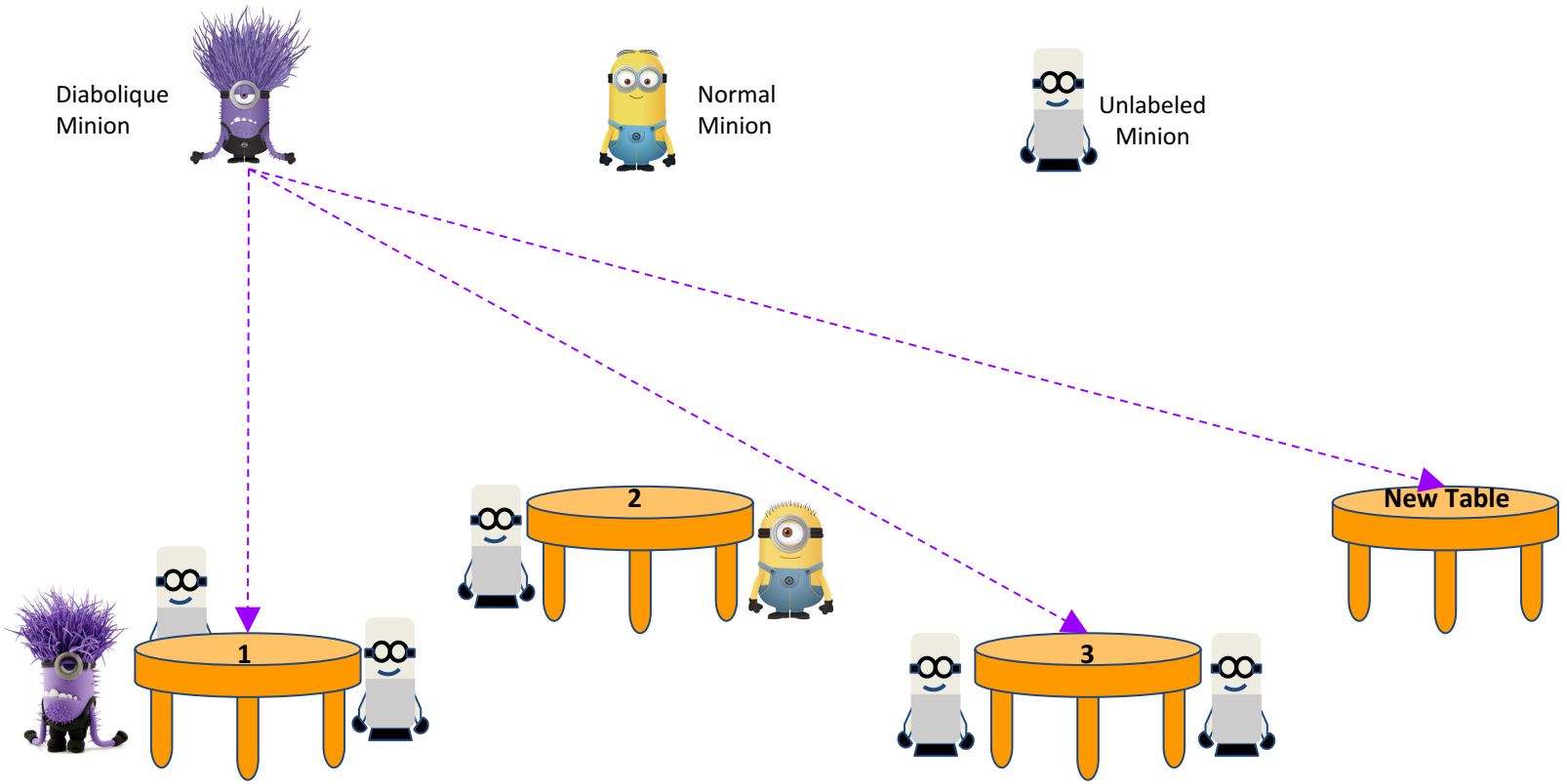
How many Gaussian components are to be used?

Let the data speak for itself.

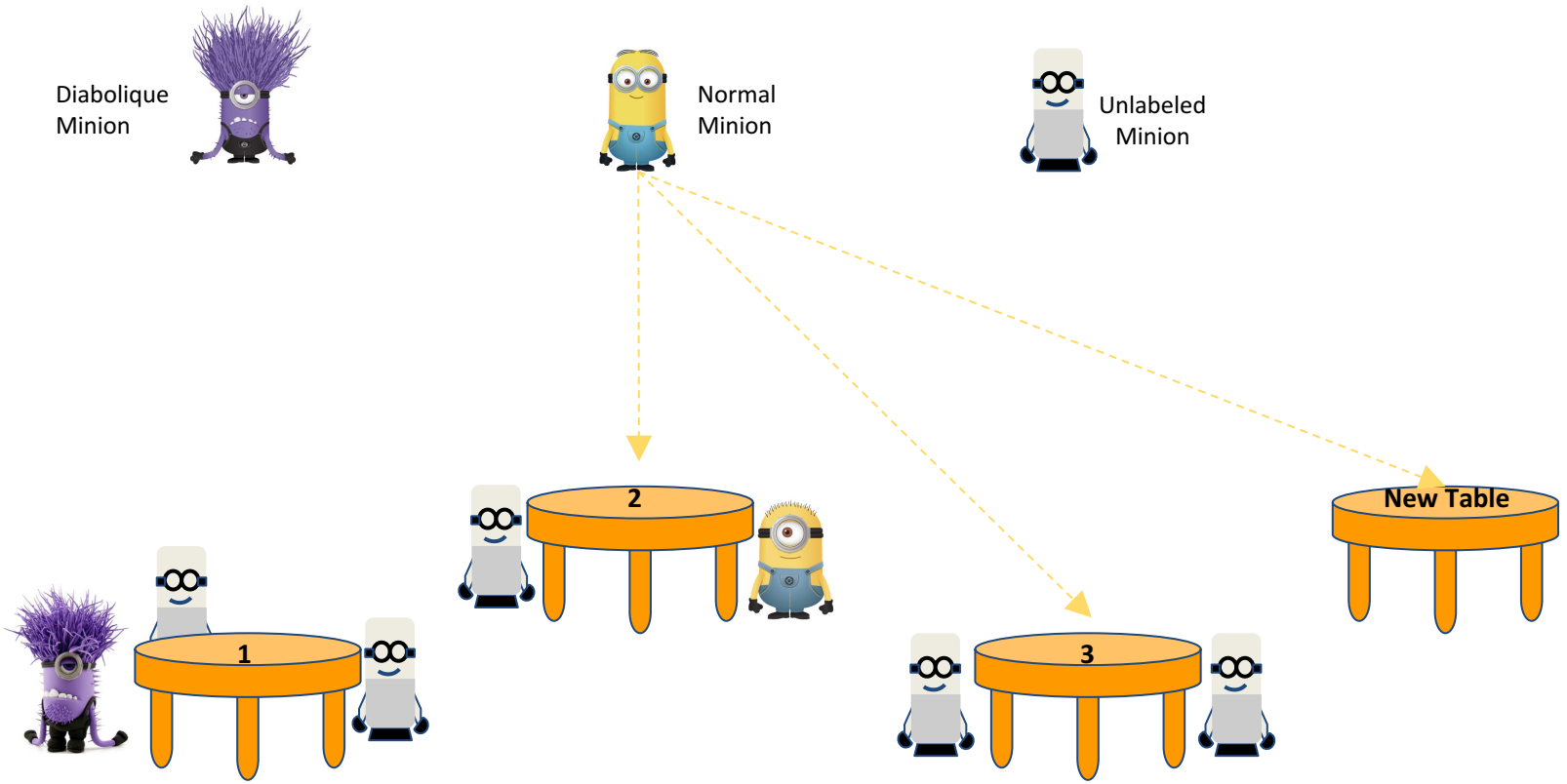




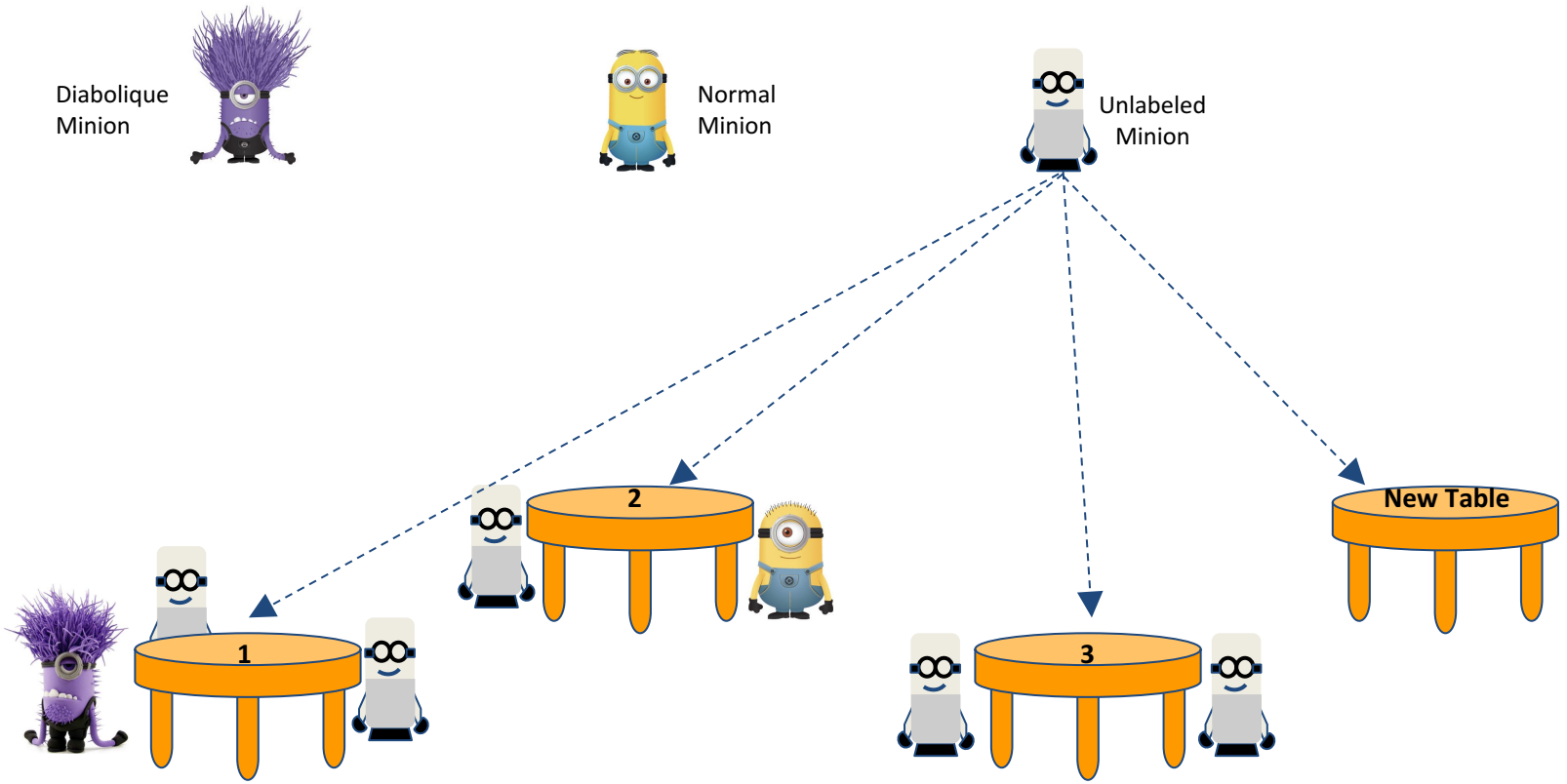
How SsVGMM works ?



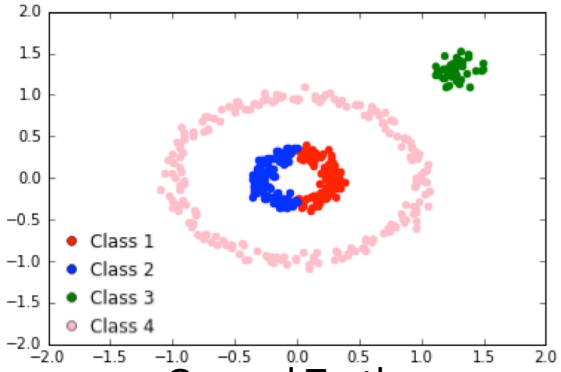
How SsVGMM works ?



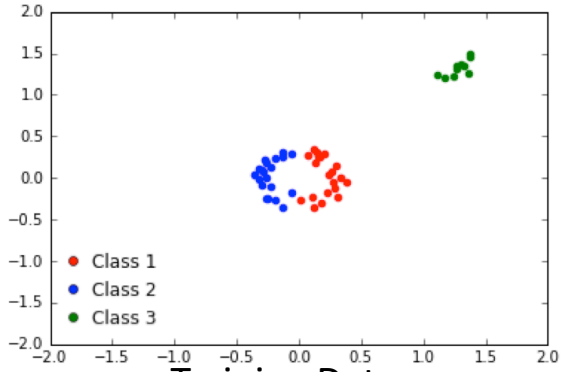
How SsVGMM works ?



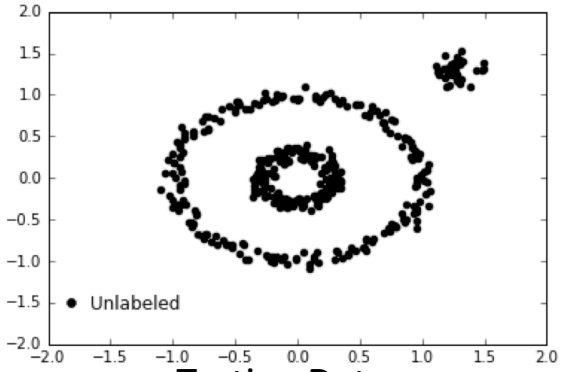
SsVGMM demonstration with synthetic data



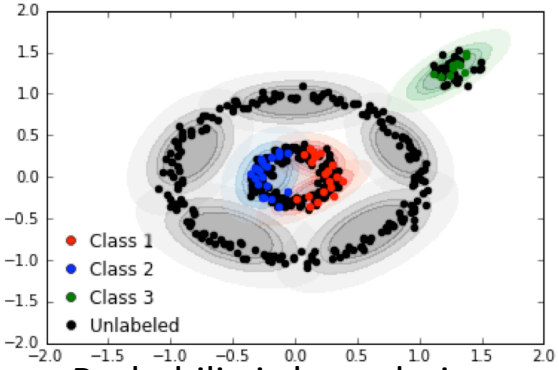
Ground Truth



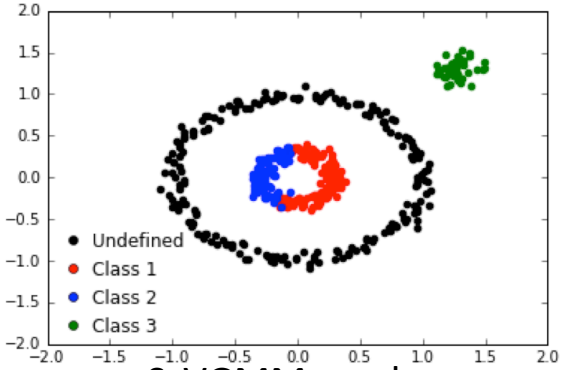
Training Data



Testing Data

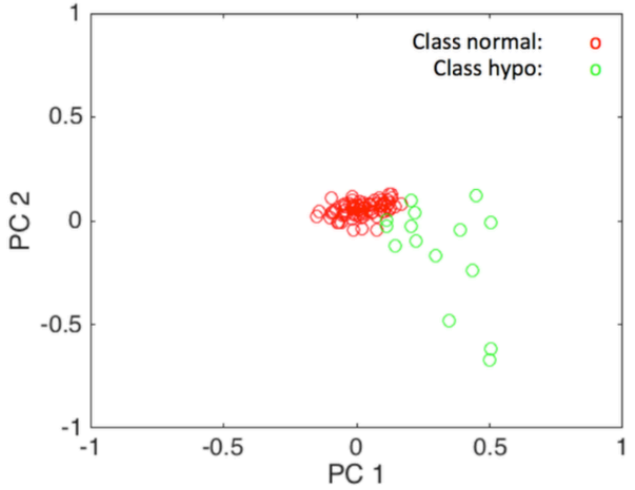


Probabilistic boundaries

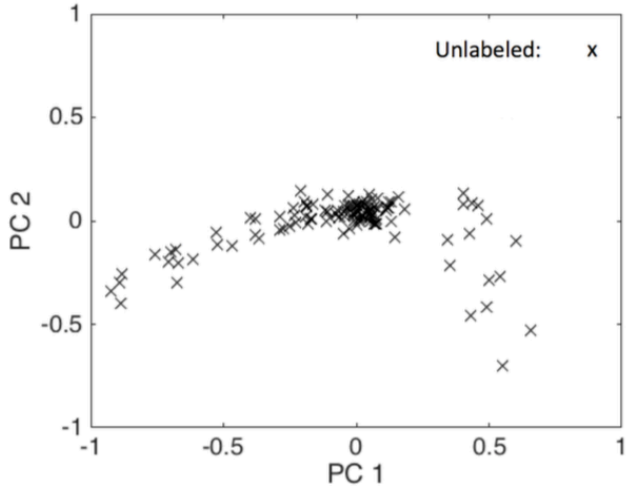


SsVGMM result

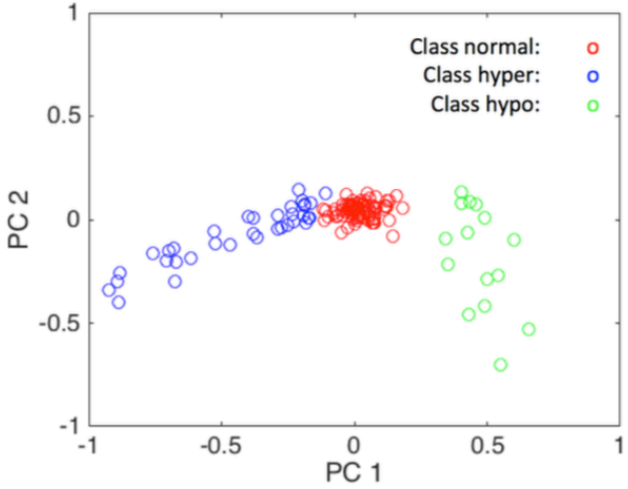
Classification with novelty detection on a thyroid disease data (1)



(a) Labeled data in the training set

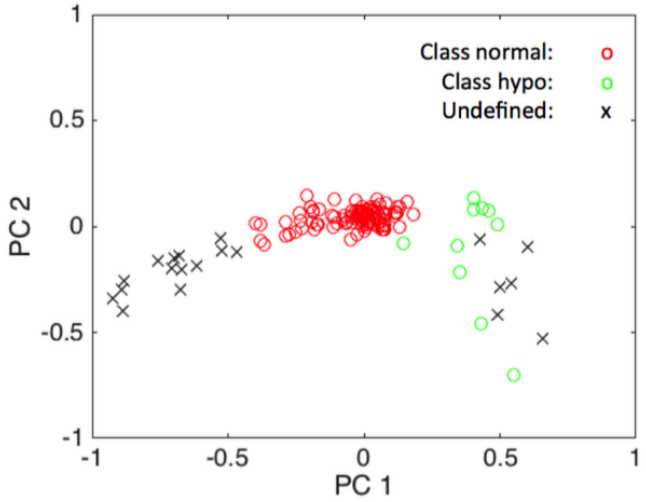


(b) Unlabelled data in the testing set

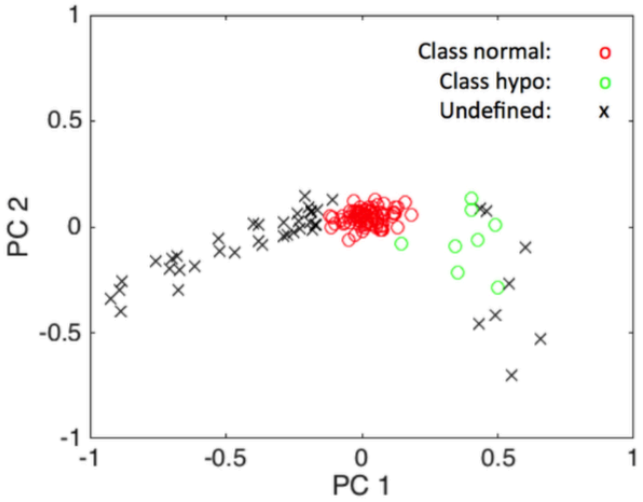


(c) True label of the testing set

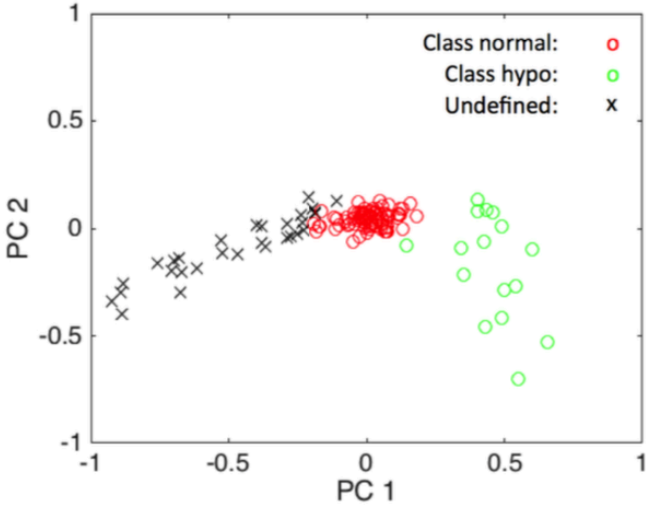
Classification with novelty detection on a thyroid disease data (2)



(d) Fine-tuned KNFST

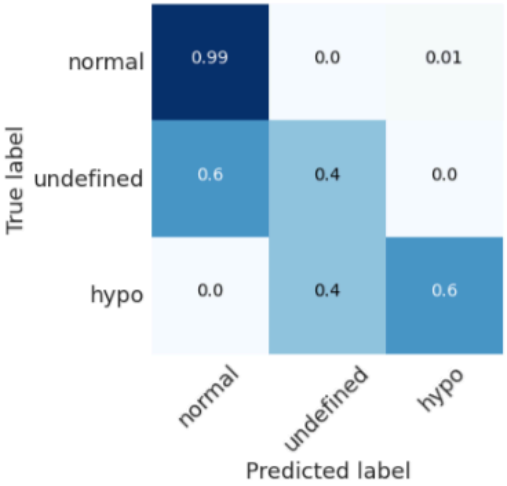


(e) Fine-tuned OSVM

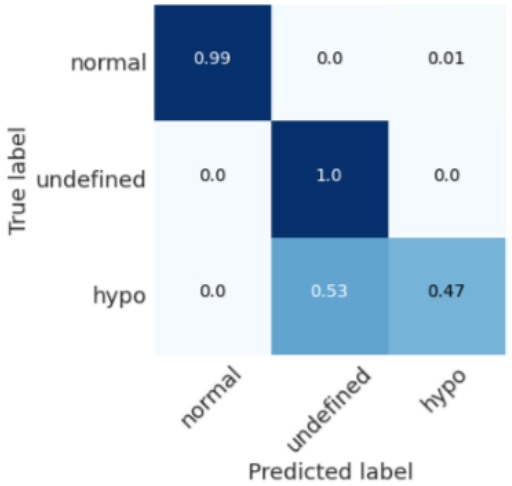


(f) SsVGMM

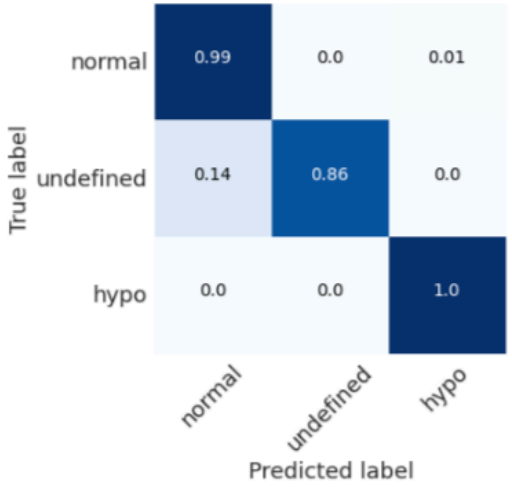
Classification with novelty detection on a thyroid disease data (3)



(g) Confusion matrix from KNFST



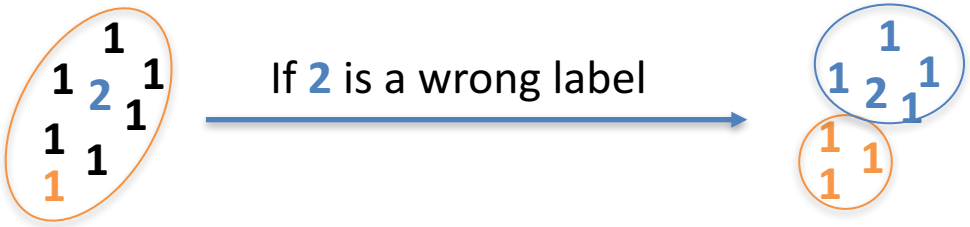
(h) Confusion matrix from OSVM



(i) Confusion matrix from SsVGMM

Limitations of SsVGMM

1.We have to assume that all given labels are correct.



Future work:
making the model to be more robust

2. SsVGMM cannot work for high dimensional data.

Possible solution:
dimensionality reduction

Summary

Aim:

To classify medical data when undefined classes exist.

Method:

Proposing SsVGMM to perform classification with undefined-class (novelty) detection.

Results:

Outperforming existing methods when predefined labels are all correct.