Novelty Detection in HEP Data Analysis

Ying-Ying Li

HKUST

Based on arXiv: 1807.10261

in collaboration with Jan Hajer, Tao Liu and He Wang

April 11, 2019

Advanced Workshop on Accelerating the Search for Dark Matter with Machine Learning

Supervised Learning

highly efficient in analysing signal events with complex topologies

Q:

- Given that new physics scenarios may share similar final states, can we search for them simultaneously and more efficiently

Case I: di-top partner production vs Z′ production (decay to top pair)
Case II: exotic Higgs decays (rich decay modes)

- Also, given the null results at LHC, new physics could be very unexpected.
- Supervised learning, being model-dependent, is incapable for these tasks.

Novelty Detection

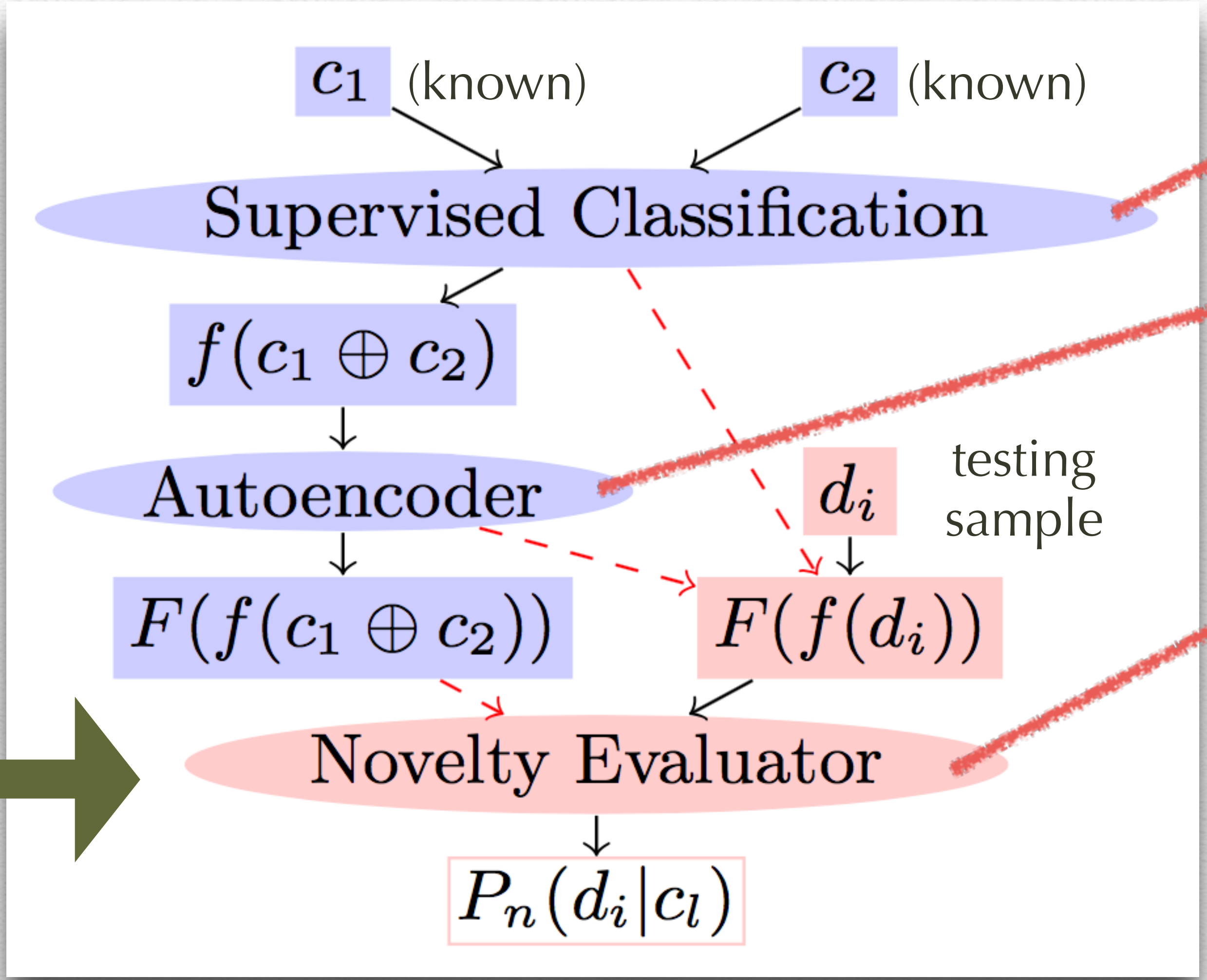
nice review [M.A.F.Pimental. et al] (2014)

Is ``model"- independent, complementary to supervised learning,
Allows us to detect new physics without a prior knowledge about it.

Novelty Detection in HEP Data Analysis

Workflow for Novelty Detection

[J.Hajer, YYL, T. Liu, H. Wang] (2018)



- Step 1: (SM/background) feature learning
- Step 2: dimension reducing of feature space (auto-encoder)
- Step 3: novelty evaluating of testing data
- Analyze detection sensitivity based on novelty response of testing data

Allow us to detect new physics model-independently

Novelty Detection in HEP Data Analysis

Novelty Evaluators: traditional wisdom

[H.Kriegel, P.Kroger, E.Schubert and A.Zimek] (2009)

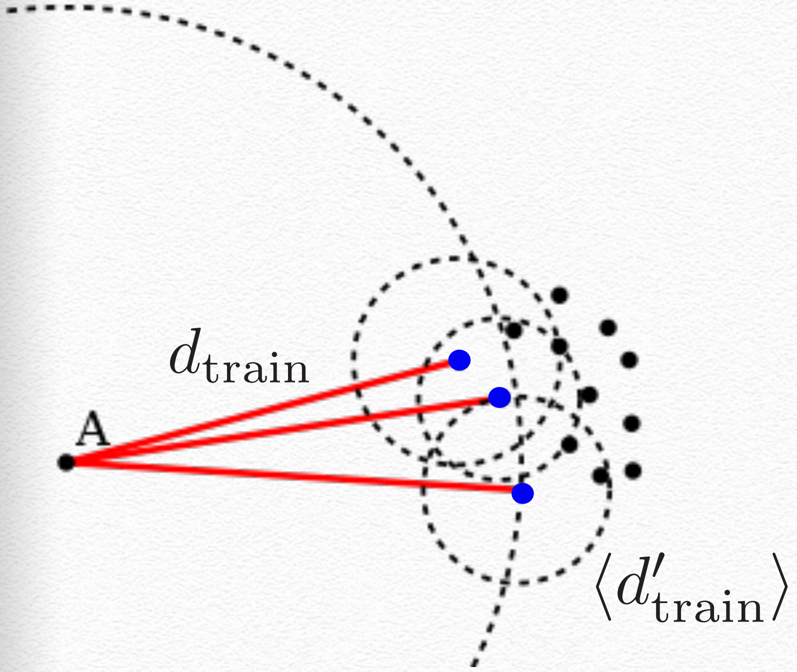
[R.Socher, M.Ganjoo, C.D.Manning and A.Ng] (2013)

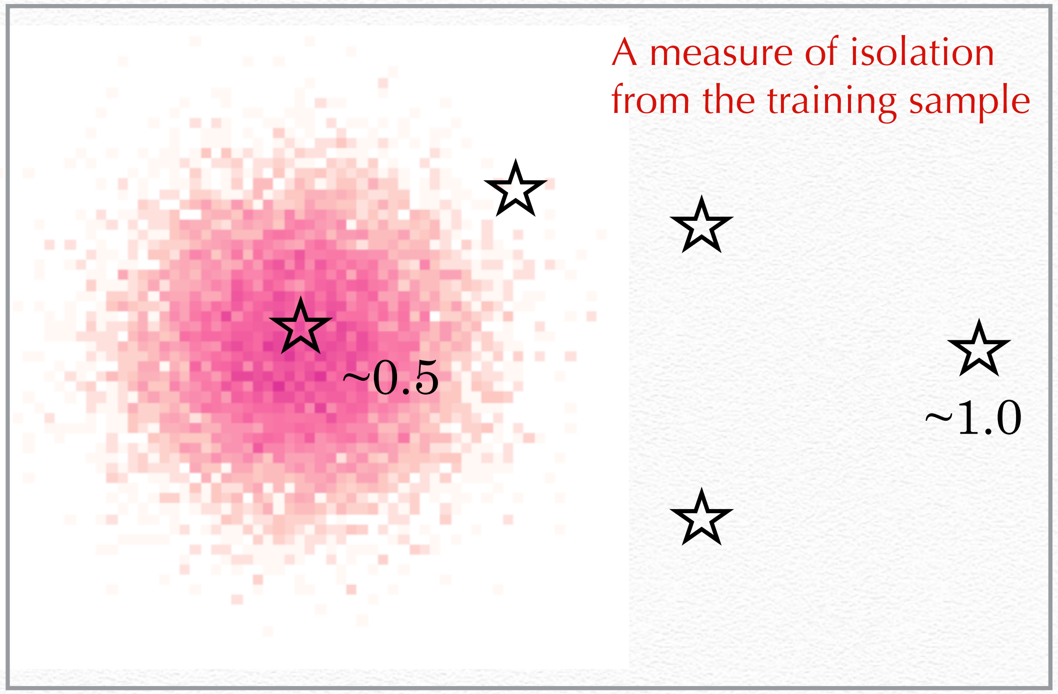
$$\Delta_{\text{trad}} = \frac{d_{\text{train}} - \langle d'_{\text{train}} \rangle}{\langle d'^2_{\text{train}} \rangle^{1/2}}$$

$$\mathcal{O} = \frac{1}{2}\left(1 + \text{erf}\left(\frac{c\Delta}{\sqrt{2}}\right)\right)$$

Novelty measure: range unnormalised

Novelty score: $0 \leq \mathcal{O} \leq 1$

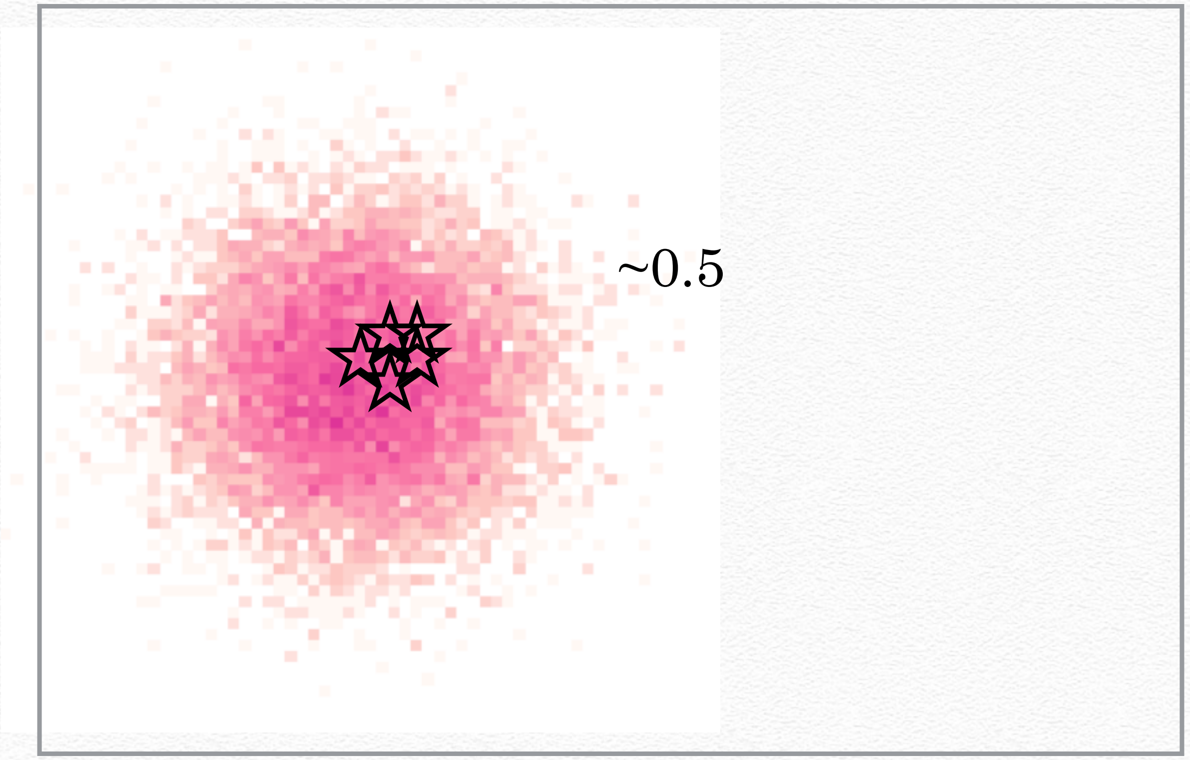




Novelty Detection in HEP Data Analysis

Novelty Evaluators: traditional wisdom

$$\Delta_{\mathrm{trad}} = \frac{d_{\mathrm{train}} - \langle d'_{\mathrm{train}} \rangle}{\langle d'^2_{\mathrm{train}} \rangle^{1/2}}$$



[H.Kriegel, P.Kroger, E.Schubert and A.Zimek] (2009)

[R.Socher, M.Ganjoo, C.D.Manning and A.Ng] (2013)

$$\mathcal{O} = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{c\Delta}{\sqrt{2}}\right)\right)$$

- However, this design is insensitive to the clustering of the testing data with unknown pattern
- Recall: the clustering features such as resonances, shape, etc., could be important for BSM physics detection
- The testing data of unknown pattern with such features are scored low, unlesss they are away from the training data distribution!

Novelty Detection in HEP Data Analysis

New Novelty Measure

[J.Hajer, YYL, T. Liu, H. Wang] (2018)

$$\Delta_{\text{trad}} = \frac{d_{\text{train}} - \langle d'_{\text{train}} \rangle}{\langle d'^2_{\text{train}} \rangle^{1/2}} \qquad \Delta_{\text{new}} = \frac{d_{\text{test}}^{-m} - d_{\text{train}}^{-m}}{d_{\text{train}}^{-m/2}}$$

d_{train} mean distance of a testing data point to its k nearest neighbors in the training dataset

d_{test} mean distance of a testing data point to its k nearest neighbors in the testing dataset

m dimension of the feature space

Novelty response is evaluated by comparing local densities in the training and testing datasets

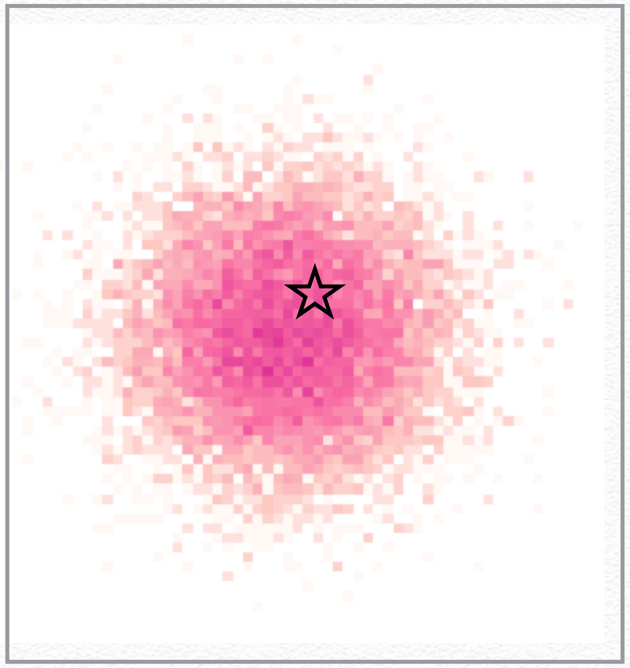
Approximately statistical interpretation : $\Delta_{\text{new}} \propto \frac{S}{\sqrt{B}}\Big|_{\text{local bin}}$

Novelty Detection in HEP Data Analysis

New Novelty Measure

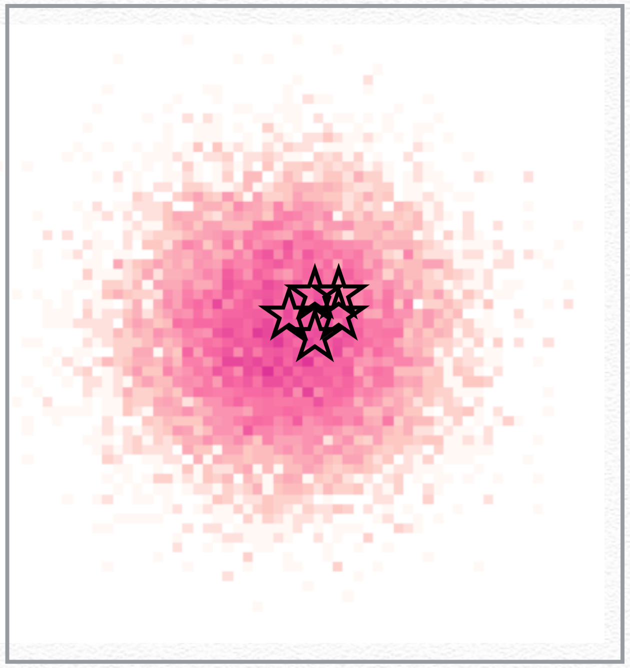
[J.Hajer, YYL, T. Liu, H. Wang] (2018)

$$\Delta_{\mathrm{trad}} = \frac{d_{\mathrm{train}} - \langle d'_{\mathrm{train}} \rangle}{\langle d'^2_{\mathrm{train}} \rangle^{1/2}} \qquad \Delta_{\mathrm{new}} = \frac{d_{\mathrm{test}}^{-m} - d_{\mathrm{train}}^{-m}}{d_{\mathrm{train}}^{-m/2}}$$



Training dataset

VS



Testing dataset

- Big density difference => high score
- Small density difference => low score
- => a measure of clustering

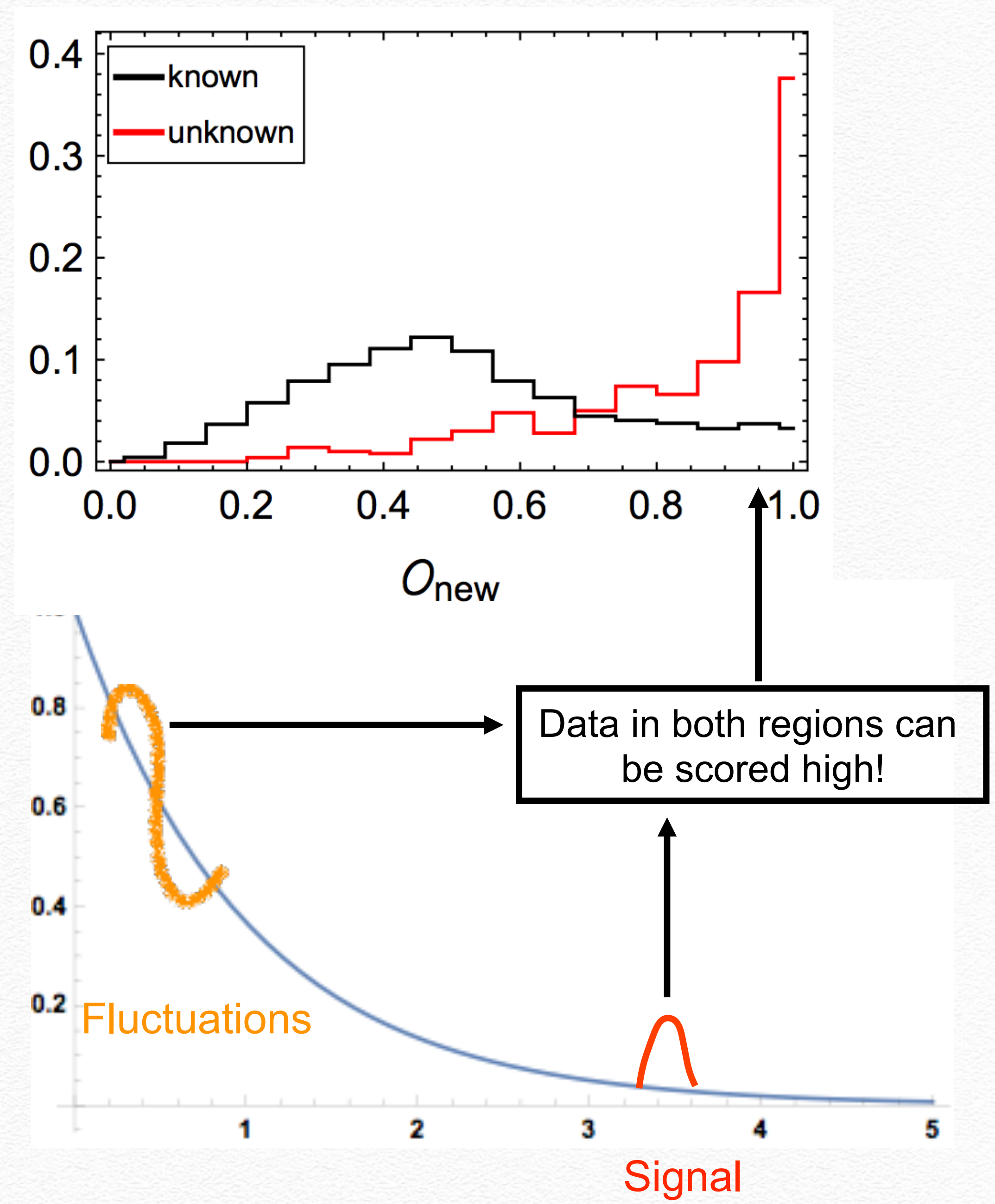
Novelty Detection in HEP Data Analysis

Addressing Look Elsewhere Effect

$$\Delta_{\mathrm{new}} = \frac{d_{\mathrm{test}}^{-m} - d_{\mathrm{train}}^{-m}}{d_{\mathrm{train}}^{-m/2}}$$

Without a priori knowledge on the BSM physics, novelty detection might suffer from a large ``Look Elsewhere Effect (LEE)'', given the feature space to probe!

[J.Hajer, YYL, T. Liu, H. Wang] (2018)



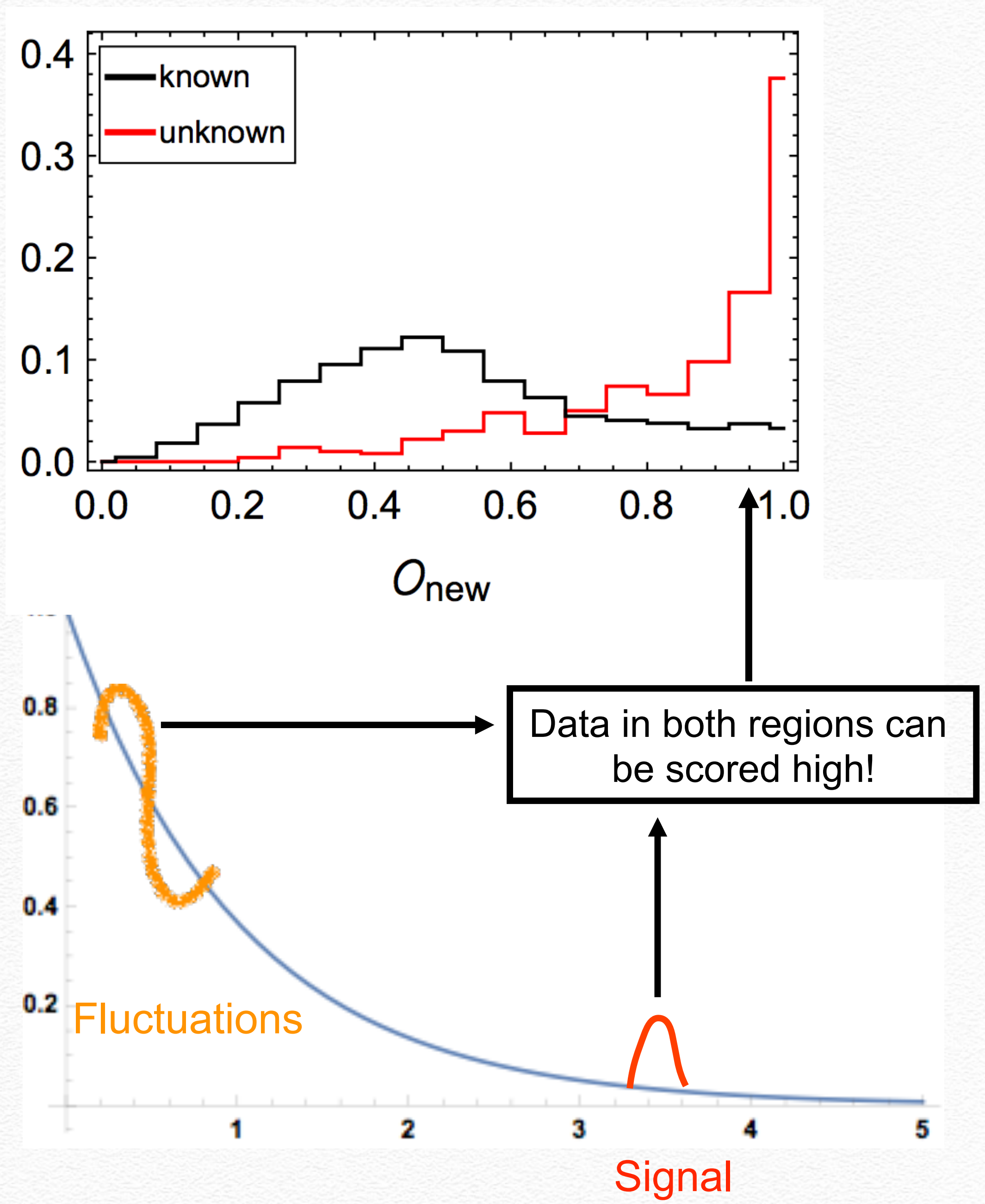
Novelty Detection in HEP Data Analysis

Addressing Look Elsewhere Effect

$$\Delta_{\rm new} = \frac{d_{\rm test}^{-m} - d_{\rm train}^{-m}}{d_{\rm train}^{-m/2}}$$

Without a priori knowledge on the BSM physics, novelty detection might suffer from a large ``Look Elsewhere Effect (LEE)", given the feature space to probe!

[J.Hajer, YYL, T. Liu, H. Wang] (2018)

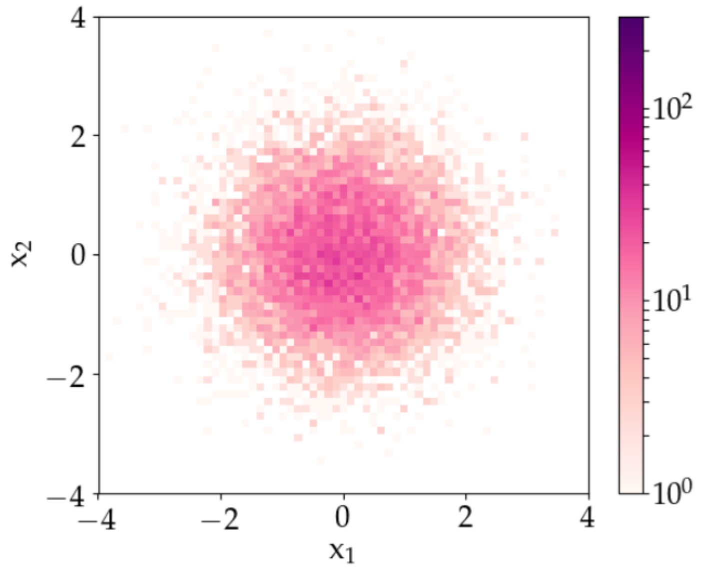


To compensate for high-scoring (by O_new) of known-pattern data from high-density region

$$\mathcal{O}_{\rm comb} = \sqrt{\mathcal{O}_{\rm trad}\mathcal{O}_{\rm new}}$$

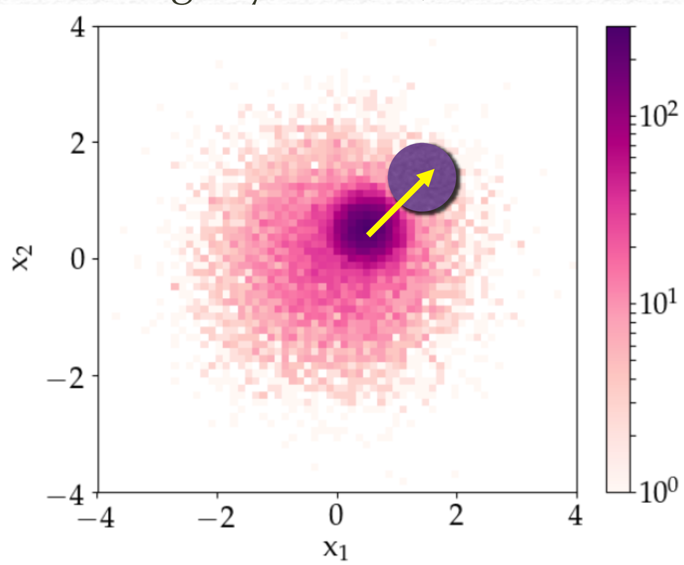
Addressing Look Elsewhere Effect

[J.Hajer, YYL, T. Liu, H. Wang] (2018)

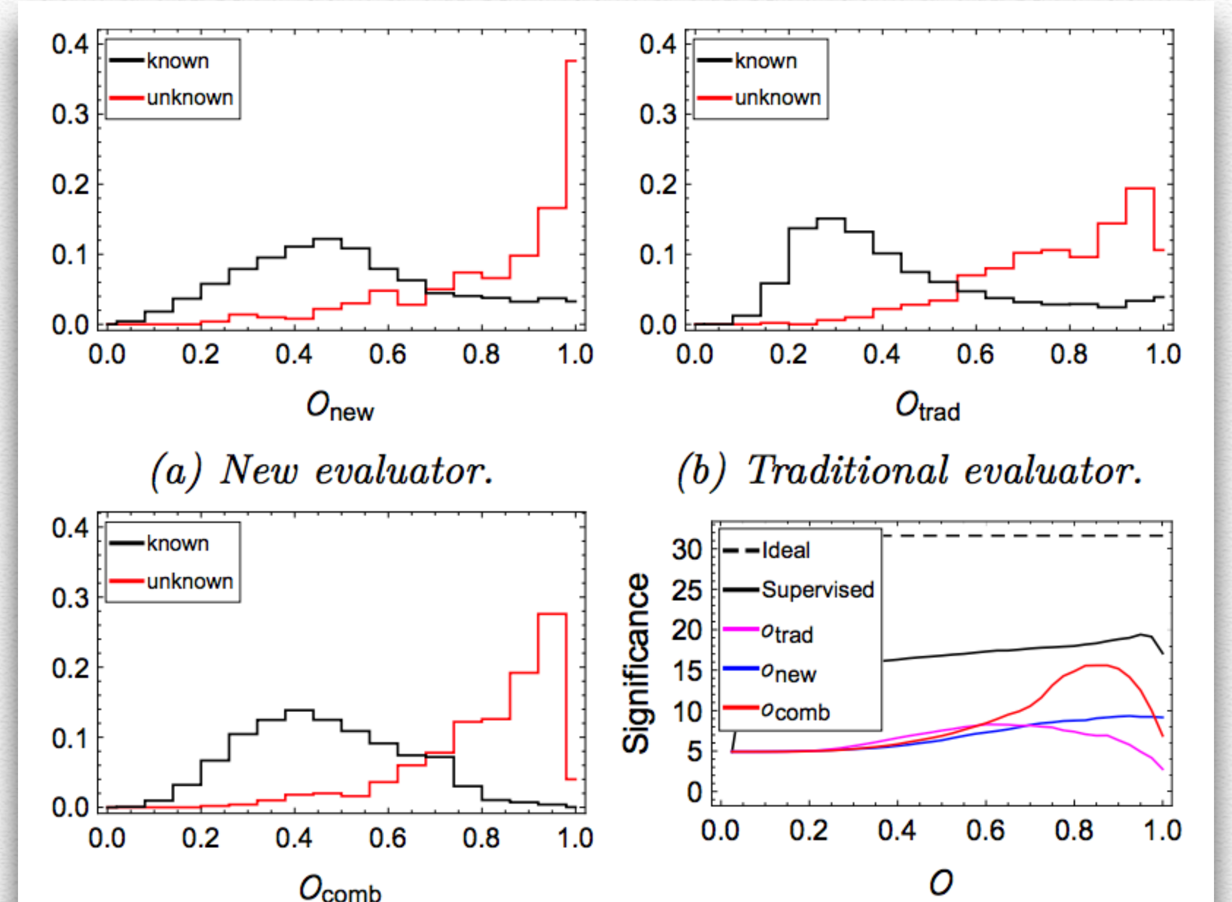


(a) Training data.

Center slightly shifted, with S/B=1/20



(b) Testing data.



(a) New evaluator. *(b) Traditional evaluator.* *(c) Combined evaluator.* *(d) Significance.*

O_comb based analysis yields more than 50% improvement in detection sensitivity!

Benchmark Analysis

[J.Hajer, YYL, T. Liu, H. Wang] (2018)

Analysis one: di-top(leptonic) production at LHC

- $pp \to \bar{t}_l t_l$, $\sigma = 11.5\,\mathrm{fb}$,
- $pp \to t_l \bar{b} W_l^{\pm}$, $\sigma = 0.365\,\mathrm{fb}$,
- $pp \to Z_b Z_l$, $\sigma = 0.0765\,\mathrm{fb}$.

$\mathbf{X_1}$: $pp \to \overline{T}T \to W_l^+ W_l^- \bar{b}b$

$\mathbf{X_2}$: $pp \to Z' \to \bar{t}t$

Analysis two: exotic Higgs decays at e+e- collider

- $e^+e^- \to hZ \to Z^*_{\mathrm{inv}} Z_{\bar{b}b} l^+ l^-$ $\sigma = 0.00686\,\mathrm{fb}$,
- $e^+e^- \to hZ \to Z^*_{\bar{b}b} Z_{\mathrm{inv}} l^+ l^-$ $\sigma = 0.00259\,\mathrm{fb}$.

$\mathbf{Y_1}$: $h \to \widetilde{\chi}_1 \widetilde{\chi}_2 \to \widetilde{\chi}_1 \widetilde{\chi}_1 a$.

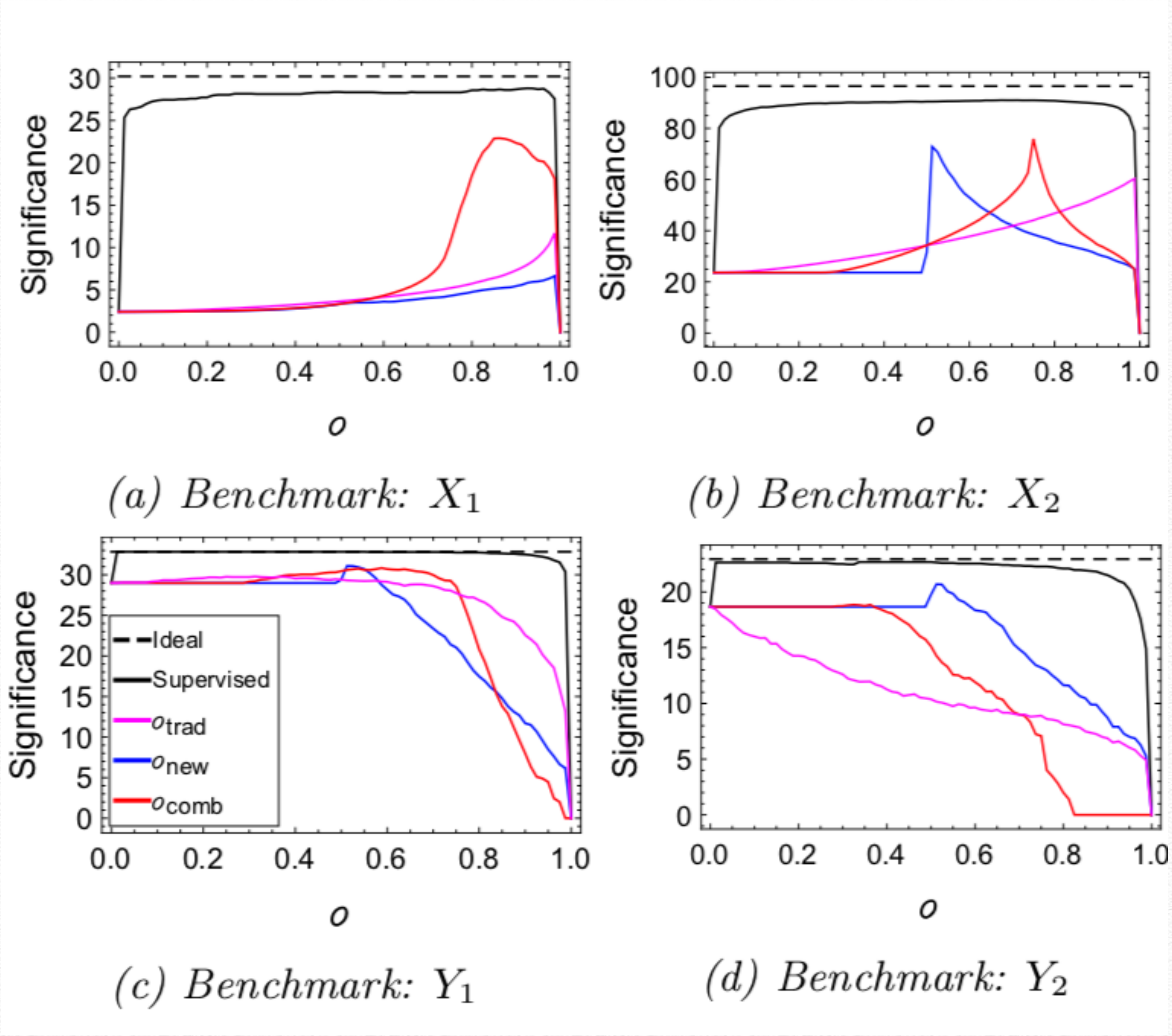
$\mathbf{Y_2}$: $h \to Za$

	Parameter values	$\sigma(fb)$
$X1$	$m_T = m_{\overline{T}}$ 1.2 TeV, BR$(T \to W_l^+ b) = 50\,\%$	0.152
$X2$	$m_{Z'} = 3\,\mathrm{TeV}$, $g_{Z'} = g_Z$, BR$(Z' \to \bar{t}t) = 16.7\,\%$	1.55
$Y1$	$m_{N_1} = \frac{m_{N_2}}{9} = \frac{m_a}{4} = 10\,\mathrm{GeV}$, BR$(h \to \bar{b}bE_T^{\mathrm{miss}}) = 1\,\%$	0.108
$Y2$	$m_a = 25\,\mathrm{GeV}$, BR$(h \to \bar{b}bE_T^{\mathrm{miss}}) = 1\,\%$	0.053

Novelty Detection in HEP Data Analysis

Benchmark Analysis

[J.Hajer, YYL, T. Liu, H. Wang] (2018)



(a) Benchmark: X_1 *(b) Benchmark: X_2* *(c) Benchmark: Y_1* *(d) Benchmark: Y_2*

- X1: well-modelled by the Gaussian sample!
- X2: O_comb less efficient due to one-order larger S/B
- Y1 and Y2: O_new performs universally better than the others, due to large S/B
- The sensitivities based on the algorithm designed are not far below the ones set by supervised learning

Conclusion and Outlook

- Novelty detection: searching for new physics model-independently

established the framework of novelty detection

including detector effects? further improve the algorithm, feedback, combination of three evaluators?

computing efficiency and realtime implementation

We are heading forward to the truth of our nature

gradually and passionately

Thank YOU!

Backup

Chapter 4 Novelty Detection

Signal Processing 99 (2014) 215–249

Contents lists available at ScienceDirect

Signal Processing

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

ELSEVIER

SIGNAL PROCESSING

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.