Detecting New Physics as Novelty

Xuhui Jiang

The Hong Kong University of Science and Technology

Based on arxiv:1807.10261 and current work
in collaboration with Aurelio Juste, Ying-Ying Li and Tao Liu

Email: xjiangaj at connect dot ust dot hk

Signal Processing 99 (2014) 215–249

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Review

A review of novelty detection

Marco A.F. Pimentel *, David A. Clifton, Lei Clifton, Lionel Tarassenko

Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford OX3 7DQ, UK

ARTICLE INFO

Article history:
Received 17 October 2012
Received in revised form
16 December 2013
Accepted 23 December 2013
Available online 2 January 2014

Keywords:
Novelty detection
One-class classification
Machine learning

ABSTRACT

Novelty detection is the task of classifying test data that differ in some respect from the data that are available during training. This may be seen as "one-class classification", in which a model is constructed to describe "normal" training data. The novelty detection approach is typically used when the quantity of available "abnormal" data is insufficient to construct explicit models for non-normal classes. Application includes inference in datasets from critical systems, where the quantity of available normal data is very large, such that "normality" may be accurately modelled. In this review we aim to provide an updated and structured investigation of novelty detection research papers that have appeared in the machine learning literature during the last decade.

To detect novel events from the known sample without any prior knowledge.

To detect NP model-independently

Novelty Detection:

- Isolation-based ($\mathcal{O}_{\rm iso}$). The novelty response for a given testing event is evaluated according to its isolation from the distribution of the known-pattern data in the feature space. All of the other testing events are irrelevant in this process.

Autoencoder loss:

to be determined by DNN trained on the known background.

The other events in the testing sample don't contribute to it.

Isolation-based:

Autoencoder-based: (arxiv: 1807.10261, 1808.08979, 1808.08992, 1811.10276, 1903.02032, 1905.10384, 2005.01598, 2007.01850, 2103.06595, 2105.07988)

Graph: (arxiv: 1912.10625)

- Clustering (density)-based ($\mathcal{O}_{\rm clu}$). The novelty response for a given testing event is evaluated according to the clustering around this point on top of the distribution of the known-pattern data in the feature space. The other testing events (especially the ones nearby) are potentially relevant in this process.

Neyman–Pearson lemma: $\dfrac{p_{data}(x)}{p_{bg}(x)}$

Z-like score: $\dfrac{N-B}{\sqrt{B}}$

Clustering-based:

kNN-based: (arxiv: 1807.10261, 1807.06038)

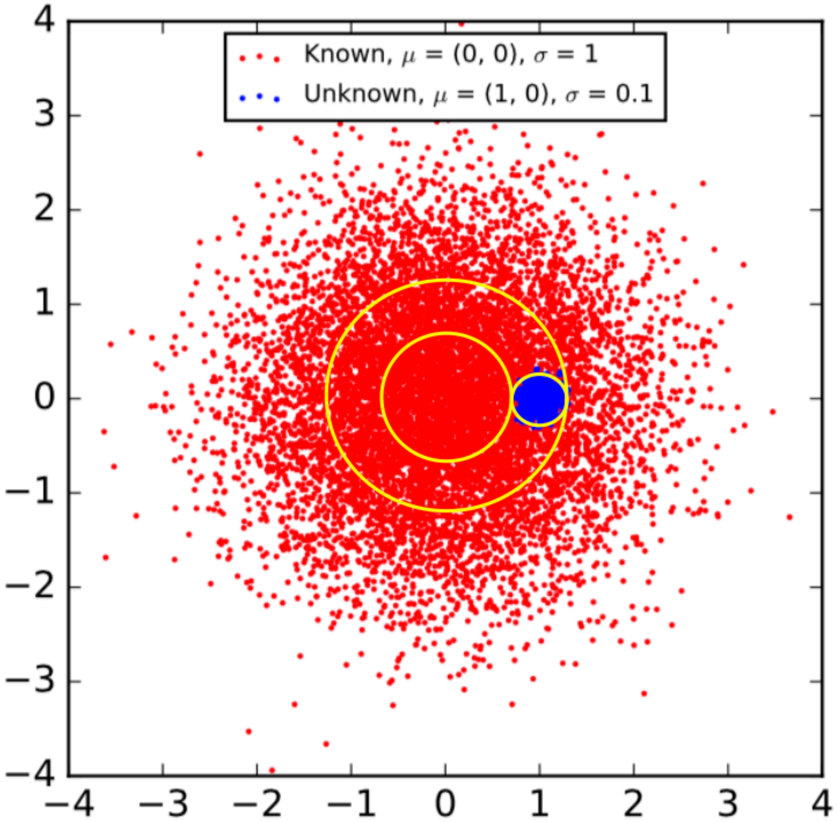
t-score (arxiv: 1806.02350, 1912.12155), ANODE (arxiv: 2001.04990), CWoLa (arxiv: 1805.02664, 1902.02634, 2005.02983), TNT (arxiv: 2002.12376), SALAD (arxiv: 2001.05001), SULU (arxiv: 2011.09863), UCluster (arxiv: 2010.07106).

Isolation-based Evaluator: k-Nearest Neighbours

arxiv:1807.10261

$$\mathcal{O}_{\text{iso}} = \frac{1}{2}\left(1 + \text{erf}\frac{\Delta_{\text{iso}}}{c\sqrt{2}}\right), \qquad \text{with} \qquad \Delta_{\text{iso}} = \frac{d_{\text{train}} - \langle d'_{\text{train}}\rangle}{\langle d'_{\text{train}}\rangle^{1/2}}$$

Novelty evaluator: in the range of [0, 1]   Novelty measure: unnormalised range



d_{train} the mean distance of a testing event to its k nearest neighbours in the training sample

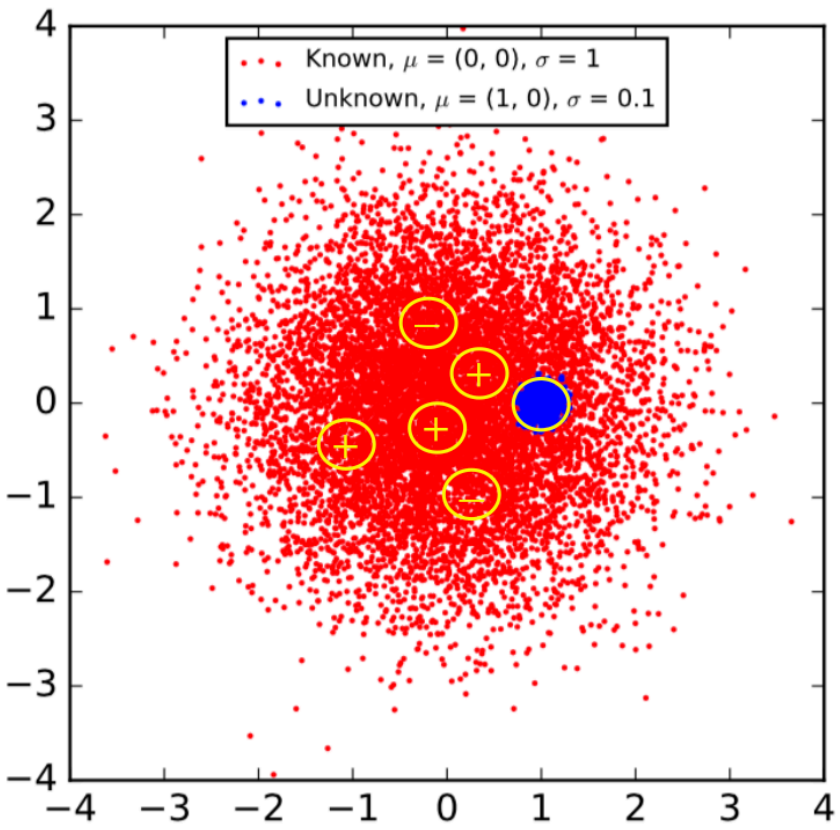
$\langle d'_{\text{train}}\rangle$ the mean distance of the training sample: average over both training sample size and k

Clustering-based Evaluator: k-Nearest Neighbours

arxiv:1807.10261

$$\mathcal{O}_{\text{clu}} = \frac{1}{2}\left(1 + \text{erf}\frac{\Delta_{\text{clu}}}{c\sqrt{2}}\right), \qquad \text{with} \qquad \Delta_{\text{clu}} = \frac{d_{\text{test}}^{-m} - d_{\text{train}}^{-m}}{d_{\text{train}}^{-m/2}}$$

Novelty evaluator: in the range of [0, 1]   Novelty measure: unnormalised range



d_{train} the mean distance of a testing event to its k nearest neighbours in the training sample

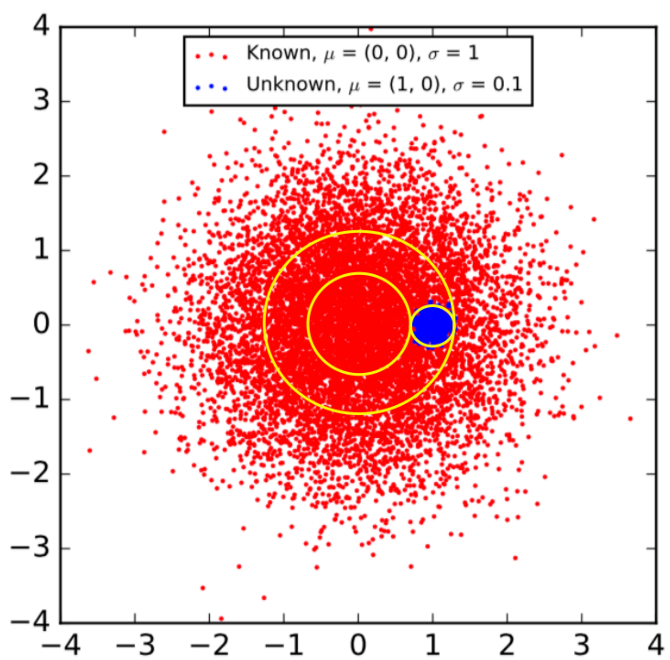
d_{test} the mean distance of a testing event to its k nearest neighbours in the testing sample

m the dimension of the feature space

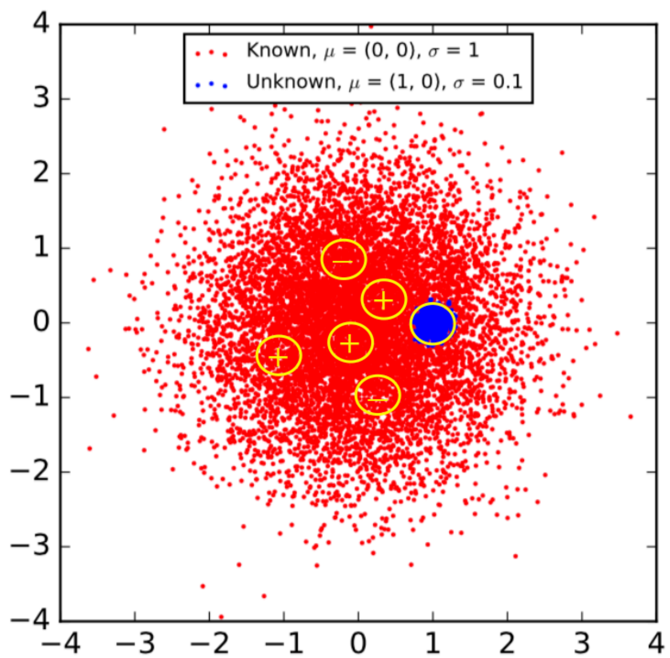
The novelty measure mimics the structure of $\frac{N-B}{\sqrt{B}}$

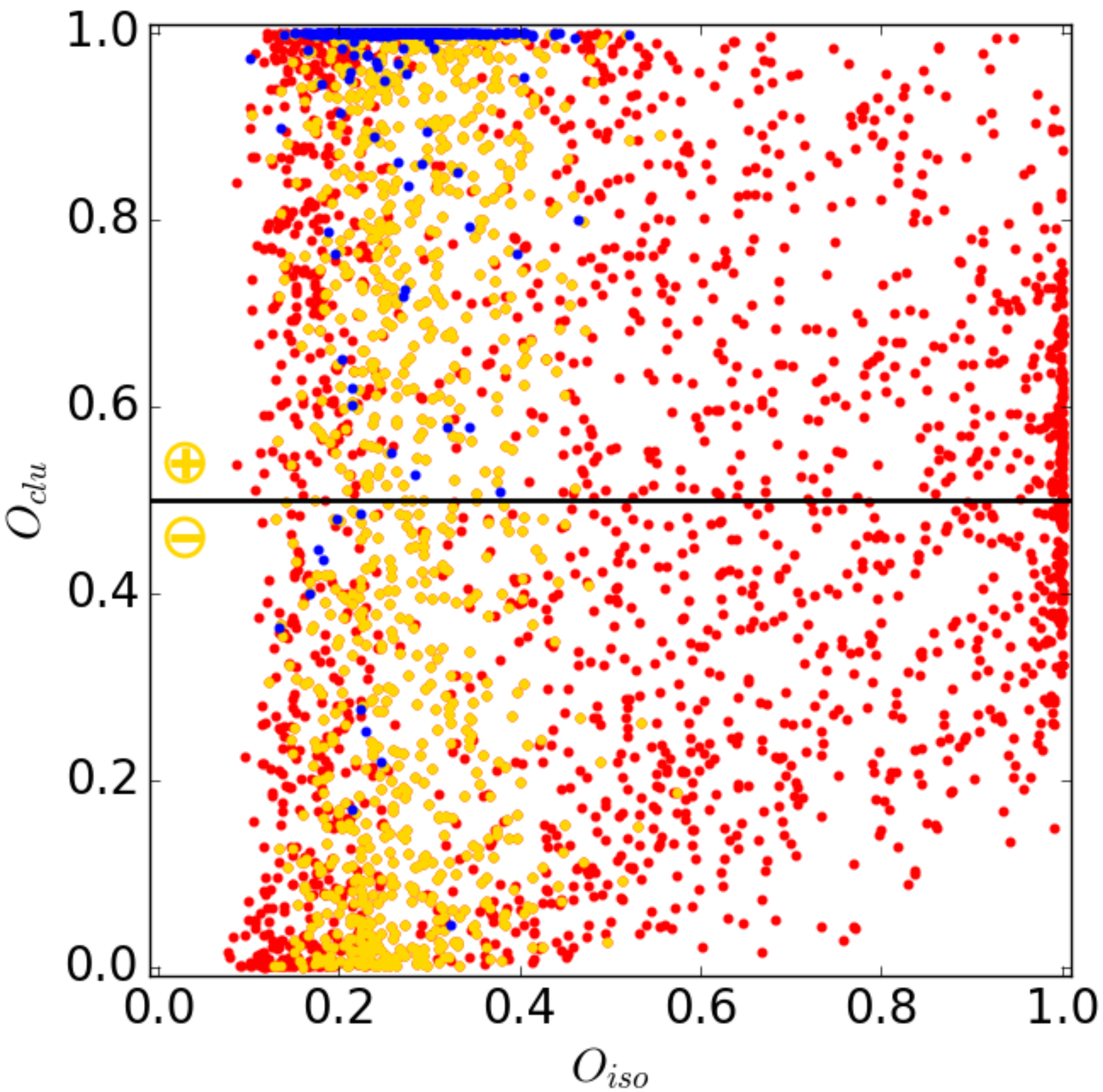
Synergy-based Evaluator

Isolation-based

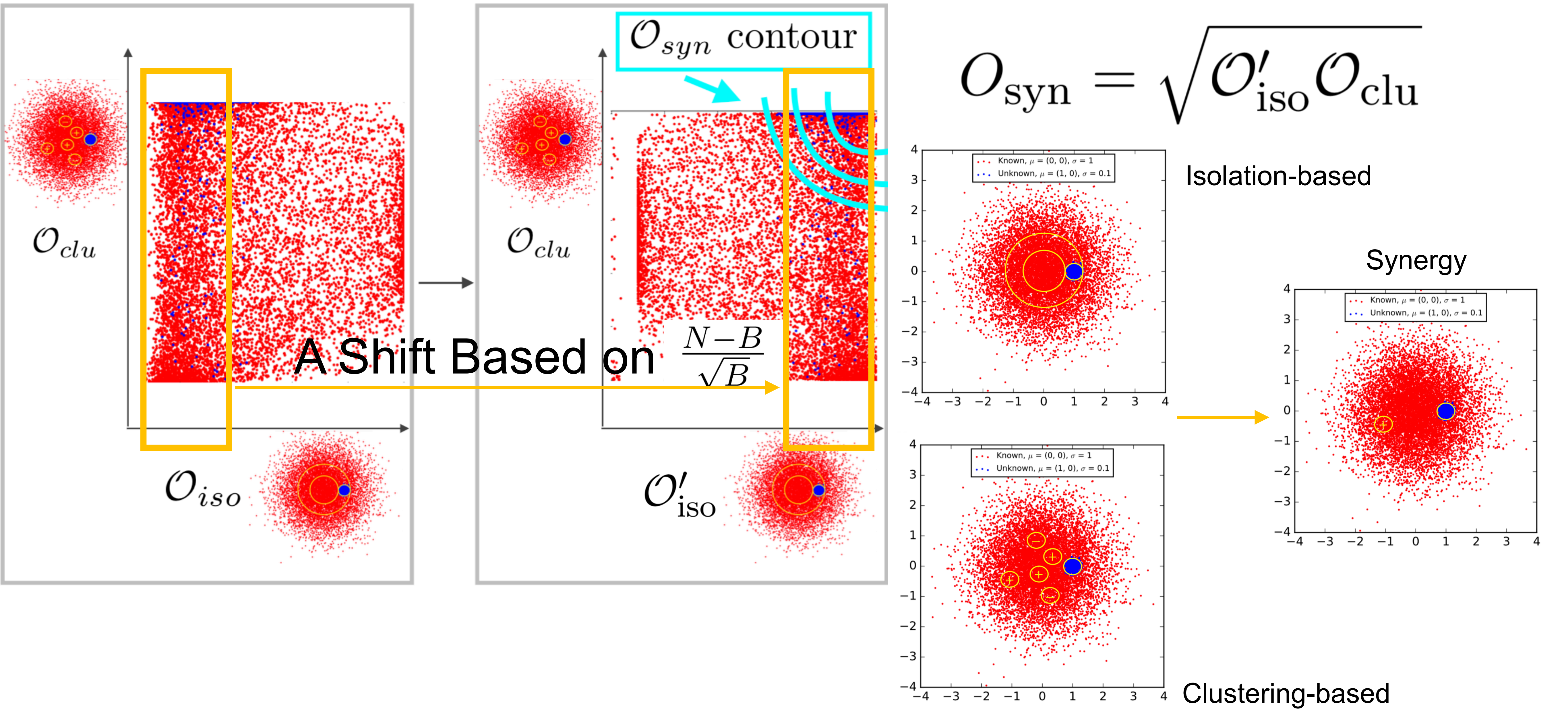


Clustering-based





Synergy-based Evaluator

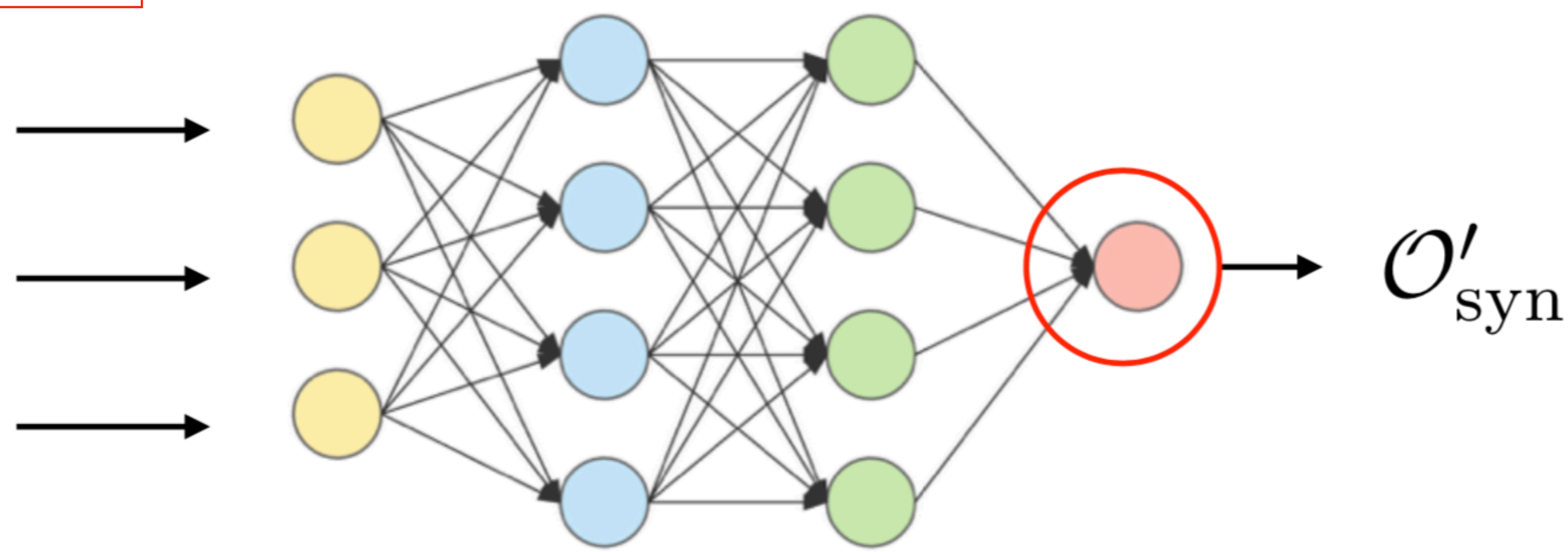


Synergy-based Evaluator

Construct a DNN and use its output as a synergy-based evaluator.

label: 0

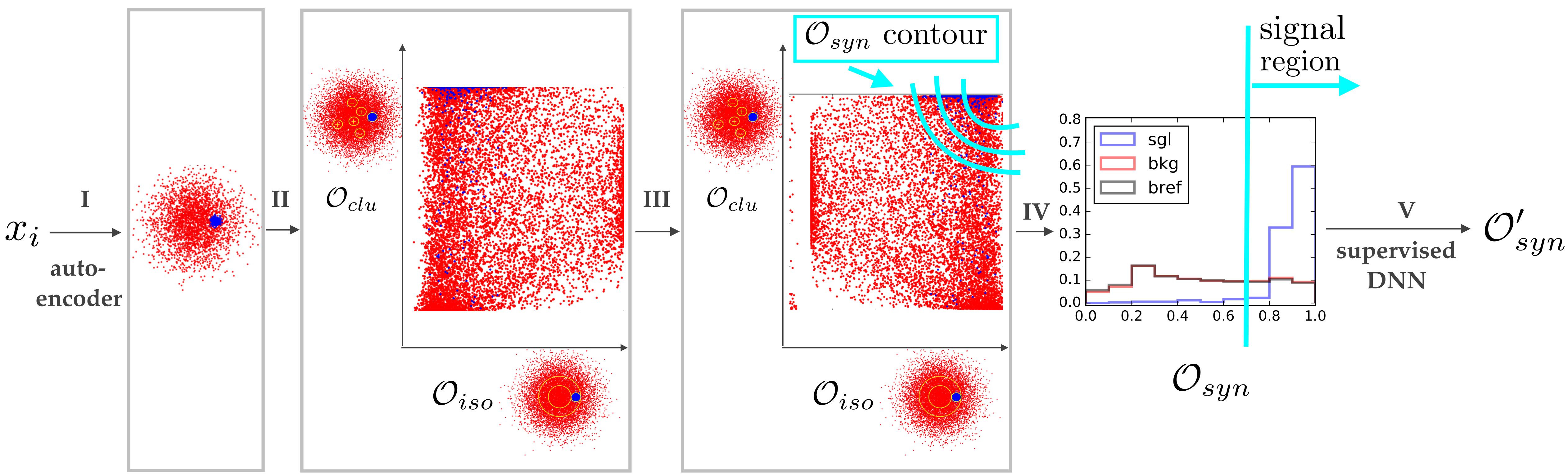
Simulated Background



$S' : O_{\mathrm{syn}} > r_0$

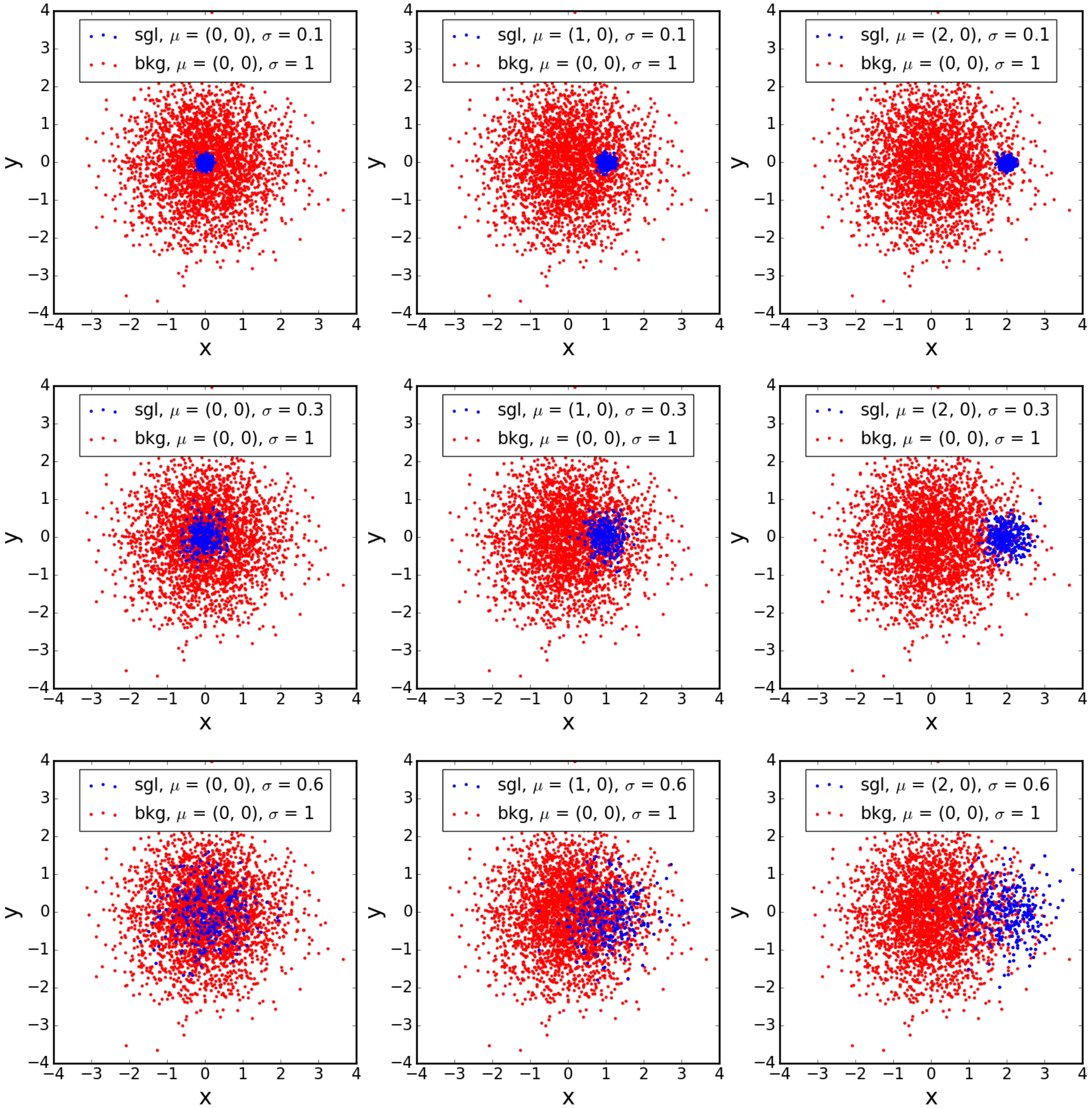
label: 1

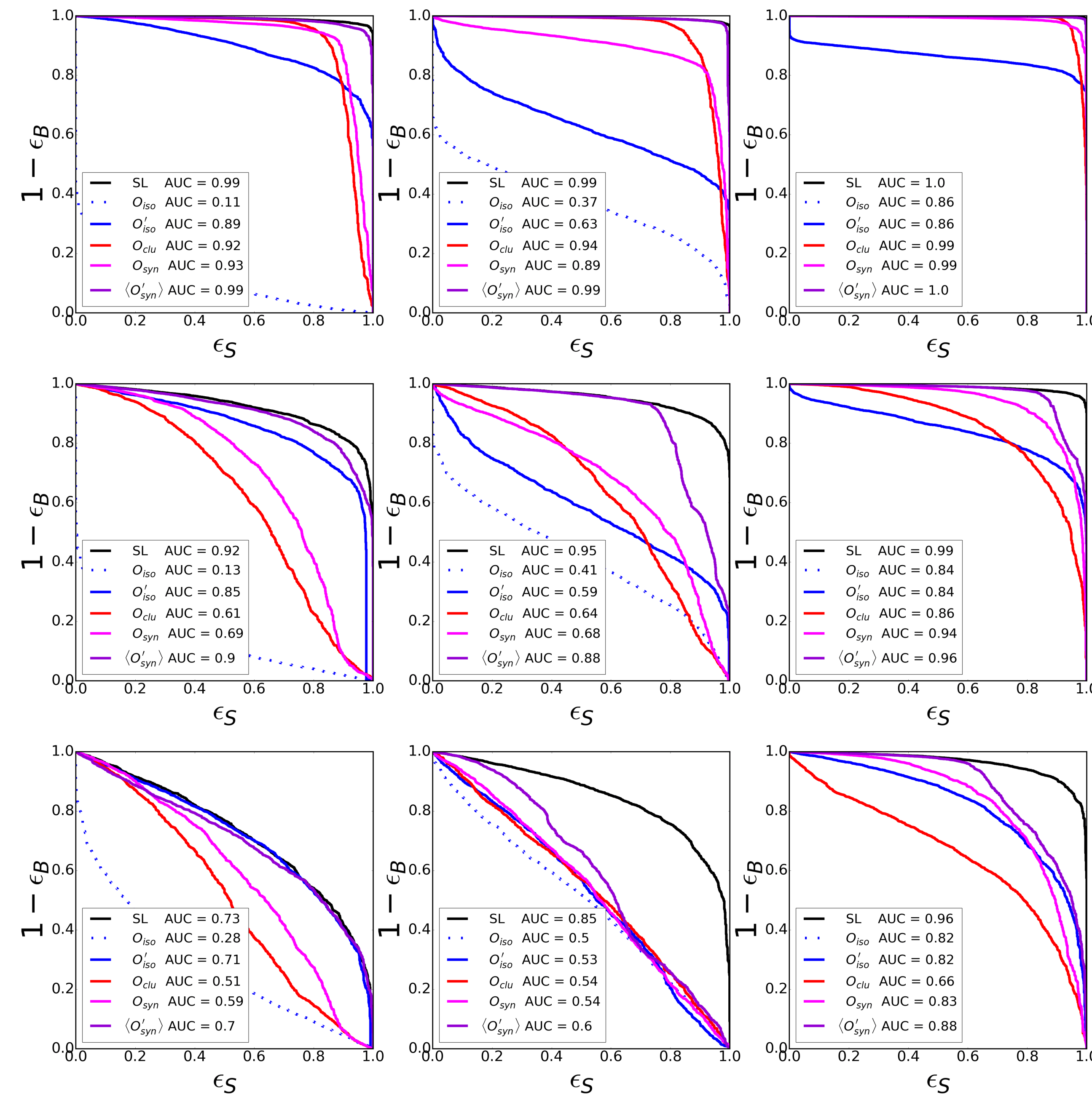
Workflow



x_i
I
auto-encoder
II
$\mathcal{O}_{clu}$
$\mathcal{O}_{iso}$
III
$\mathcal{O}_{syn}$ contour
$\mathcal{O}_{clu}$
$\mathcal{O}_{iso}$
IV
signal region
sgl
bkg
bref
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0 0.2 0.4 0.6 0.8 1.0
$\mathcal{O}_{syn}$
V
supervised DNN
$\mathcal{O}'_{syn}$

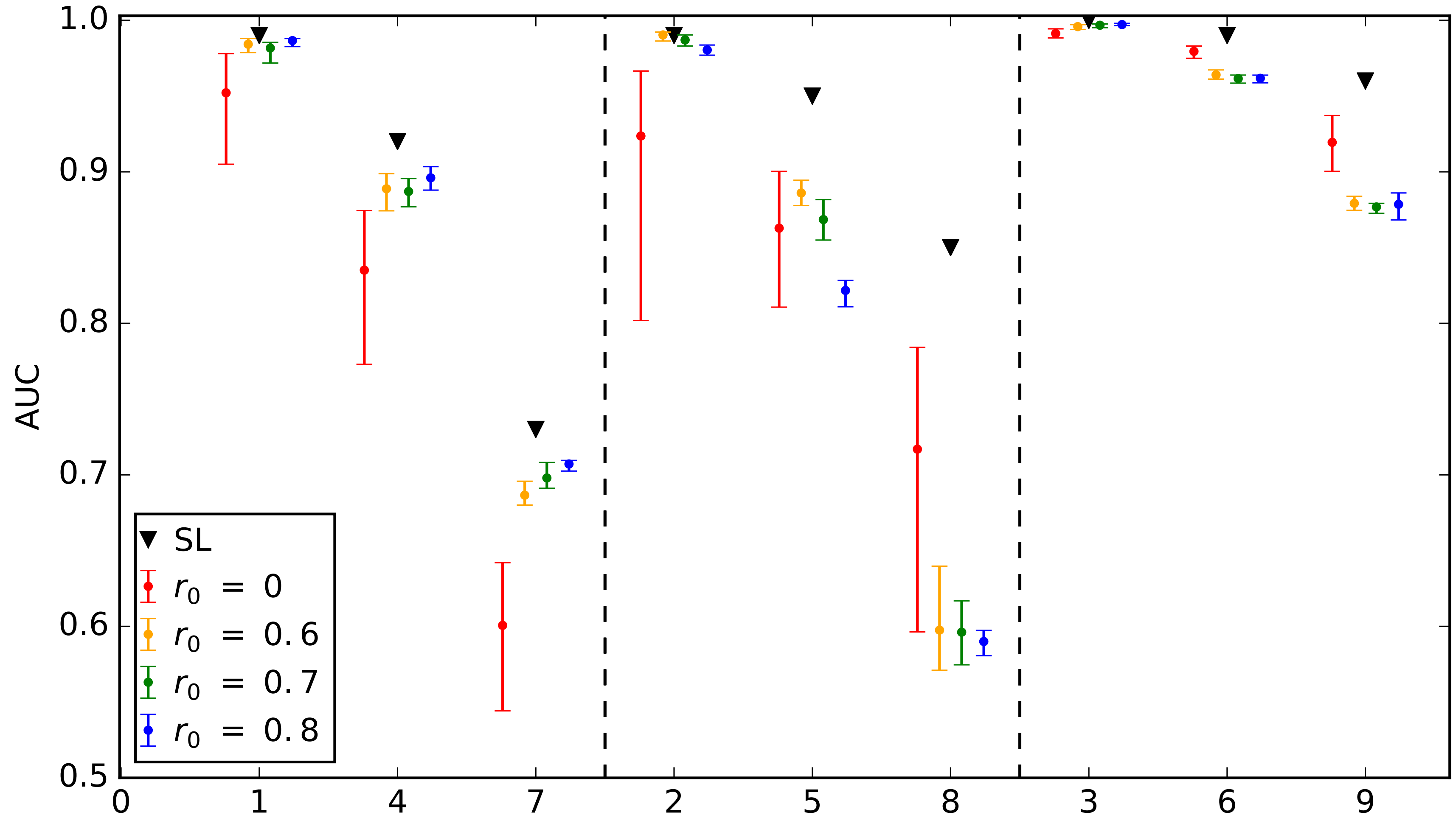
B = 100, 000   S = 1, 000





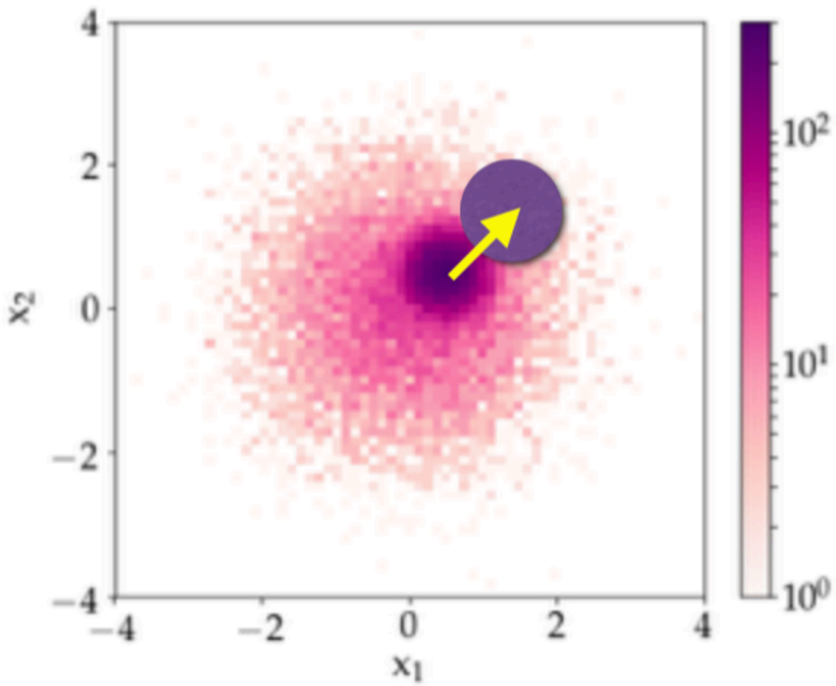
Robustness

20 classifiers with different initialisations are trained in every case.



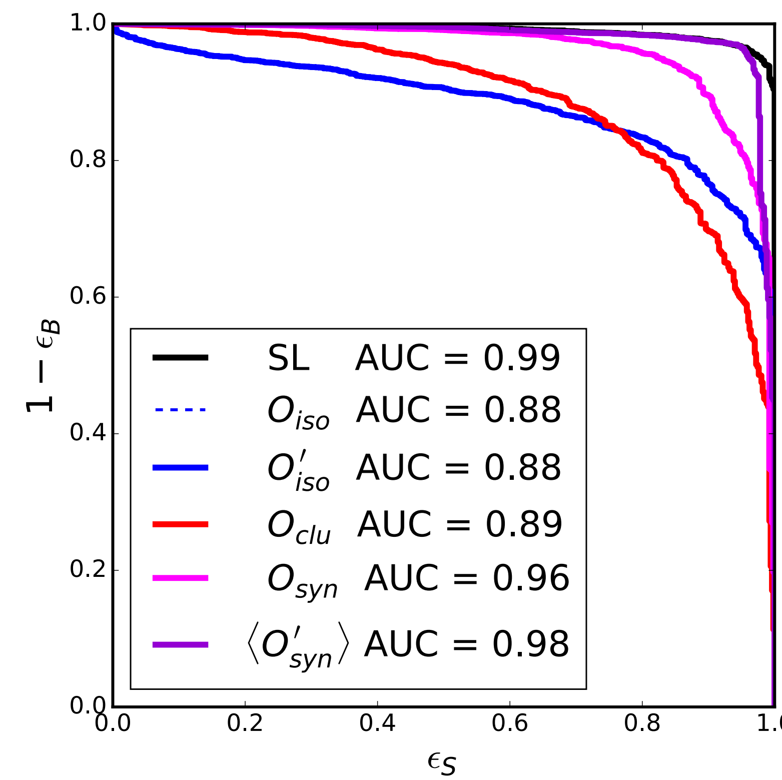
Performance comparison of $\mathcal{O}'_{\text{syn}}$ ($r_0 = 0.6,\ 0.7,\ 0.8$) with the full WSL ($r_0 = 0$)

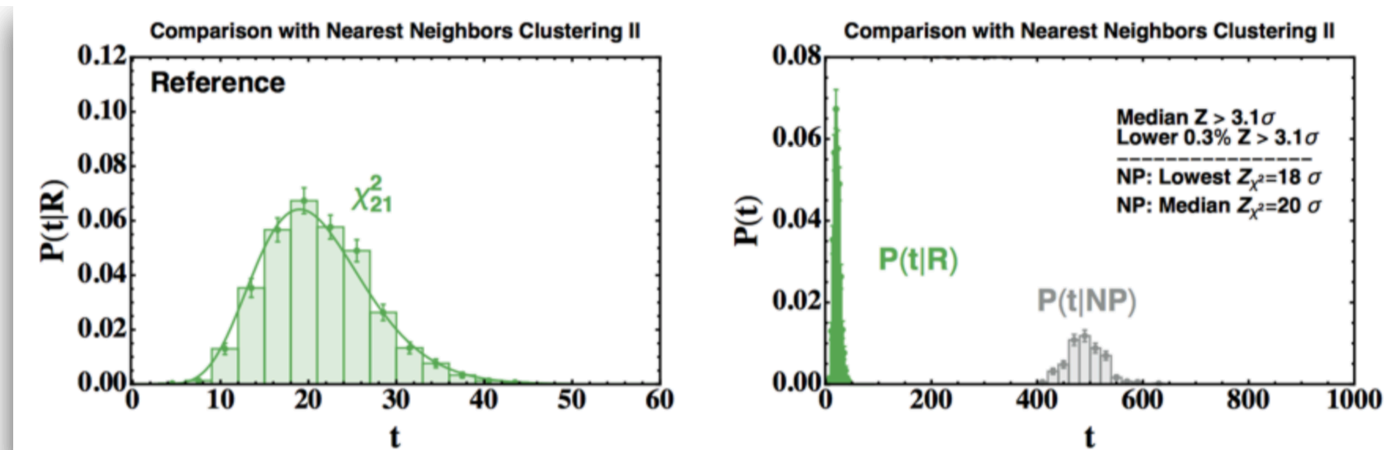
Comparison with arxiv: 1912.12155



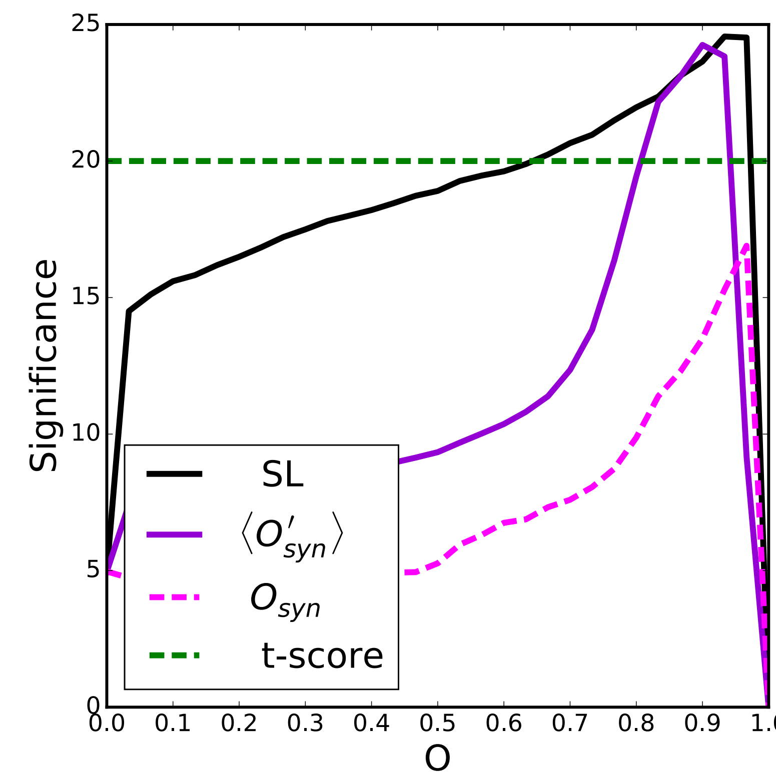
$$B : \mu = (0,0), \sigma = 1$$
$$S : \mu = (0,0), \sigma = \sqrt{0.1}$$
$$B = 10000, S = 500$$

arxiv:1807.10261





The median significance, obtained with a χ^2 approximation of the test statistic, is 20σ, while Ref. [29] quotes between 5 and 16σ for the nearest neighbors approach depending on the cut on their discriminating variable. We can conclude that both approaches are sensitive to the simple problem at hand.



arxiv:1912.12155

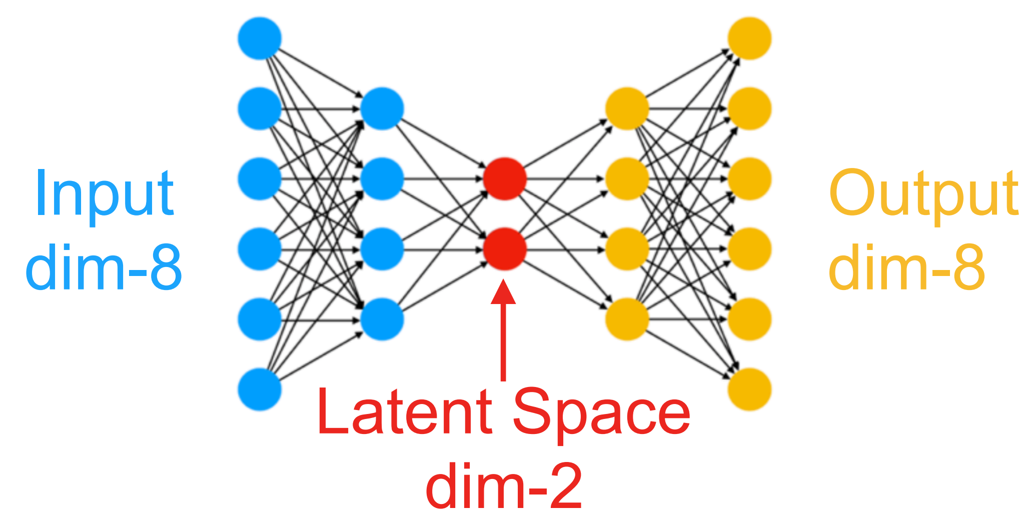
Benchmark Studies

the analysis of $t\bar{t}\gamma\gamma$ with the data of 3 ab^{-1}@13 TeV. We will focus on the di-photon kinematics only for simplicity. The potential signals include (1) the tth Higgs production with a resonance decay $h \to \gamma\gamma$; and (2) stop pair production in gravity-mediated SUSY, with a chain decay to $t\bar{t}\gamma\gamma$ + MET. They represent two typical signal patterns at colliders: resonance and broad shape.

Relational autoencoder is used for dimensionality reduction.

Input: 4-momenta of di-photons.

	Process	Matching	Event yields
Backgrounds	$t\bar{t}\gamma\gamma$	No	765
	$t\bar{t}\gamma$	Up to one jet	370
	$t\bar{t}$	Up to two jets	83
	Continuum $\gamma\gamma$	Up to two jets	1216
$t\bar{t}h$	$t\bar{t}h(\gamma\gamma)$	No	167
SUSY	$\tilde{t}\bar{\tilde{t}} \to t\bar{t}\tilde{\chi}\tilde{\chi} \to t\bar{t}\gamma\gamma + 2\tilde{G}$	No	226



preselect the events, as [34] does, by requiring two hardest photons ($p_T > 25\,\text{GeV}$, $|\eta| < 2.37$ but with an exclusion of $1.37 < |\eta| < 1.52$), one isolated lepton ($p_T > 10\,\text{GeV}$, $|\eta| < 2.7$ but with an exclusion of $1.37 < |\eta| < 1.52$ for electrons), at least two central jets ($p_T > 25\,\text{GeV}$ and $|\eta| < 2.4$) and one b-tagged jet ($p_T > 25\,\text{GeV}$ and $|\eta| < 2.4$) with a b-tagging efficiency of 70% determined by the $t\bar{t}$ sample. The event yields after the preselection are summarized in Tab. 2. In addition, we generate 1.3×10^5 background events satisfying the preselection criteria for defining the reference sample.

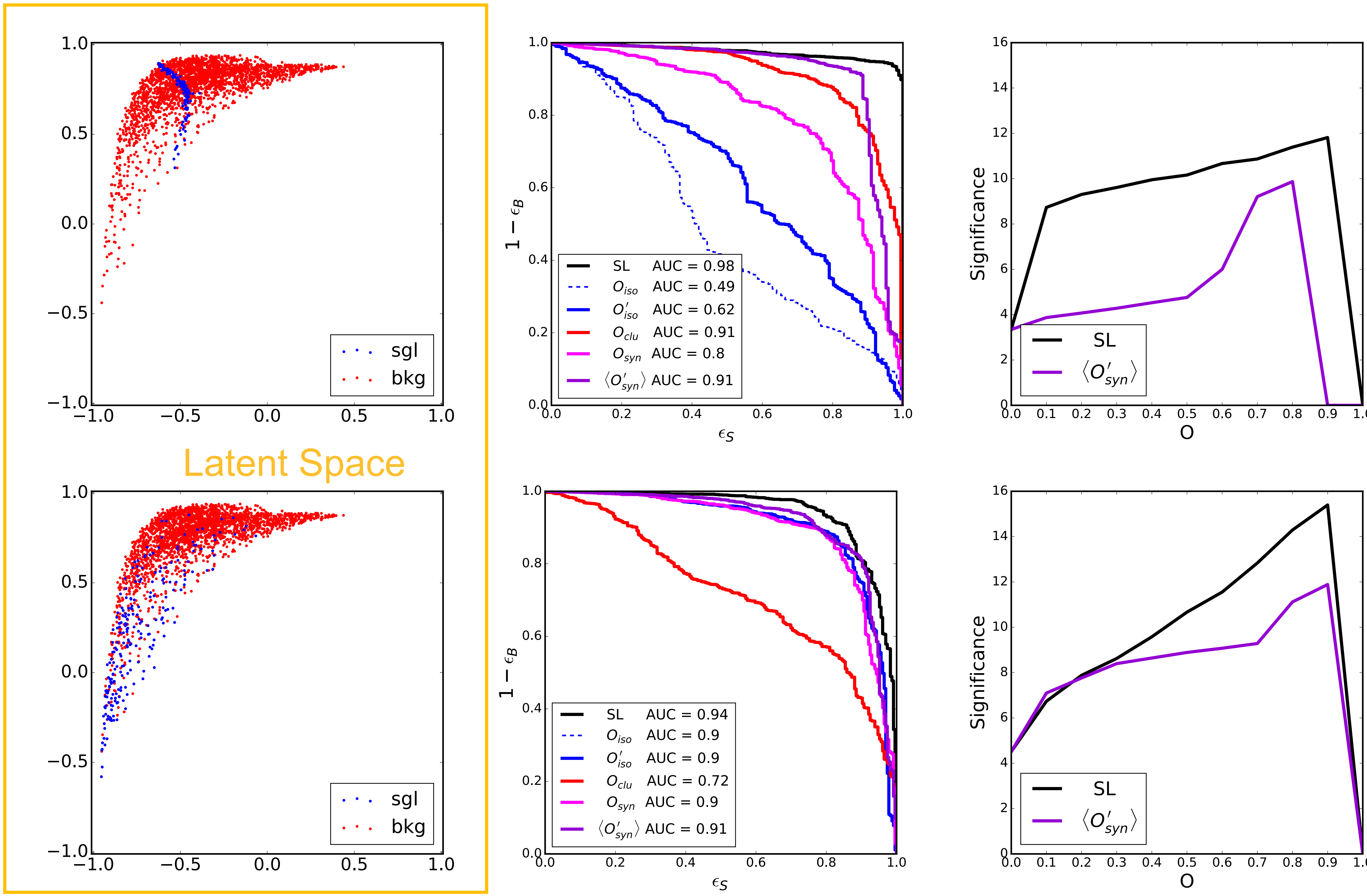
Loss function:

$$L' = |x - x'|^2 + c \times \left(\frac{m'^2_{\gamma_1}}{\sigma^2_{E_1}} + \frac{m'^2_{\gamma_2}}{\sigma^2_{E_2}} \right) + c \times \frac{(m_{\gamma\gamma} - m'_{\gamma\gamma})^2}{\sigma^2_{m_{\gamma\gamma}}}$$

c = 10

$t\bar{t}h$

SUSY



Latent Space
sgl
bkg
sgl
bkg
$1-\epsilon_B$
ϵ_S
SL AUC = 0.98
O_{iso} AUC = 0.49
O'_{iso} AUC = 0.62
O_{clu} AUC = 0.91
O_{syn} AUC = 0.8
$\langle O'_{syn} \rangle$ AUC = 0.91
$1-\epsilon_B$
ϵ_S
SL AUC = 0.94
O_{iso} AUC = 0.9
O'_{iso} AUC = 0.9
O_{clu} AUC = 0.72
O_{syn} AUC = 0.9
$\langle O'_{syn} \rangle$ AUC = 0.91
Significance
O
SL
$\langle O'_{syn} \rangle$
Significance
O
SL
$\langle O'_{syn} \rangle$

Conclusion

Isolation-based and Clustering-based evaluators manage to detect novelty along different strategies. Synergy of the 2 shines light on improving the performance of novelty detection.

We have worked on 9 two-dimensional Gaussian samples which mimic a broad range of NP scenarios. The geometry mean of the isolation-based and clustering-based evaluators is used as the synergy-based one. Then it is further optimised with a DNN, which shows encouragingly comparable AUC values with supervised learning. The robustness is tested and is compared with fully weakly supervised learning.

We have applied our algorithm onto physical benchmark studies with signal either forming a resonance or a broad shape. The analysis is done at detector level and it reaches a discovery confidence level comparable to the dedicated supervised learning search under the framework of invariant-mass-preserving-autoencoder.


Thank you.