

Использование машинного обучения в физике высоких энергий

Саров, 29 июля 2022

Ф. Ратников

НИУ ВШЭ, лаборатория Лямбда



Анализ данных

- Анализ данных — область математики и информатики, занимающаяся построением и исследованием наиболее общих математических методов и вычислительных алгоритмов извлечения знаний из экспериментальных (в широком смысле) данных
 - процесс исследования, фильтрации, преобразования и моделирования данных с целью извлечения полезной информации и принятия решений.
 - анализ данных имеет множество аспектов и подходов, охватывает разные методы в различных областях науки и деятельности.



Математическая модель

- Математическая модель является представлением некоторой реальности, выраженное математическими средствами.
- В данном контексте модели описывают соотношения и зависимости между количественными характеристиками объектов и/или явлений
- Основные компоненты:
 - собственные параметры модели - θ
 - пространство области определения модели - X
 - пространство области предсказаний модели - Y
 - собственно модель - $f(\theta) : X \rightarrow Y$
- Детерминистические модели: $y = f(x, \theta)$, $x \in X$, $y \in Y$
- Стохастические модели: $p(y | x) = f(x, y, \theta)$, $x \in X$, $y \in Y$

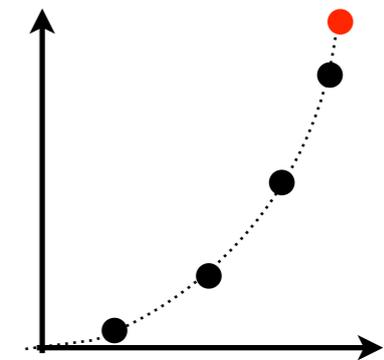
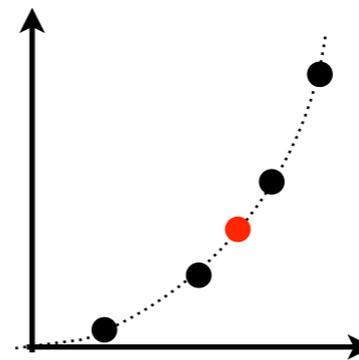
Типы математических характеристик

- Бинарные
 - да/нет
- Категориальные
 - синий/зелёный/мокрый
- Дискретные
 - 1, 2, 3, ...
- Текстовые
 - “abc”, ...
- Непрерывные
 - температура
- Произвольные сочетания
 - карта здоровья
 - поведение объекта в пространстве-времени
 - ...

Примеры из ФВЭ?

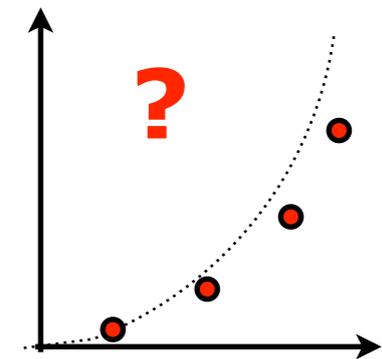
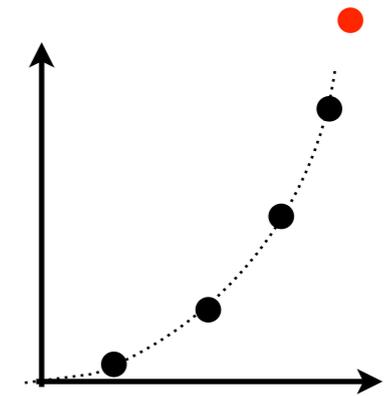
Предсказательность модели

- Модель должна уметь предсказывать наблюдаемый результат для неисследованных точек в пространстве параметров модели
 - абсолютное предсказание $x = gt^2/2$
 - интерполяция данных
 - экстраполяция данных
- Проверка правильности предсказаний - критерий качества модели

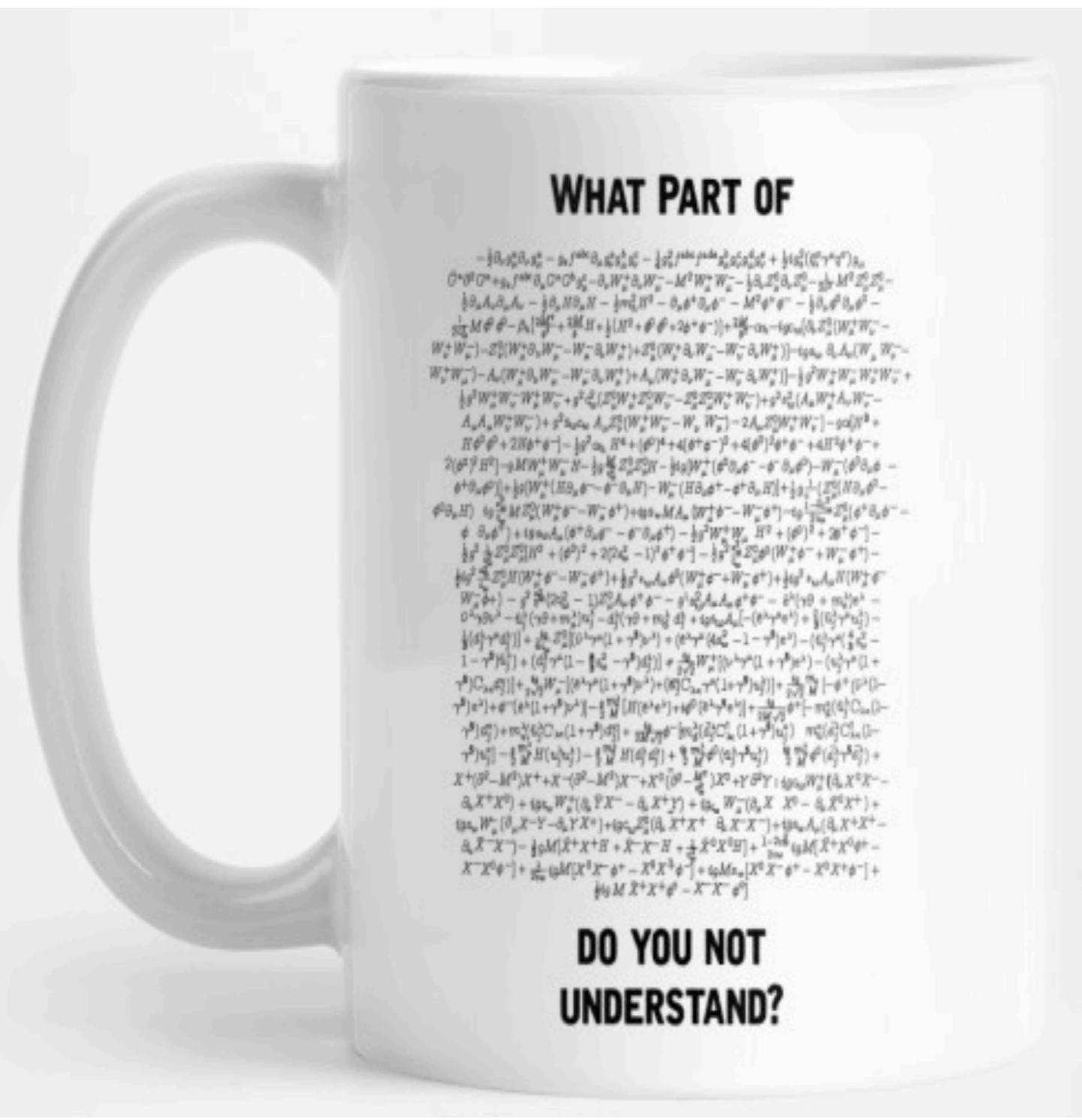


Научная модель

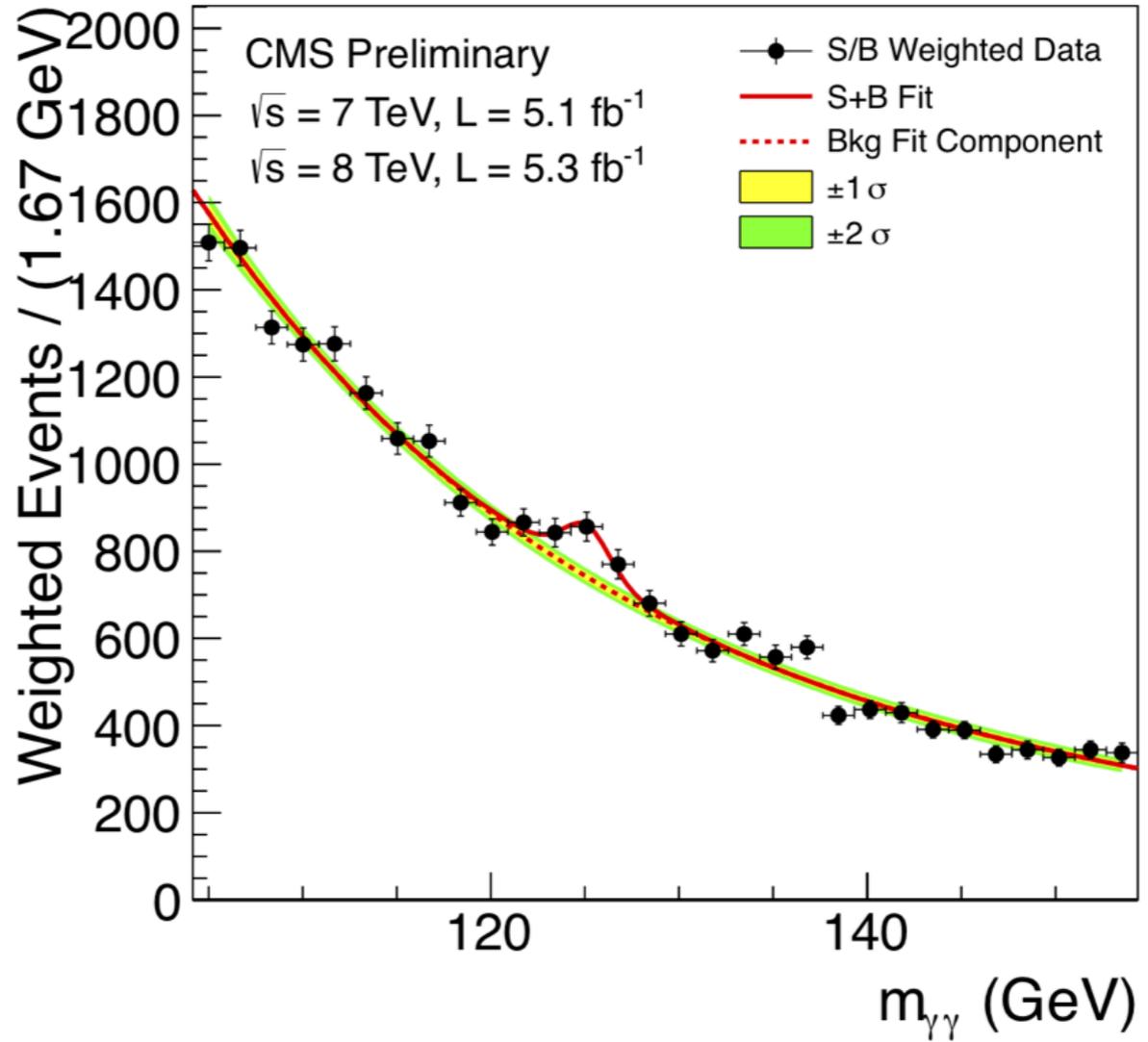
- Парадигма: модель предсказывает данные
- Специфические зависимости *a priori*, вытекающие из фундаментальных законов
- Количество значимых параметров (степеней свободы) ограничено
 - бритва Окама
- Модель можно экстраполировать вне области, в которой она прошла проверку
- Интерпретируемые предсказания
- Серьёзные проблемы в случае различия между предсказаниями и наблюдаемыми



Например, Стандартная модель



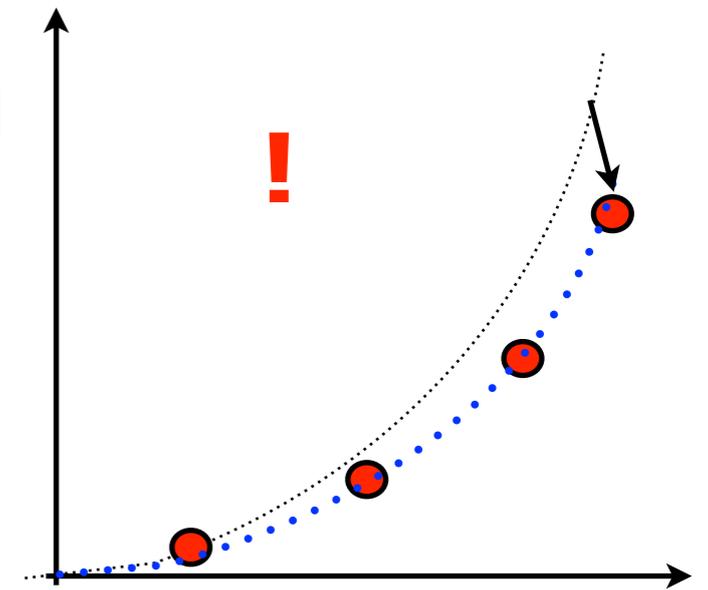
Модель



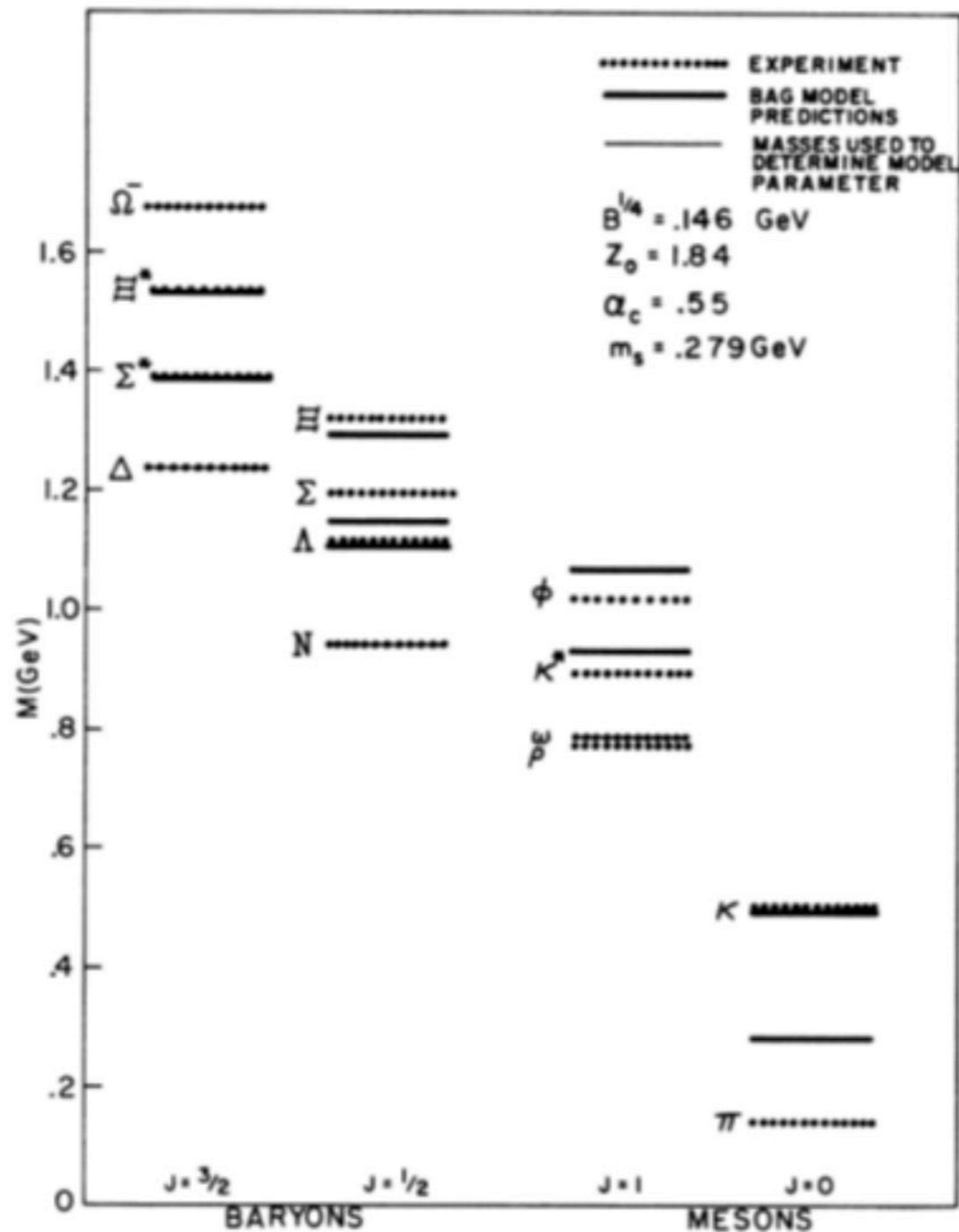
Данные

Эмпирическая (феноменологическая) модель

- Парадигма: модель предсказывает данные
- Специфические зависимости, отражающие наблюдаемые закономерности
- Количество значимых параметров (степеней свободы) увеличивается со временем
- Модель можно ограниченно экстраполировать вне области, в которой она прошла проверку
- Интерпретируемые предсказания
- Адаптация случае различия между предсказаниями и наблюдаемыми



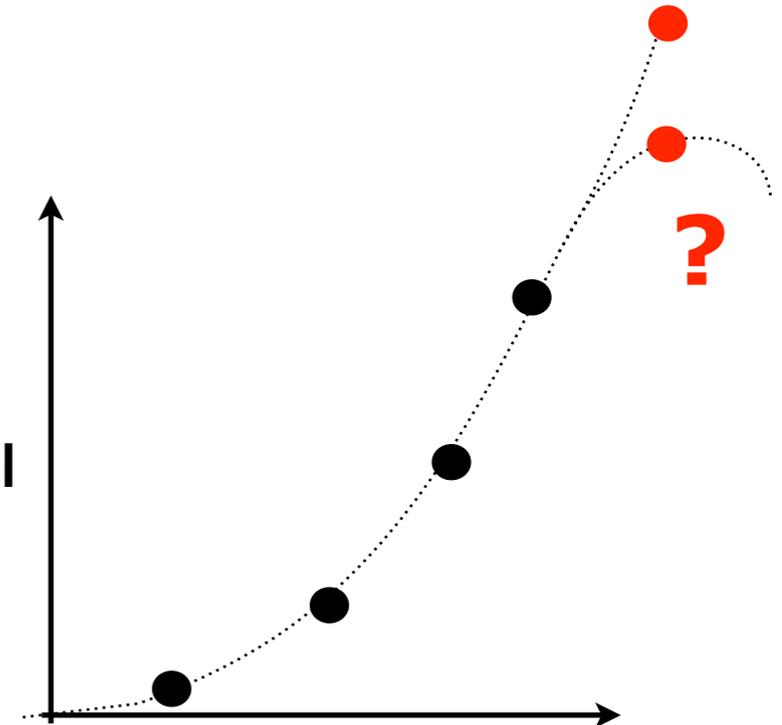
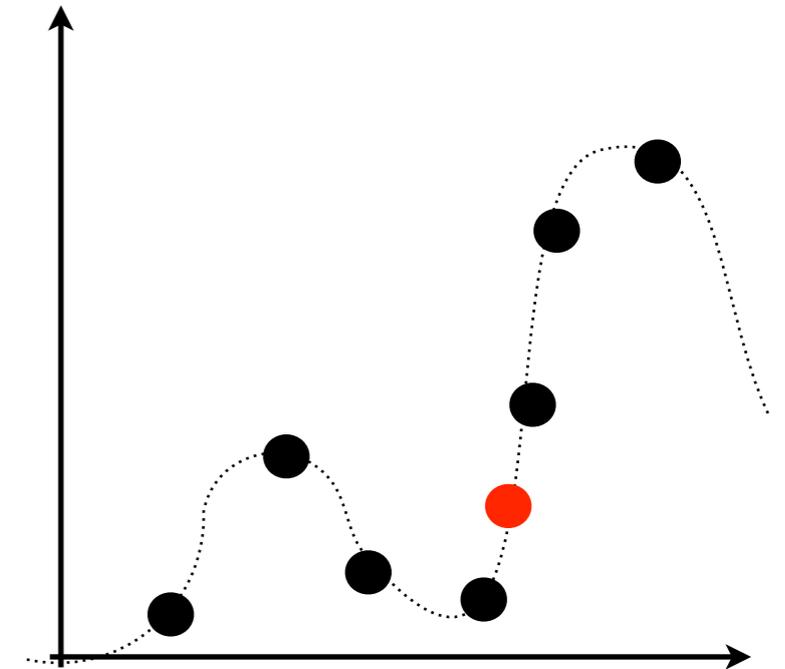
Например, модель мешков для адронов



- Phys. Rev., D, v. 12, no. 7, pp. 2060-2076 (1975)
- Предсказание масс мезонов и барионов
- модель откалибрована на $N, \Delta, \Omega, \omega$

Описательная модель

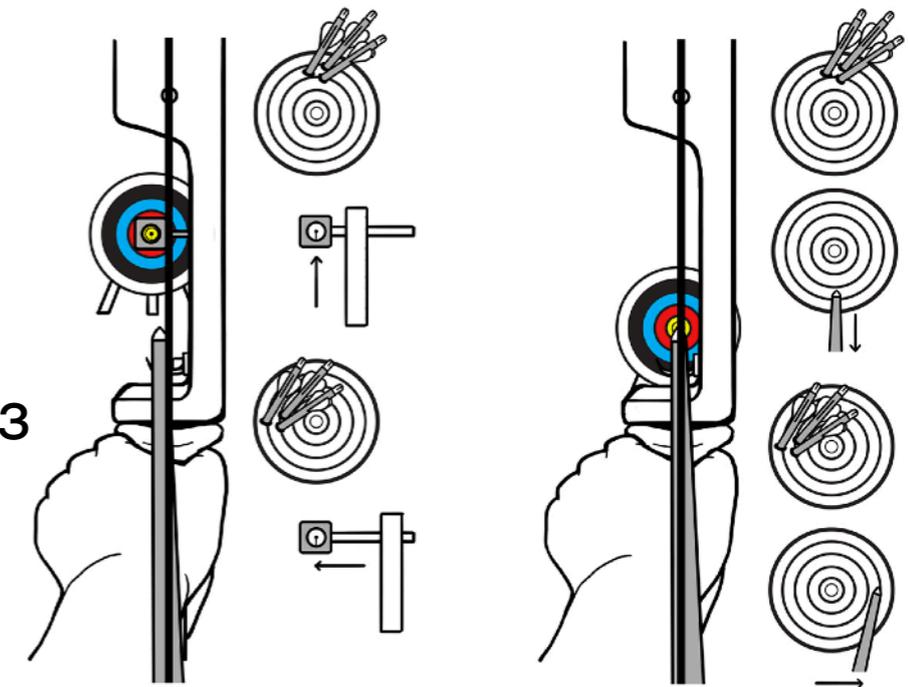
- Парадигма: модель описывает данные
 - данные первичны
 - нет существенных предположений о характере зависимостей *a priori*
- Модель имеет универсальный характер
- Количество параметров модели велико
 - параметры трудно интерпретируемы
- Экстраполяции модели нельзя доверять



Пример: стрельба из лука



- Сложное взаимодействие гравитации, аэродинамики, упругости, колебаний, ...
- Должен ли хороший стрелок быть хорошим аналитиком?
 - вовсе нет
- Возможно ли построить хорошую модель, исходя из первых принципов?
 - наврядли
- Можно ли стабильно попадать в цель?
 - вполне



Illustrated by © Jessica Emmett - www.jessica-emmett.com
This image can be used for personal and non-commercial use with credit. This image is a basic guide only, always consult a qualified archery instructor.

Другие примеры описательных моделей

- Подгонка данных полиномами/сплайнами
- Разложение в ряд/интеграл Фурье
- **Машинное обучение**

Что такое “Машинное Обучение”

Машинное обучение (англ. *machine learning*, ML) — класс методов **искусственного интеллекта**, характерной чертой которых является не прямое решение задачи, а обучение за счёт применения решений множества сходных задач. Для построения таких методов используются средства **математической статистики, численных методов, математического анализа, методов оптимизации, теории вероятностей, теории графов**, различные техники работы с **данными в цифровой форме**.



- МО - обобщенная задача аппроксимации
 - истинная зависимость \rightarrow выборка данных \rightarrow аппроксимация зависимости
 - $f(\mathbf{x}) \rightarrow X_{train} = \{\mathbf{x}_i\} \rightarrow a(\theta, \mathbf{x})$
 - в пространстве больших размерностей \mathbf{x} , i , θ

Что такое “Машинное Обучение”

- МО - обобщенная задача аппроксимации
 - истинная зависимость \rightarrow выборка данных \rightarrow аппроксимация зависимости
 - $f(\mathbf{x}) \rightarrow X_{train} = \{\mathbf{x}_i\} \rightarrow a(\theta, \mathbf{x})$
 - в пространстве больших размерностей \mathbf{x}, i, θ
 - $dim(\mathbf{x}) \sim 10 \dots 10^4 \dots$ “большие данные”
 - $dim(i) \sim 10^5 \dots 10^9 \dots$ “много данных”
 - $dim(\theta) \sim 10^3 \dots 10^6 \dots$ “сложные данные”
- Ключевые вопросы
 - метрика качества аппроксимации
 - семейство параметризованных аппроксимирующих функций $a(\theta, \mathbf{x})$
 - метод поиска оптимальной функции аппроксимации
 - $a_0(\mathbf{x}) = a(\theta_{best}, \mathbf{x})$

Что аппроксимируем?

- Истинная зависимость $f(\mathbf{x})$
 - X - множество известных признаков объекта
 - $\mathbf{x} \in X$
 - Y - множество предсказываемых признаков объекта
- **Детерминистская модель:** предсказываем признак
 - $y = f(\mathbf{x}), \mathbf{x} \in X, y \in Y$
- **Простая вероятностная модель:** предсказываем среднее значение и неопределенность признака
 - $(y, \sigma_y) = f(\mathbf{x}), \mathbf{x} \in X, y \in R^{(n)}$
- **Частная вероятностная модель:** предсказываем вероятностное распределение признака для данного набора известных признаков
 - $f(y, \mathbf{x}) = p(y | \mathbf{x}), \mathbf{x} \in X, y \in Y$ Условная генеративная модель
- **Полная вероятностная модель:** предсказываем совместное вероятностное распределение признака и набора известных признаков
 - $f(y, \mathbf{x}) = p(y, \mathbf{x}), \mathbf{x} \in X, y \in Y$ Безусловная генеративная модель

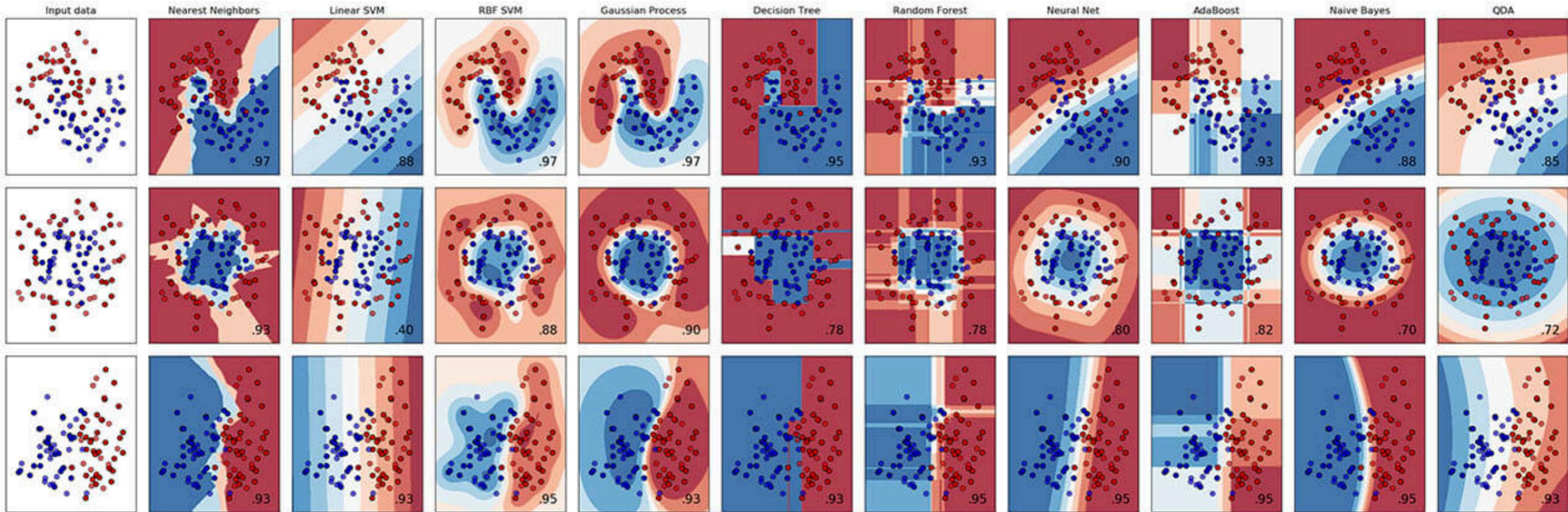
Типы машинного обучения

- С учителем
 - “я могу внятно сказать и показать, чего хочу”
 - научить модель воспроизводить наши решения
- Без учителя
 - “сделай, не знаю что, но чтобы было красиво”
 - научить модель сама извлекать нужную информацию из данных
- С подкреплением
 - “обучись ‘хорошему’ поведению с помощью кнута и пряника”
 - научить модель правильным действиям в различных ситуациях
- ...

Типичные задачи машинного обучения

- С учителем
 - классификация
 - регрессия
 - генерация
- Без учителя
 - кластеризация
 - автоматическое выявление взаимосвязей
- С подкреплением
 - управление
 - игры

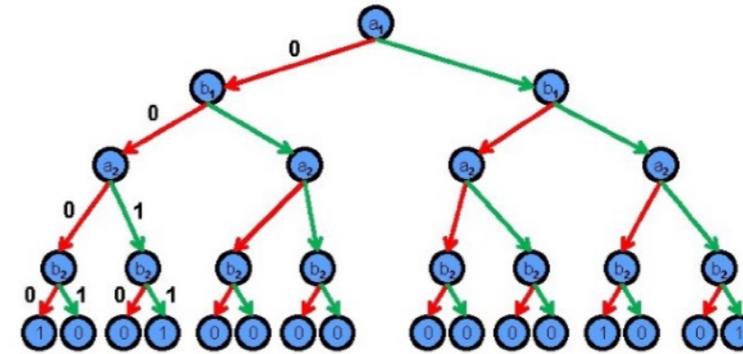
Задача классификации



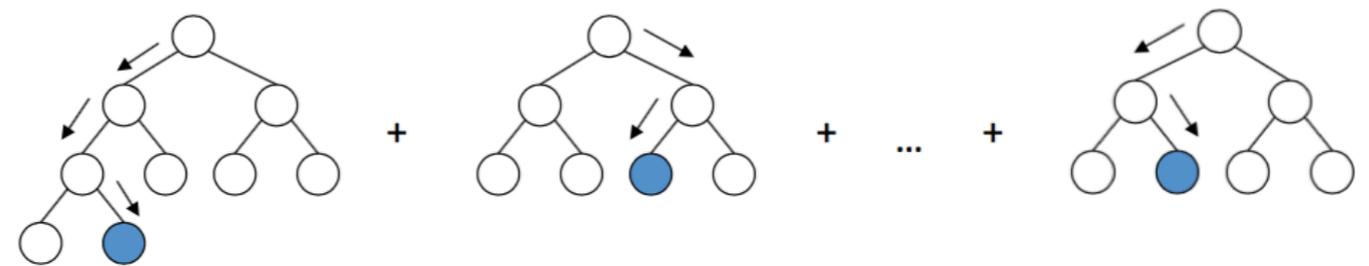
- Автоматический поиск наилучшей разделяющей поверхности в многопараметрическом подпространстве общих функций в многомерном пространстве характеристик объектов

Семейства универсальных параметризованных функций

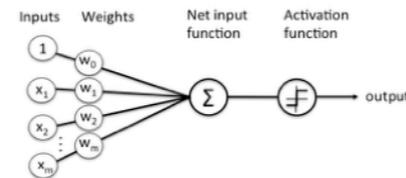
- деревья



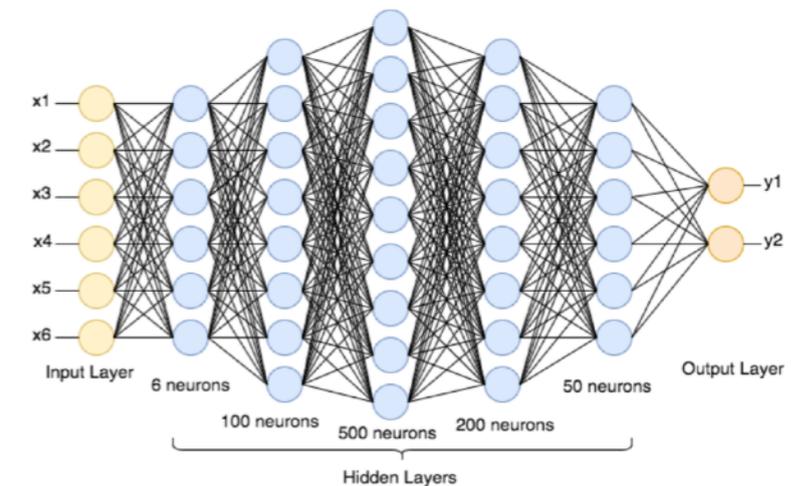
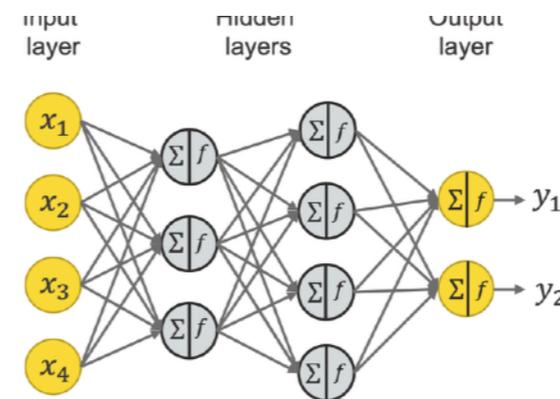
- ансамбли деревьев



- нейронные сети



- глубокие нейронные сети



Почему нейронные сети?

- Теорема Колмогорова — Арнольда (1957)
 - любая “достаточно хорошая” функция многих переменных представляется в виде композиций функций одной переменной и операции сложения

$$f(\mathbf{x}) = f(x_1, \dots, x_n) = \sum_{q=0}^{2n} \Phi_q \left(\sum_{p=1}^n \phi_{q,p}(x_p) \right)$$

- единственной фундаментальной функцией многих переменных является сложение

Почему нейронные сети?

УДК 519.7

Горбань А.Н. Обобщенная аппроксимационная теорема и вычислительные возможности нейронных сетей // Сиб. журн. вычисл. математики / РАН. Сиб. отд-ние. — Новосибирск, 1998. — Т. 1, № 1. — С. 11–24.

Исследуются вычислительные возможности искусственных нейронных сетей. В связи с этим намечались возврат к классической постановке вопроса о представлении функций многих переменных с помощью суперпозиций и сумм функций одного переменного и новая редакция этого вопроса (ограничение одной произвольно выбранной нелинейной функцией одного переменного).

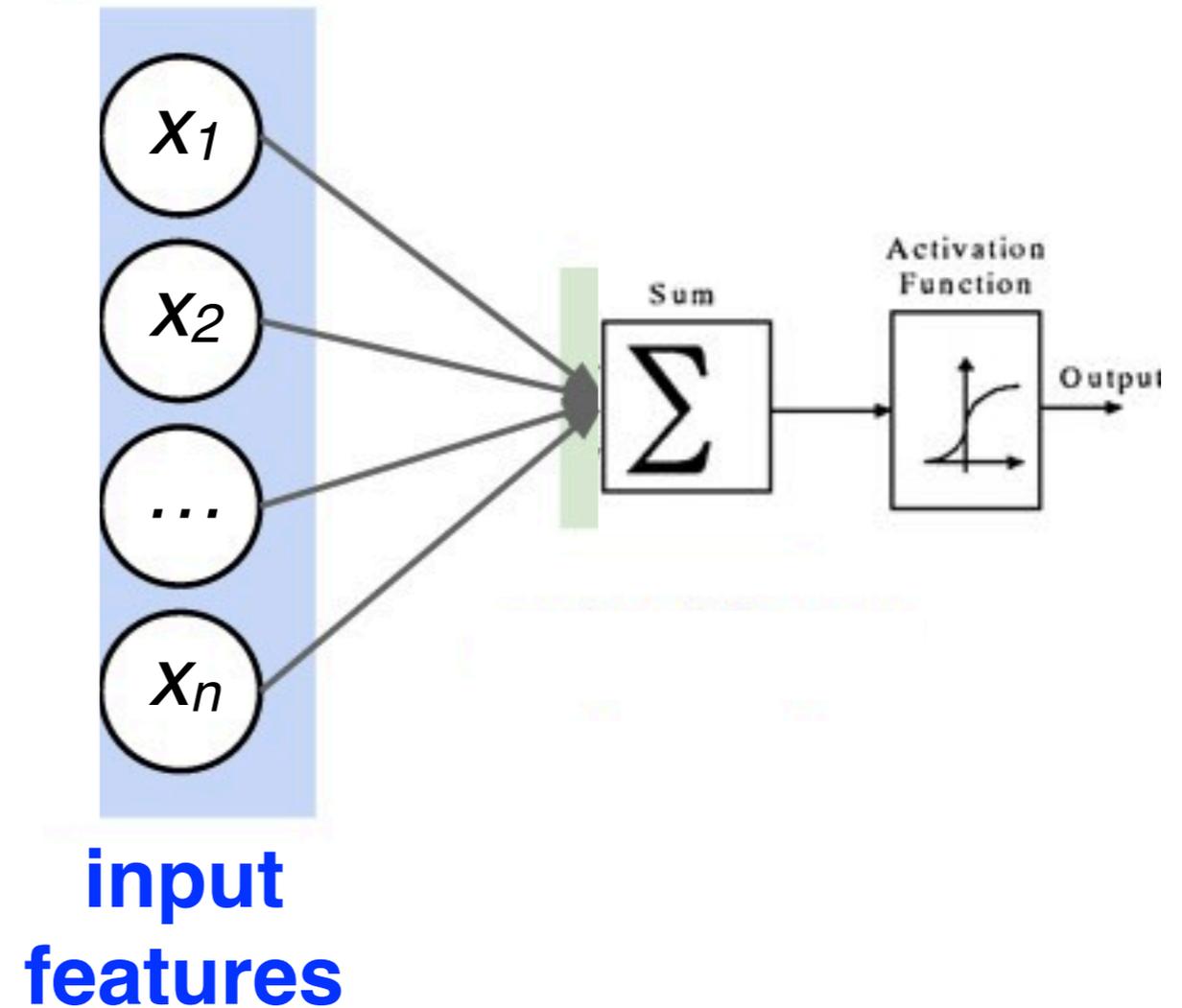
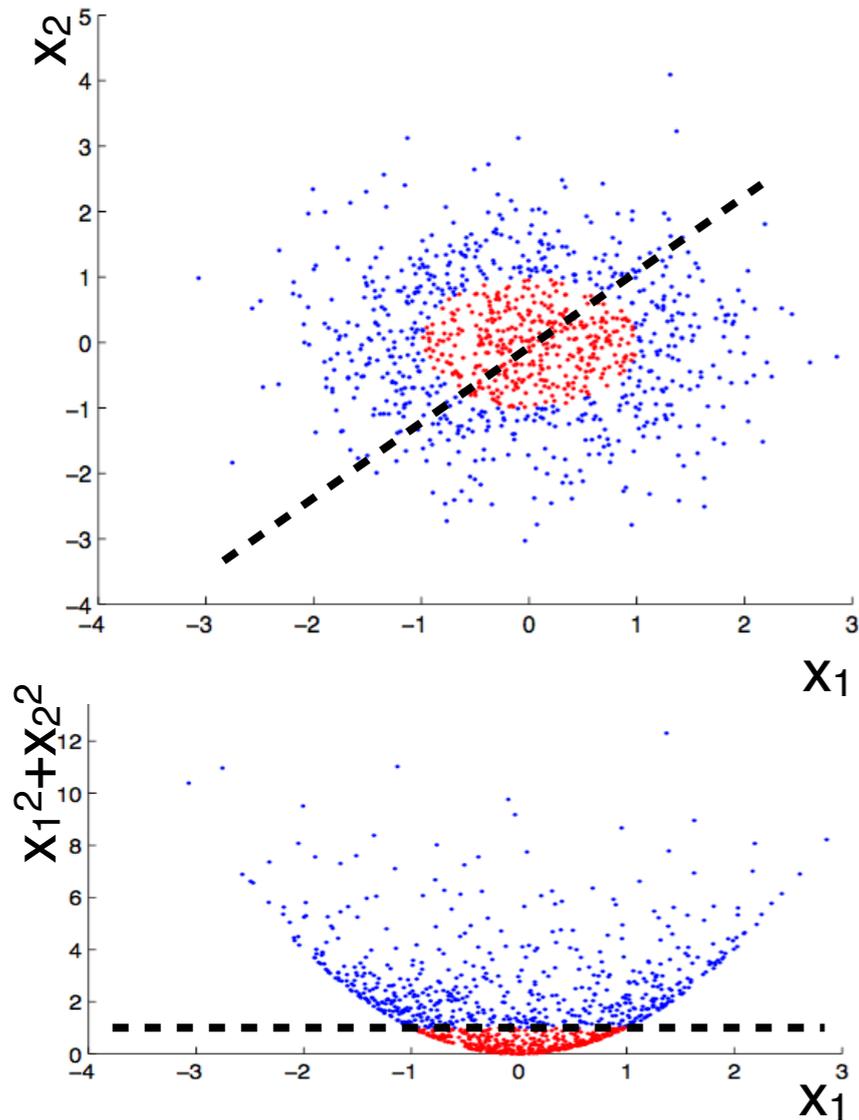
Показано, что можно получить сколь угодно точное приближение любой непрерывной функции многих переменных, используя операции сложения и умножения на число, суперпозицию функций, линейные функции, а также одну произвольную непрерывную нелинейную функцию одного переменного.

Для многочленов получен алгебраический вариант теоремы.

Для нейронных сетей полученные результаты означают, что от функции активации нейрона требуется только нелинейность.

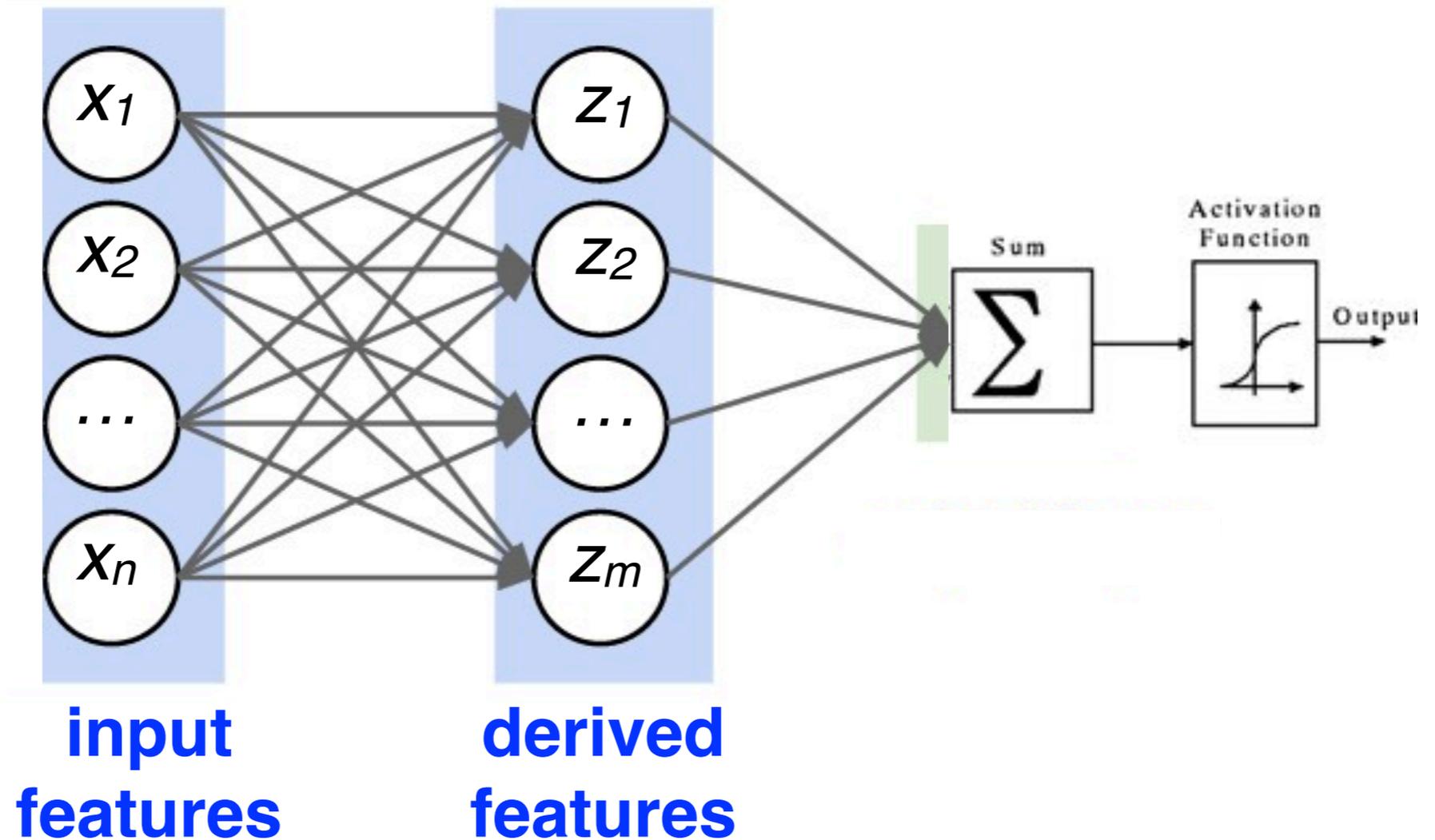
- достаточно одной нелинейной функции активации
- стандартная архитектура современных нейронных сетей
- требуемые глубина и ширина сети не определены

Логистическая регрессия



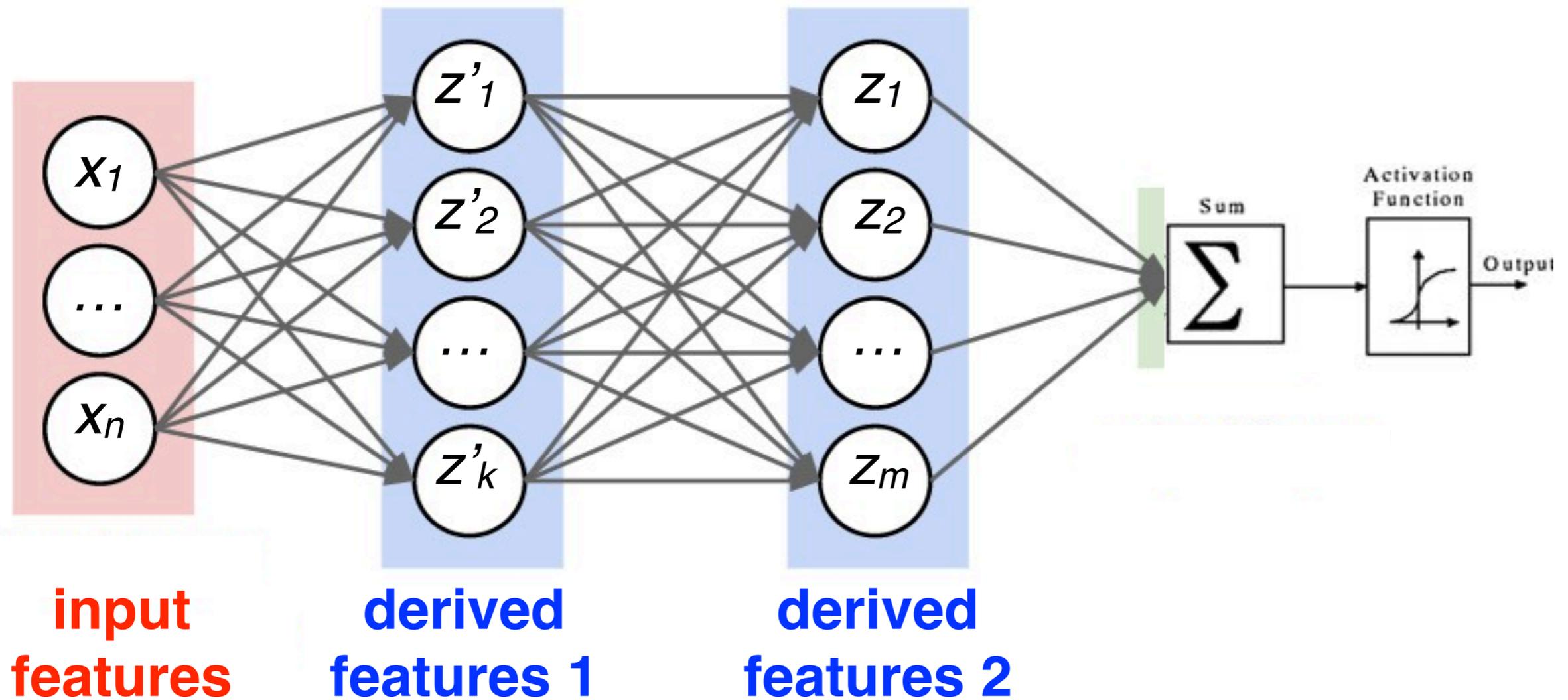
- **Линейная логистическая регрессия**
 - требует, чтобы разделяющая поверхность была плоской в пространстве исходных признаков

Послойное конструирование признаков



- Линейная логистическая регрессия
 - используем промежуточный слой, чтобы сконструировать признаки, подходящие для линейной логистической регрессии

Многослойное конструирование признаков



- Линейная логистическая регрессия
 - используем промежуточные слои, чтобы сконструировать признаки, подходящие для конструирования признаков, подходящих для линейной логистической регрессии

Практические применения МО



15 октября 2021, 07:01

Систему Face Pay для оплаты проезда запустили на всех станциях метро

Поделиться [f](#) [vk](#) [ok](#) [t](#) [g](#)



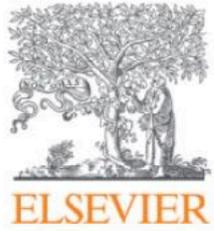
Типичные задачи в ФВЭ

- Физический анализ, оптимизация отделения сигнала от фона (MVA)
- Быстрый отбор событий в триггере
- Реконструкция в детекторах
- Идентификация частиц в детекторе
- Поиск аномалий
 - технические аномалии: качество данных
 - физические аномалии: поиск новой физики
- Ускорение МС генераторов
- Ускорение симуляции детекторов
- Оптимизация детекторов
- Управление ускорителем

Использование МО для задач ФВЭ

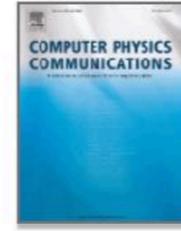
- Ключевые вопросы
 - метрика качества аппроксимации
 - семейство параметризованных аппроксимирующих функций $a(\theta, \mathbf{x})$
 - метод поиска оптимальной функции аппроксимации
 - $a_0(\mathbf{x}) = a(\theta_{best}, \mathbf{x})$
- Переиспользуем и адаптируем методики и подходы, разработанные сообществом МО

Первые применения в ФВЭ



Computer Physics Communications

Volume 49, Issue 3, June 1988, Pages 429-448



Bruce Denby, 1988



Neural networks and cellular automata in experimental high energy physics

B. Denby

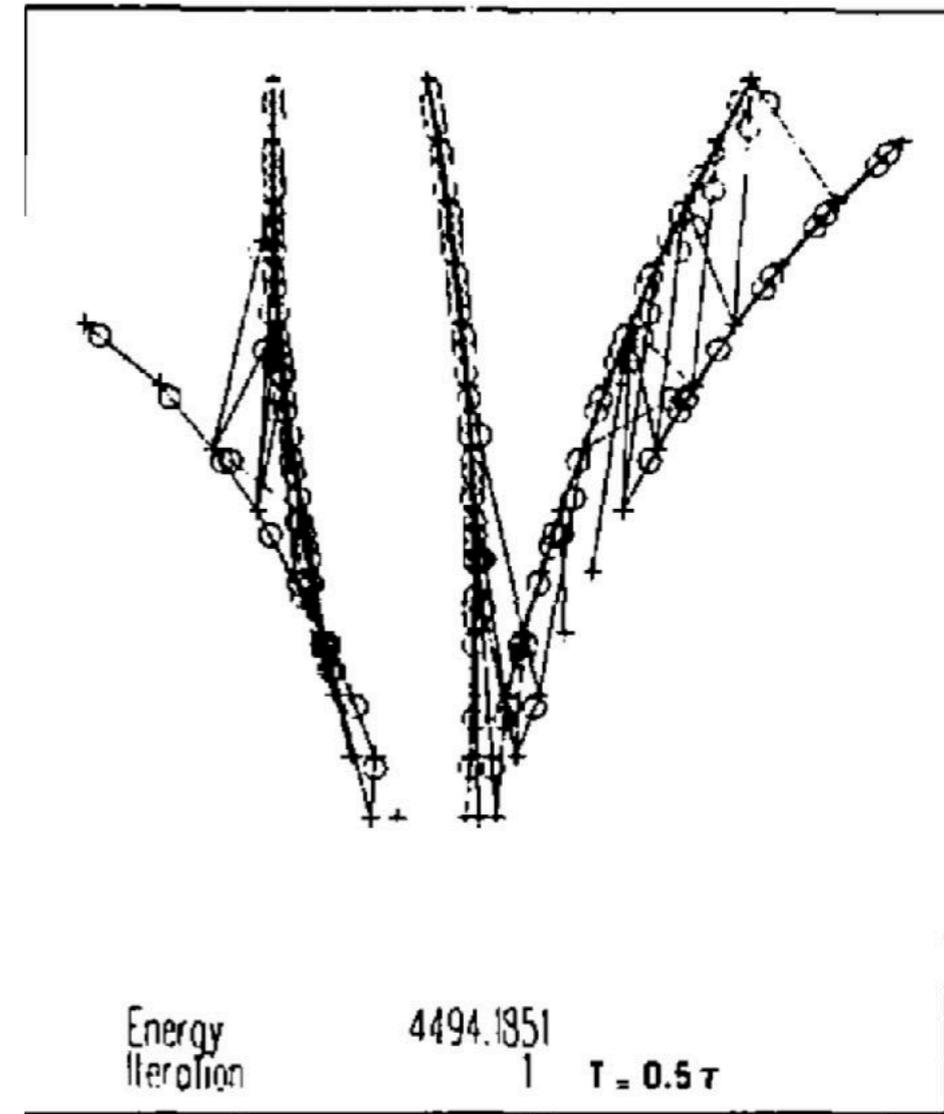
Show more

[https://doi.org/10.1016/0010-4655\(88\)90004-5](https://doi.org/10.1016/0010-4655(88)90004-5)

Get rights and content

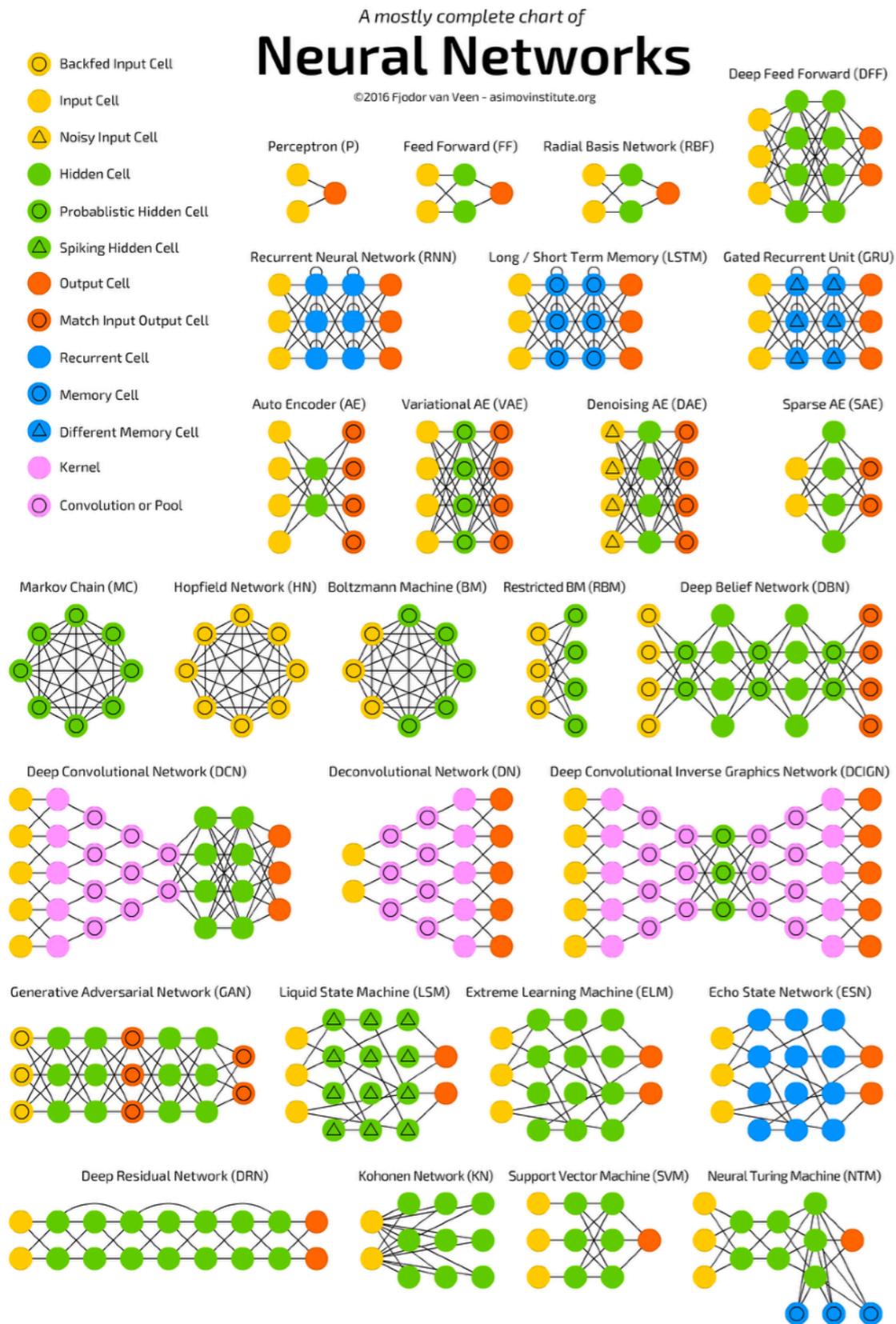
Abstract

Within the past few years, two novel computing techniques, cellular automata and neural networks, have shown considerable promise in the solution of problems of a very high degree of complexity, such as turbulent fluid flow, image processing, and pattern recognition. Many of the problems faced in experimental high energy physics are also of this



- Возможность использования МО в задачах поиска треков и кластеров в детекторах ФВЭ.

МО - универсальный подход



- собрать модель, адекватную задаче
- обучить стандартными методами
- профит!

12 месяцев в arXiv

Showing 1–50 of 51 results

hep-ex

Search v0.5.6 released 2020-02-24

Feedback

Query: order: -announced_date_first; size: 50; hide_abstracts: True; date_range: from 2020-12-01 to 2021-12-13; classification: Physics (grp_physics)::High Energy Physics - Experiment (hep-ex); include_cross_list: True; terms: AND title=machine learning

Refine query New search

50 results per page. Sort results by Announcement date (newest first) Go

1 2

Next

1. [arXiv:2112.05043](#) [pdf, other] [hep-ph](#) [hep-ex](#)
Reconstructing partonic kinematics at colliders with Machine Learning
Authors: David F. Rentería-Estrada, Roger J. Hernández-Pinto, German F. R. Sborlini, Pia Zurita
Submitted 9 December, 2021; originally announced December 2021.
Comments: 28 pages + appendices, 16 figures, 7 tables
Report number: DESY 21-211
2. [arXiv:2112.03769](#) [pdf, other] [hep-ph](#) [hep-ex](#) [physics.data-an](#) [stat.ML](#)
Machine Learning in the Search for New Fundamental Physics
Authors: Georgia Karagiorgi, Gregor Kasieczka, Scott Kravitz, Benjamin Nachman, David Shih
Submitted 7 December, 2021; originally announced December 2021.
Comments: Preprint of article submitted to Nature Reviews Physics, 19 pages, 1 figure
3. [arXiv:2112.01494](#) [pdf, other] [physics.ins-det](#) [hep-ex](#) [physics.data-an](#)
Improvement of the NOVA Near Detector Event Reconstruction and Primary Vertexing through the Application of Machine Learning Methods
Authors: Zakaria Elkarghli
Submitted 22 November, 2021; originally announced December 2021.
Comments: 43 pages
4. [arXiv:2111.14589](#) [pdf, other] [hep-ph](#) [hep-ex](#) [nucl-ex](#) [nucl-th](#)
The information content of jet quenching and machine learning assisted observable design
Authors: Yue Shi Lai, James Mulligan, Mateusz Płoskoń, Felix Ringer
Submitted 8 December, 2021; v1 submitted 29 November, 2021; originally announced November 2021.
Comments: Added data availability statement

Showing 1–24 of 24 results

hep-th

Search v0.5.6 released 2020-02-24

Feedback

Simple

Query: order: -announced_date_first; size: 50; hide_abstracts: True; date_range: from 2020-12-01 to 2021-12-13; classification: Physics (grp_physics)::High Energy Physics - Theory (hep-th); include_cross_list: True; terms: AND title=machine learning

Simple

Refine query New search

50 results per page. Sort results by Announcement date (newest first) Go

1. [arXiv:2112.06350](#) [pdf, other] [hep-th](#) [math.AG](#) [stat.ML](#)
Machine Learning Calabi-Yau Hypersurfaces
Authors: David S. Berman, Yang-Hui He, Edward Hirst
Submitted 12 December, 2021; originally announced December 2021.
Comments: 32 pages, 45 figures
2. [arXiv:2112.04950](#) [pdf, other] [hep-th](#)
Machine Learning to the Orientifold Calabi-Yau with String Vacua
Authors: Xin Gao, Hao Zou
Submitted 9 December, 2021; originally announced December 2021.
Comments: 26 pages, 4 figures, 2 tables
3. [arXiv:2112.00460](#) [pdf, other] [hep-lat](#) [cs.LG](#) [hep-ph](#) [hep-th](#) [nucl-th](#)
Machine learning Hadron Spectral Functions in Lattice QCD
Authors: Shi-Yang Chen, Heng-Tong Ding, Fei-Yi Liu, Gabor Papp, Chun-Bin Yang
Submitted 1 December, 2021; originally announced December 2021.
Comments: 9 pages, 7 figures. Talk presented at the 38th International Symposium on Lattice Field Theory (Lattice 2021), 26-30 July, 2021, Zoom/Gather@Massachusetts Institute of Technology
4. [arXiv:2111.05216](#) [pdf, other] [hep-lat](#) [cond-mat.stat-mech](#) [hep-th](#)
Machine learning approaches to the QCD transition
Authors: Andrea Palermo, Lucio Anderlini, Maria Paola Lombardo, Andrey Kotov, Anton Trunin
Submitted 9 November, 2021; originally announced November 2021.
Comments: Proceedings of the 38th international symposium on Lattice Field Theory, LATTICE2021
5. [arXiv:2110.13521](#) [pdf, other] [hep-lat](#) [cs.LG](#) [hep-ph](#) [hep-th](#) [nucl-th](#)
Machine learning spectral functions in lattice QCD
Authors: S. -Y. Chen, H. -T. Ding, F. -Y. Liu, G. Papp, C. -B. Yang
Submitted 26 October, 2021; originally announced October 2021.



12 месяцев в arXiv

Showing 1-50 of 51 results **hep-ex** — Showing 1-24 of 24 results **hep-th**

Search v0.5.6 released 2020-02-24 Feedback Simple Search

Query: order: -announced_date_first; size: 50; 2020-12-01 to 2021-12-13; classification: Physics Experiment (hep-ex); include_cross_list: True;

Refine query New search

50 results per page. Sort

Showing 1-18 of 18 results **CS** Search v0.5.6 released 2020-02-24 Feedback Simple Search

Query: order: -announced_date_first; size: 50; hide_abstracts: True; date_range: from 2020-12-01 to 2021-12-13; classification: Computer Science (cs); include_cross_list: True; terms: AND title=high energy physics

Refine query New search

50 results per page. Sort results by Announcement date (newest first) Go

1. [arXiv:2112.05043](#) [pdf, other] **hep-ph** cs.LG
Function Approximation for High-Energy Physics: Comparing Machine Learning and Interpolation Methods
Authors: Ibrahim Chahrour, James D. Wells
Submitted 8 December, 2021; v1 submitted 29 November, 2021; originally announced November 2021.
Comments: 30 pages, 17 figures, added a few references
2. [arXiv:2107.07304](#) [pdf, ps, other] **cs.DB** hep-ex doi 10.1051/epjconf/202125102066
Exploring Object Stores for High-Energy Physics Data Storage
Authors: Javier López-Gómez, Jakob Blomer
Submitted 15 July, 2021; originally announced July 2021.
Comments: Accepted for Proceedings of 25th International Conference on Computing in High-Energy and Nuclear Physics
3. [arXiv:2107.02908](#) [pdf, other] **hep-ph** cs.LG hep-ex
Particle Convolution for High Energy Physics
Authors: Chase Shimmin
Submitted 5 July, 2021; originally announced July 2021.
Comments: To be presented at ML4Jets 2021
4. [arXiv:2105.08960](#) [pdf, other] **hep-ex** cs.LG doi 10.1051/epjconf/202125103042
Physics Validation of Novel Convolutional 2D Architectures for Speeding Up High Energy Physics Simulations
Authors: Florian Rehm, Sofia Vallecorsa, Kerstin Borrás, Dirk Krücker
Submitted 19 May, 2021; originally announced May 2021.
Comments: Paper published at vCHEP2021 conference

1. [arXiv:2112.05043](#) [pdf, other] **hep-ph**
Reconstructing partonic kinematics from detector observables
Authors: David F. Rentería-Estrada, ...
Submitted 9 December, 2021; originally announced December 2021.
Comments: 28 pages + appendices, 16 figures
Report number: DESY 21-211

2. [arXiv:2112.03769](#) [pdf, other] **cs.LG**
Machine Learning in the Search for New Physics
Authors: Georgia Karagiorgi, Gregor ...
Submitted 7 December, 2021; originally announced December 2021.
Comments: Preprint of article submitted to ...

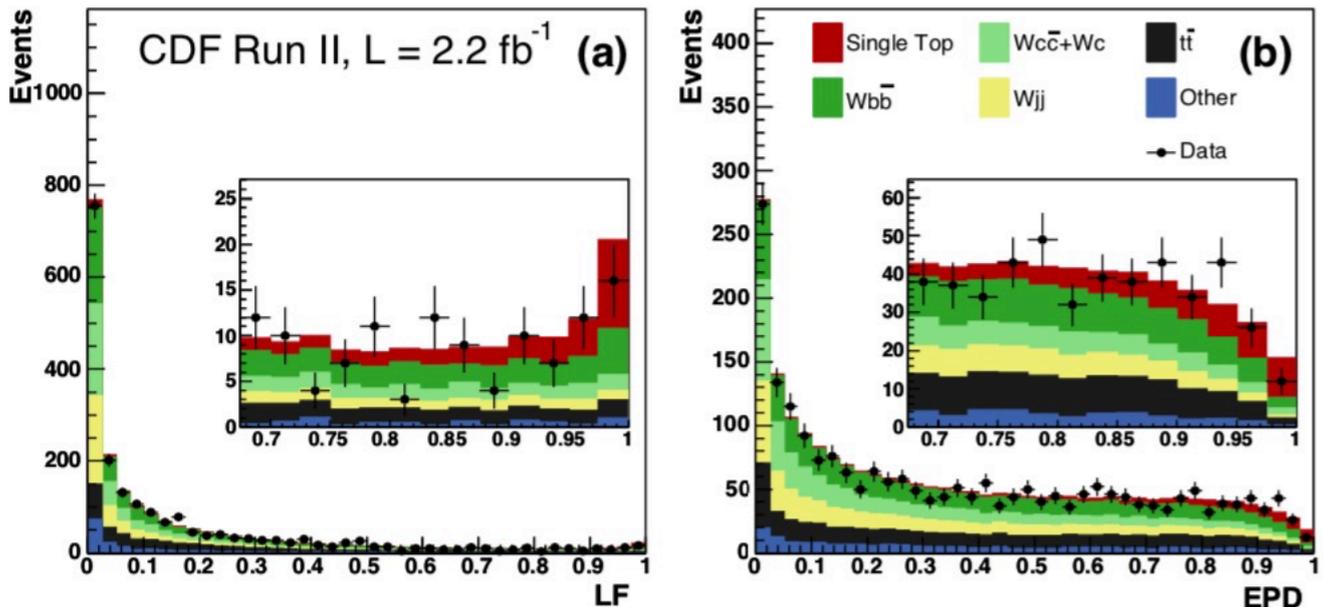
3. [arXiv:2112.01494](#) [pdf, other] **hep-ph**
Improvement of the NOVA through the Application of Machine Learning
Authors: Zakaria Elkarghli
Submitted 22 November, 2021; originally announced November 2021.
Comments: 43 pages

4. [arXiv:2111.14589](#) [pdf, other] **hep-ph**
The information content of particle detector design
Authors: Yue Shi Lai, James Mulligan
Submitted 8 December, 2021; v1 submitted December 2021.
Comments: Added data availability statement

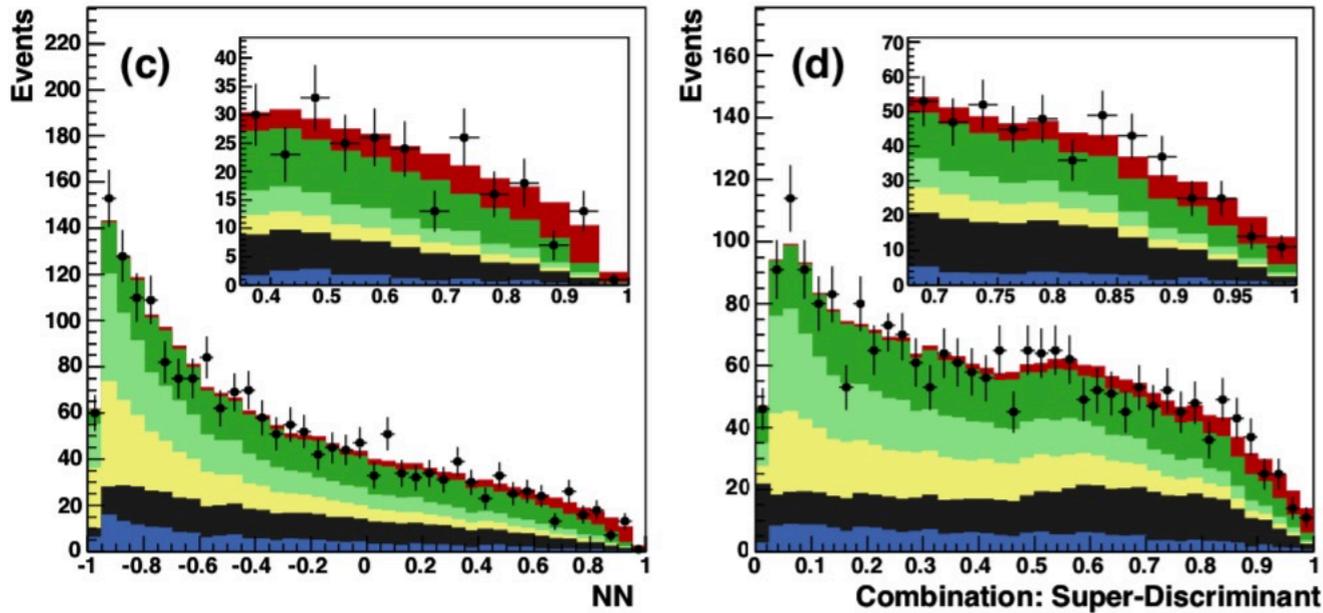
Vacua
nucl-th
D
Yang
attice Field Theory (Lattice 2021), 26-30 July.
otov, Anton Trunin
TICE2021
nucl-th



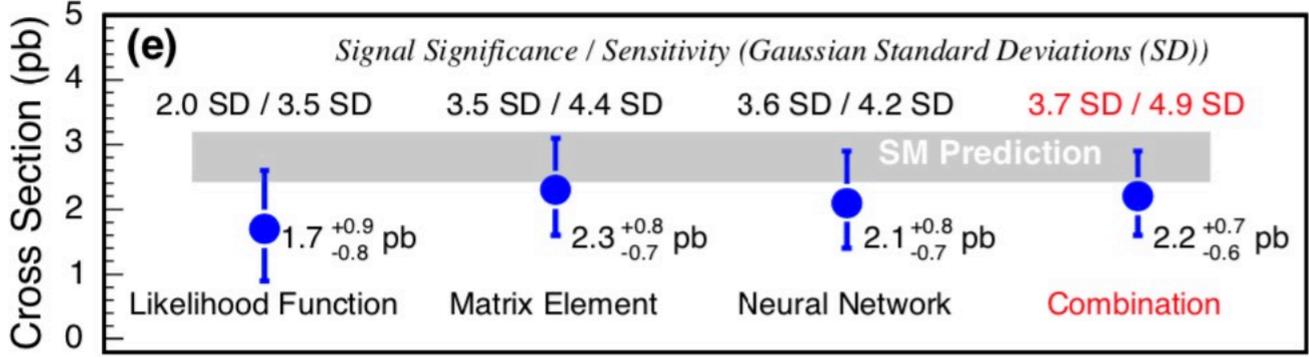
Открытие рождения одиночного t-кварка (2008)



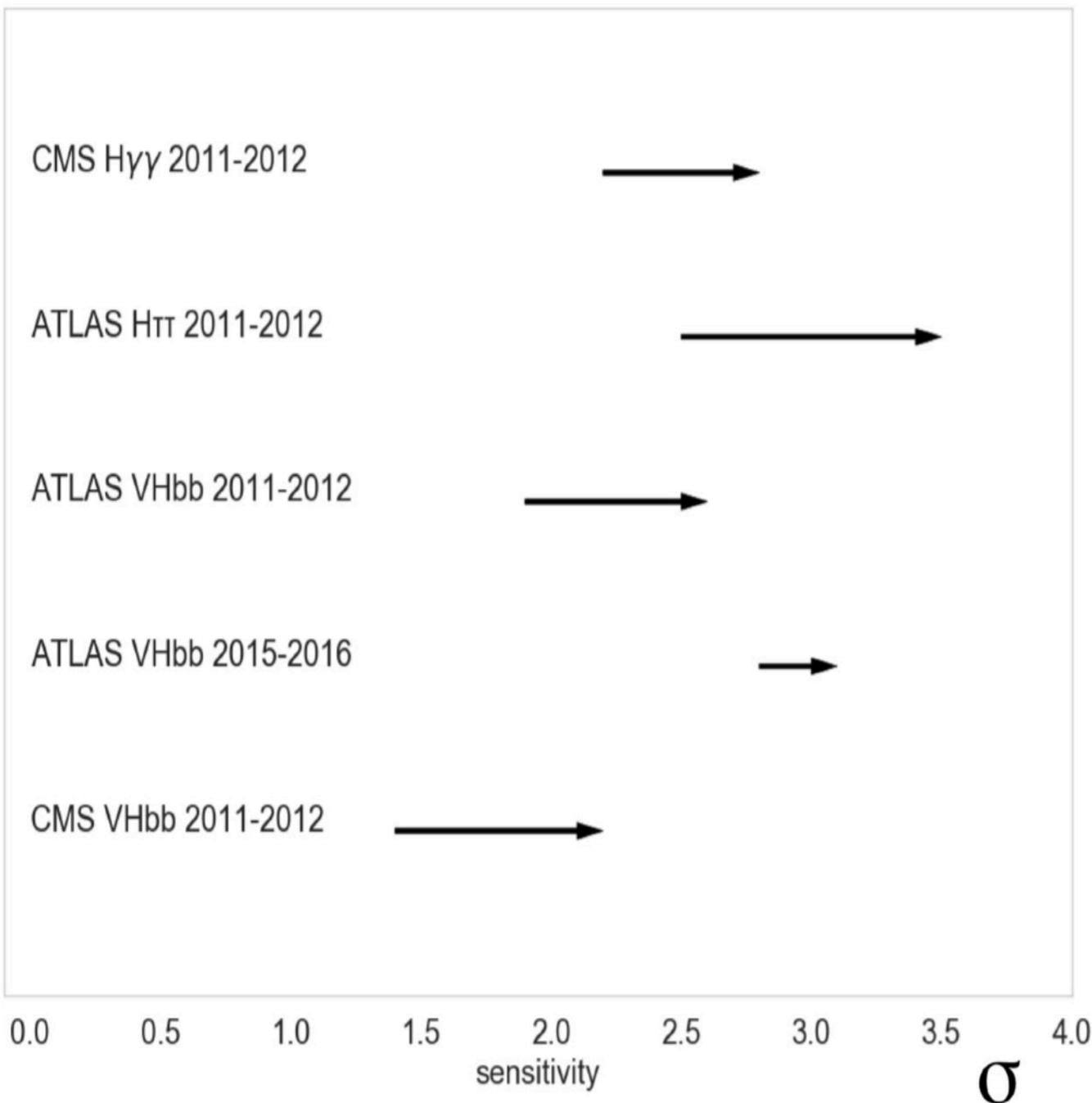
arXiv:0809.2581



- Проявление эффекта в распределении выхода классификатора, а не физического параметра



Оптимизация анализа данных. LHC



- Активное использование MVA в экспериментах LHC
 - ~ 50% прирост значимости для измерений бозона Хиггса
- В 2020 трудно найти физический анализ без использования MVA для разделения сигнала и фона
 - типична совместная оптимизация ~10 критериев физического отбора

Произвольная свежая статья LHCb

EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH (CERN)



CERN-EP-2022-097
LHCb-PAPER-2022-008
July 8, 2022

Search for the lepton-flavour violating decays $B^0 \rightarrow K^{*0} \mu^\pm e^\mp$ and $B_s^0 \rightarrow \phi \mu^\pm e^\mp$

LHCb collaboration

Abstract

A search for the lepton-flavour violating decays $B^0 \rightarrow K^{*0} \mu^\pm e^\mp$ and $B_s^0 \rightarrow \phi \mu^\pm e^\mp$ is presented, using proton-proton collision data collected by the LHCb detector at the LHC, corresponding to an integrated luminosity of 9 fb^{-1} . No significant signals are observed and upper limits of

$$\begin{aligned}\mathcal{B}(B^0 \rightarrow K^{*0} \mu^+ e^-) &< 5.7 \times 10^{-9} \quad (6.9 \times 10^{-9}), \\ \mathcal{B}(B^0 \rightarrow K^{*0} \mu^- e^+) &< 6.8 \times 10^{-9} \quad (7.9 \times 10^{-9}), \\ \mathcal{B}(B^0 \rightarrow K^{*0} \mu^\pm e^\mp) &< 10.1 \times 10^{-9} \quad (11.7 \times 10^{-9}), \\ \mathcal{B}(B_s^0 \rightarrow \phi \mu^\pm e^\mp) &< 16.0 \times 10^{-9} \quad (19.8 \times 10^{-9})\end{aligned}$$

are set at 90% (95%) confidence level. These results constitute the world's most stringent limits to date, with the limit on the decay $B_s^0 \rightarrow \phi \mu^\pm e^\mp$ the first being set. In addition, limits are reported for scalar and left-handed lepton-flavour violating New Physics scenarios.

arXiv:2207.04005v1 [hep-ex] 8 Jul 2022

Произвольная свежая статья LHCb

EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH (CERN)



hadronic $B^0 \rightarrow K^* \pi^+ \pi^-$ and $B_s^0 \rightarrow \phi \pi^+ \pi^-$ decays (with misidentification of the pions as leptons).

Background from combinations of random tracks (combinatorial background) is reduced using a boosted decision tree (BDT) [50] classifier trained with the AdaBoost algorithm [51] as implemented in the TMVA software package [52]. The BDT classifier is trained separately for the two signal decays using calibrated simulation as signal. Data from the $B_{(s)}^0$ upper mass sideband region [5600, 6700] MeV/ c^2 are used as a proxy for the background. The classifier is trained using a k -folding approach and its performance is verified using standard cross-validation techniques [53]. The classifier uses the (transverse) momentum of the $B_{(s)}^0$ candidate, its vertex fit quality and flight distance significance, the angle between the $B_{(s)}^0$ momentum and the vector connecting the associated PV and the $B_{(s)}^0$ decay vertex, and the χ_{IP}^2 of the $B_{(s)}^0$ candidate and the final-state particles. The selection criterion on the classifier output is chosen according to the Punzi figure of merit $\varepsilon_{\text{sig}} / (3/2 + \sqrt{N_{\text{comb}}})$ [54]. Here, ε_{sig} denotes the signal efficiency and N_{comb} the expected combinatorial background yield, which is extrapolated from the upper mass sideband using the reconstructed same-sign lepton samples $K^+ \pi^- \mu^\pm e^\pm$ and $K^+ K^- \mu^\pm e^\pm$. Relative to the previously described selection criteria, the BDT requirement results in a signal efficiency of 55–80%, depending on the signal mode, and a rejection for combinatorial background of larger than 99%.

The normalisation modes, $B^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-) K^{*0}$ and $B_s^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-) \phi$, are reconstructed and selected in a way that is as similar as possible to the corresponding

arXiv:2207.04005v1 [hep-ex] 8 Jul 2022

vio

A
is
th
ar

ar
st
In
N

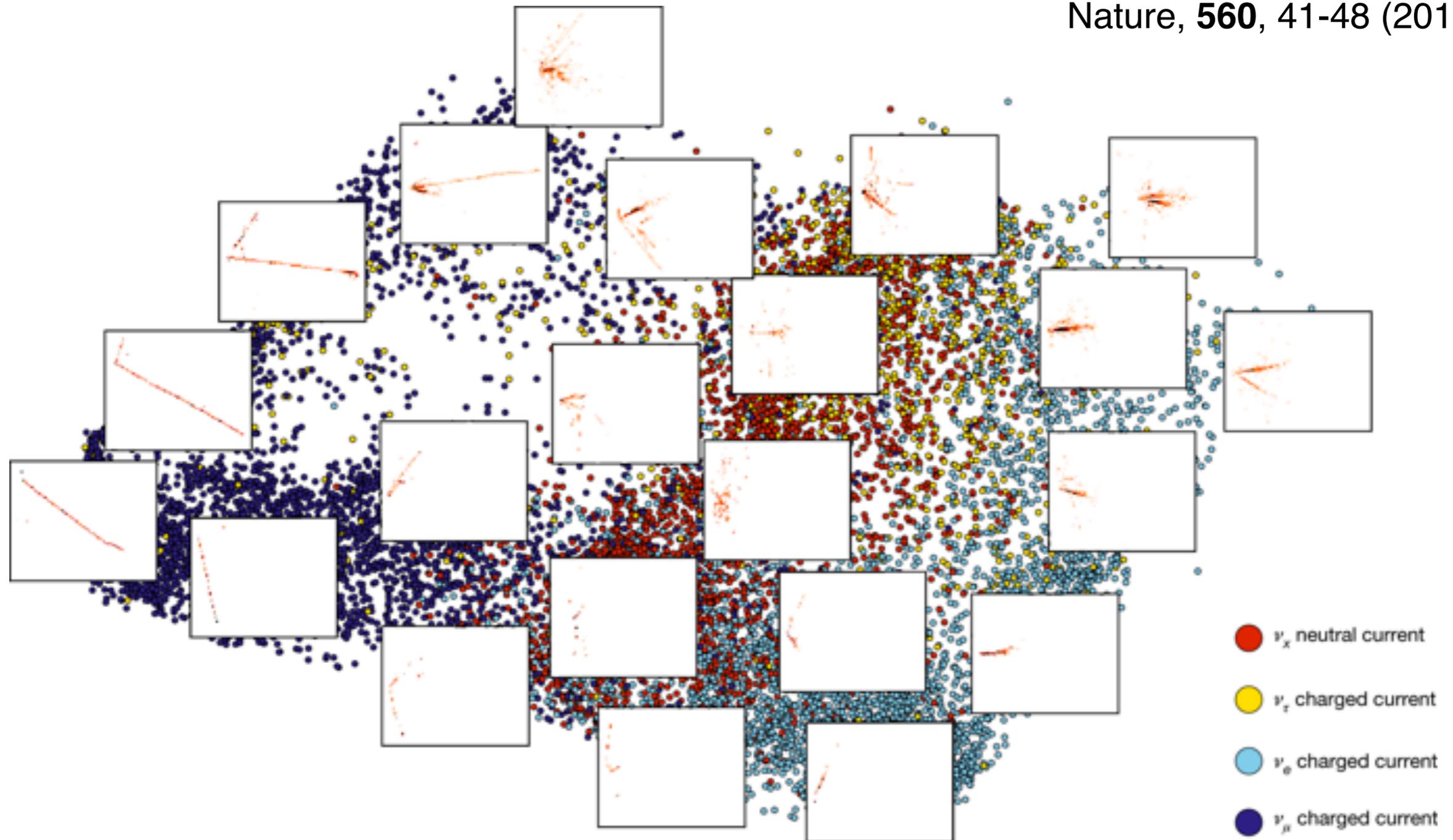


Данные для обучения

- Калибровочные данные, набранные в эксперименте
 - наиболее реалистичное представление
 - содержат фоновые вклады
 - покрывают не весь фазовый объем анализируемых данных
- MC данные
 - чистое представление, полностью под контролем
 - покрывают все фазовое пространство
 - нефизические различия между MC и реальными данными может быть использовано обученным алгоритмом для дискриминации
 - требует специального внимания

Классификация распадов

Nature, **560**, 41-48 (2018)



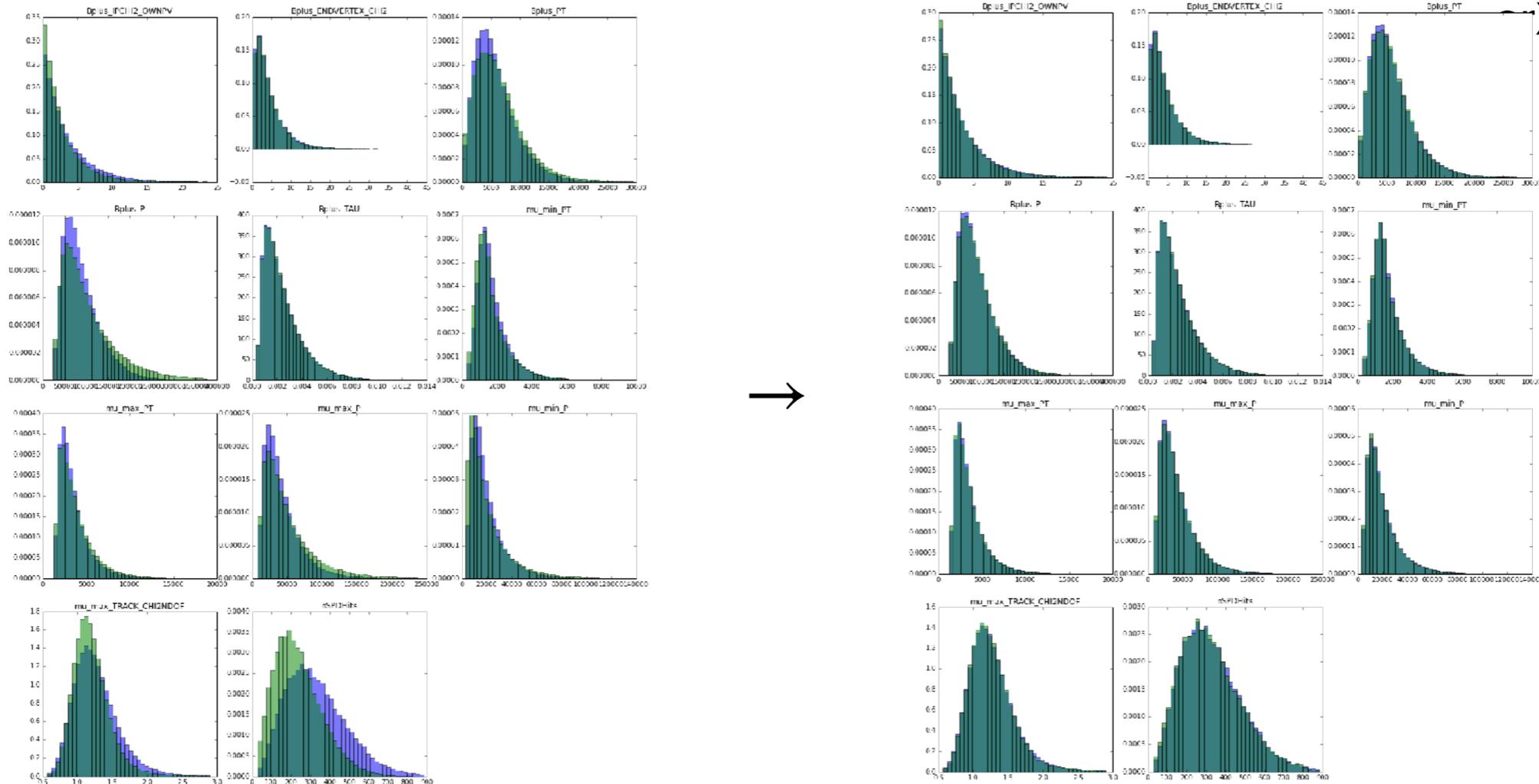
- Нейтринный эксперимент NOvA
- классификация разумно разбивает данные на домены

Неопределенности

- Проистекающие от специфических свойств обучаемого алгоритма (недо/переобучение, ограниченный класс функций, ...)
 - алгоритмы МО универсальны
 - различные представления алгоритмов слабо коррелированы
 - сравнение результатов обученных алгоритмов различных архитектур дает представление о систематиках, привнесенных алгоритмом
- Проистекающие от специфических свойств обучающих данных
 - валидация на чистых (пусть локальных) данных
 - $J/\psi \rightarrow \mu^+ \mu^-$
 - подмешивание сигнала к реальным данным
 - замена μ на e в событии

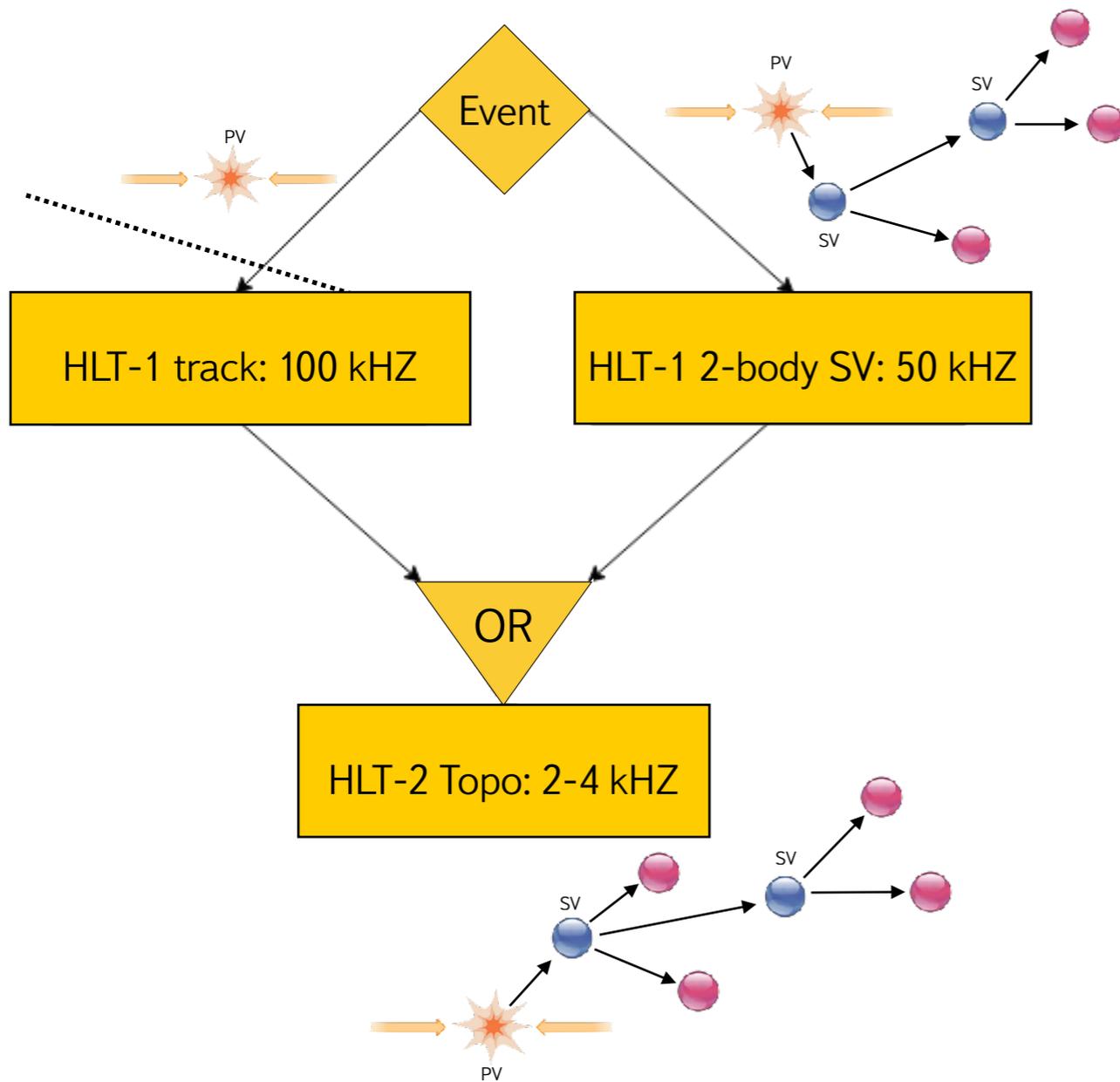
Перевзвешивание МС к данным

Xiv:1608.05806



- Задача: перевзвесить события симуляции так, чтобы воспроизводились определённые распределения в данных
 - без проблем для 1D
 - непросто для больших размерностей и ограниченной статистики
- Естественная задача для МО: перевзвесить так, чтобы распределения стали неразличимыми для классификатора

Классификация в триггерах

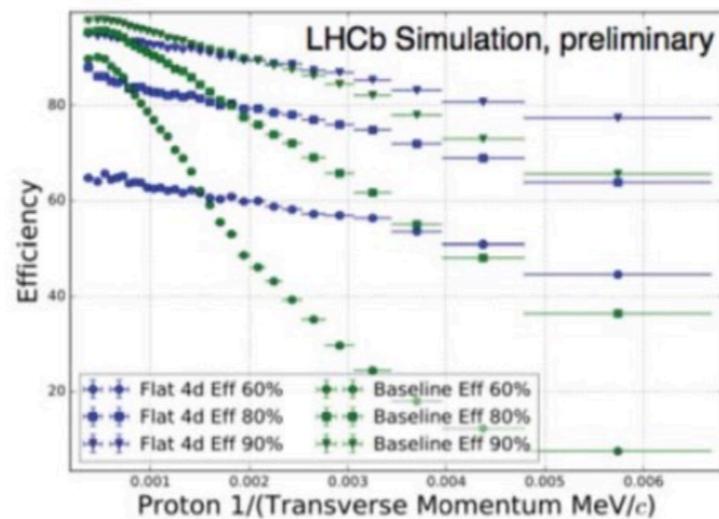
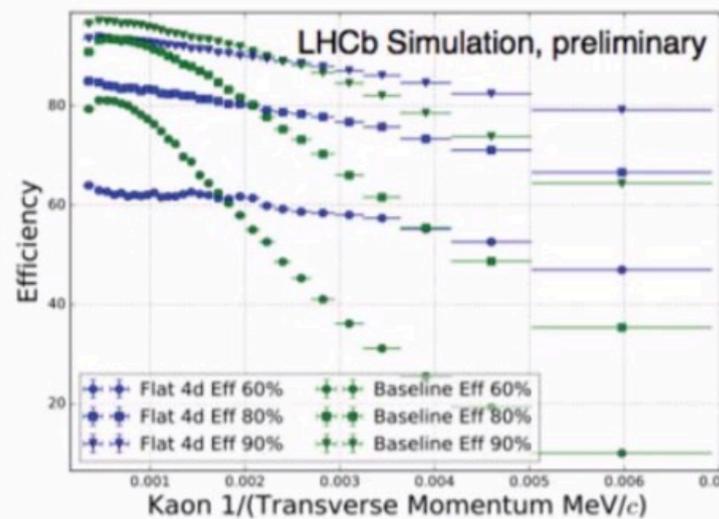
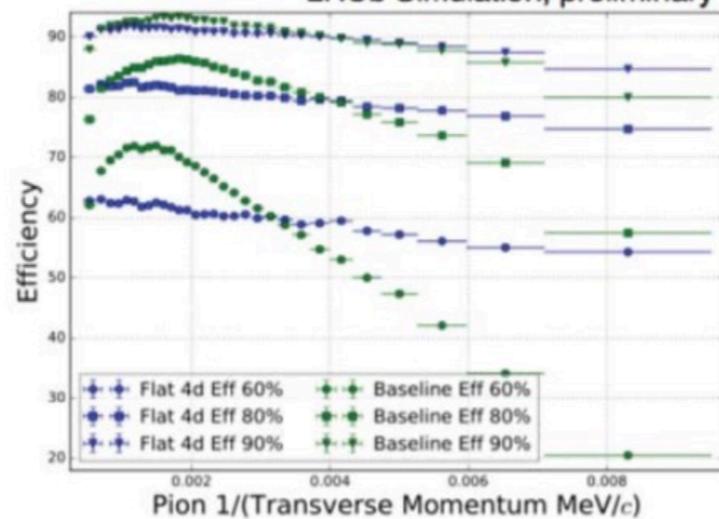
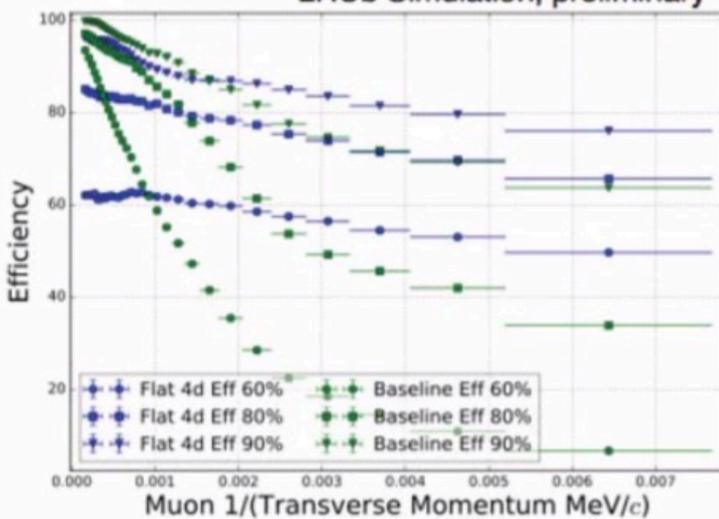
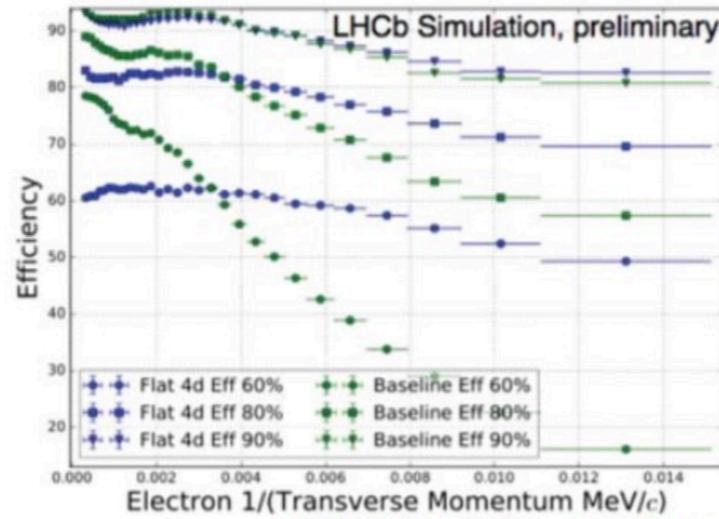
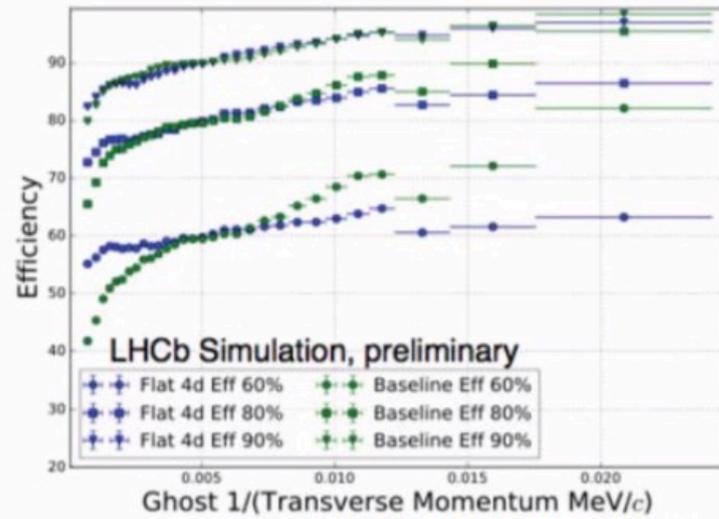


arXiv:1510.00572

mode	2.5 kHz	4. kHz
$B^0 \rightarrow K^*[K^+\pi^-]\mu^+\mu^-$	1.64	1.72
$B^+ \rightarrow \pi^+K^-K^+$	1.59	1.65
$B_s^0 \rightarrow D_s^-[K^+K^-\pi^-]\mu^+\nu_\mu$	1.14	1.47
$B_s^0 \rightarrow \psi(1S)[\mu^+\mu^-]K^+K^-\pi^+\pi^-$	1.62	1.71
$B_s^0 \rightarrow D_s^-[K^+K^-\pi^-]\pi^+$	1.46	1.52
$B^0 \rightarrow D^+[K^-\pi^+\pi^+]D^-[K^+\pi^-\pi^-]$	1.40	1.86

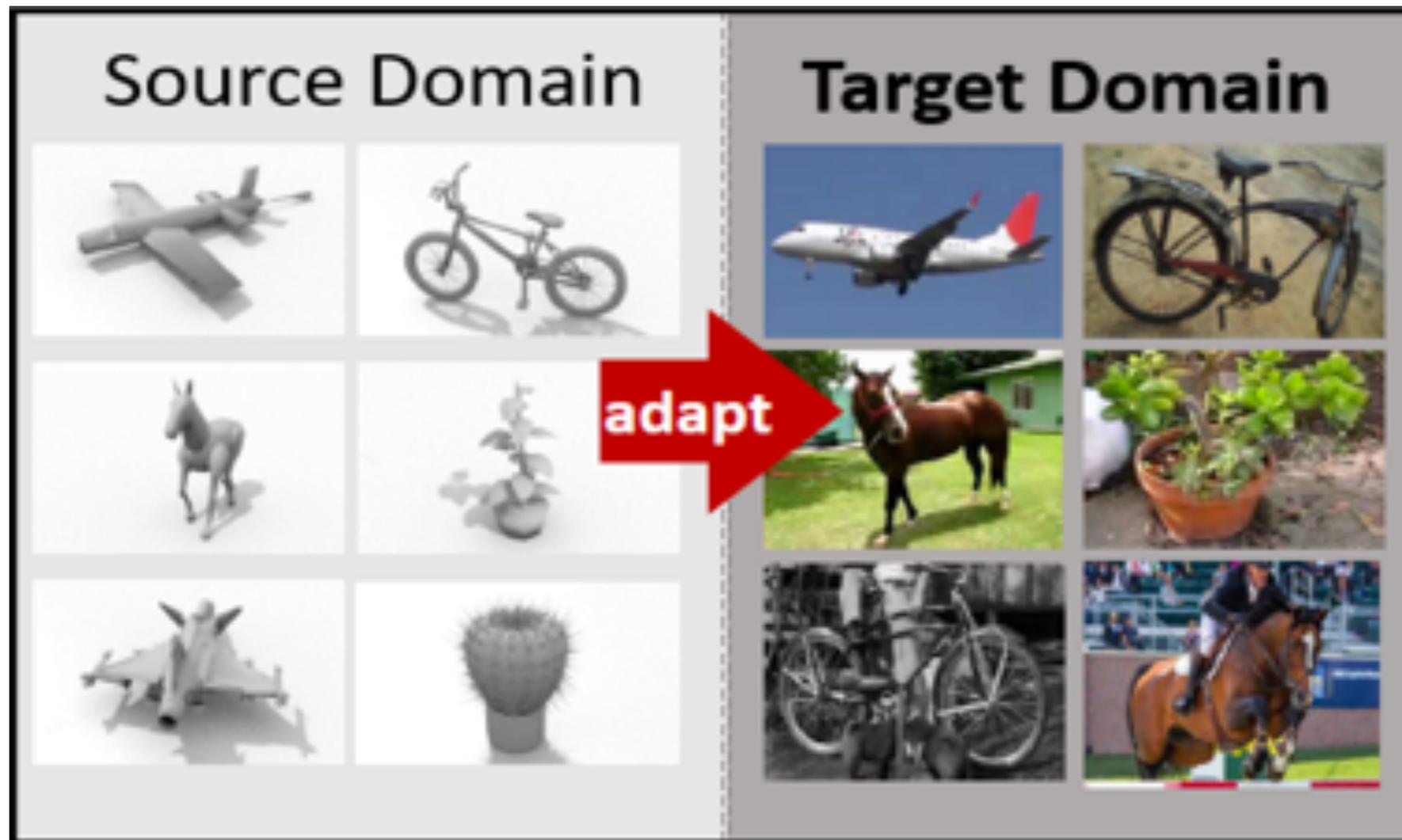
- 70% данных LHCb отбирается основанными на применении МО триггерами
- сбалансированное обучение позволяет значительно повысить эффективность обученных алгоритмов

Наложение специфических требований



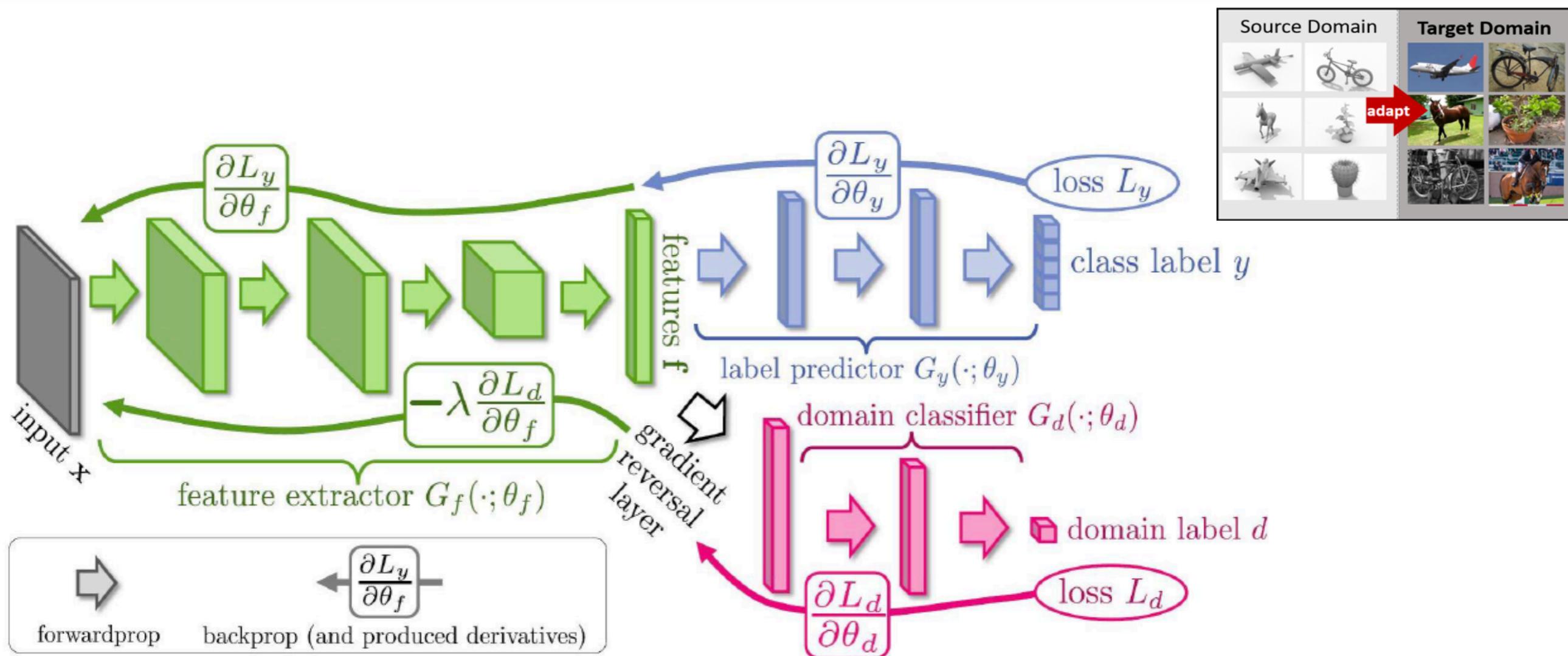
- Пожелание иметь плоскую эффективность идентификации

Доменная адаптация



- Сделать алгоритм зависящим от существенных свойств объектов, но нечувствительным к деталям обучающего набора данных

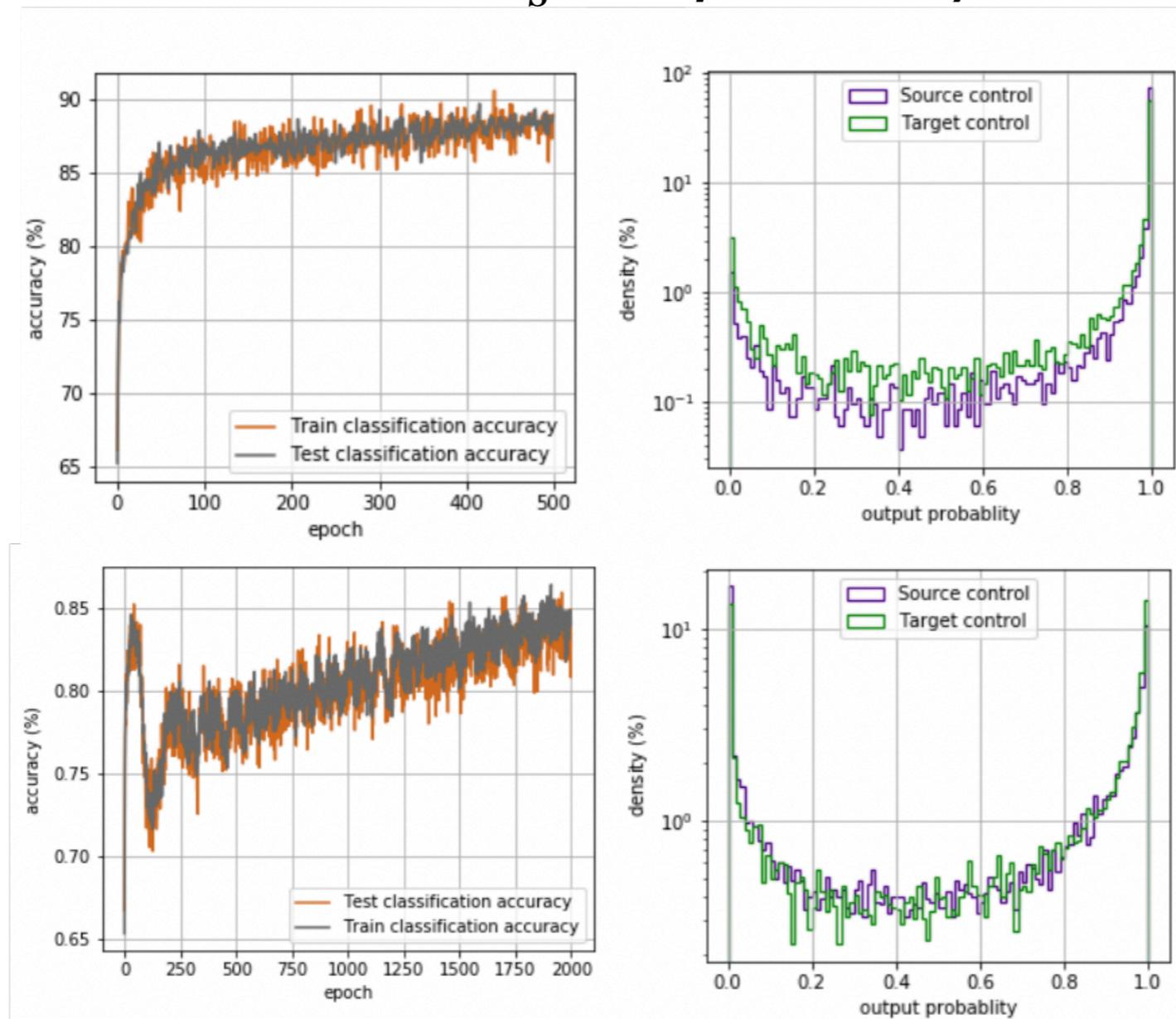
Доменная адаптация



- Сделать алгоритм зависящим от существенных свойств объектов, но нечувствительным к деталям обучающего набора данных

Доменная адаптация в ФВЭ

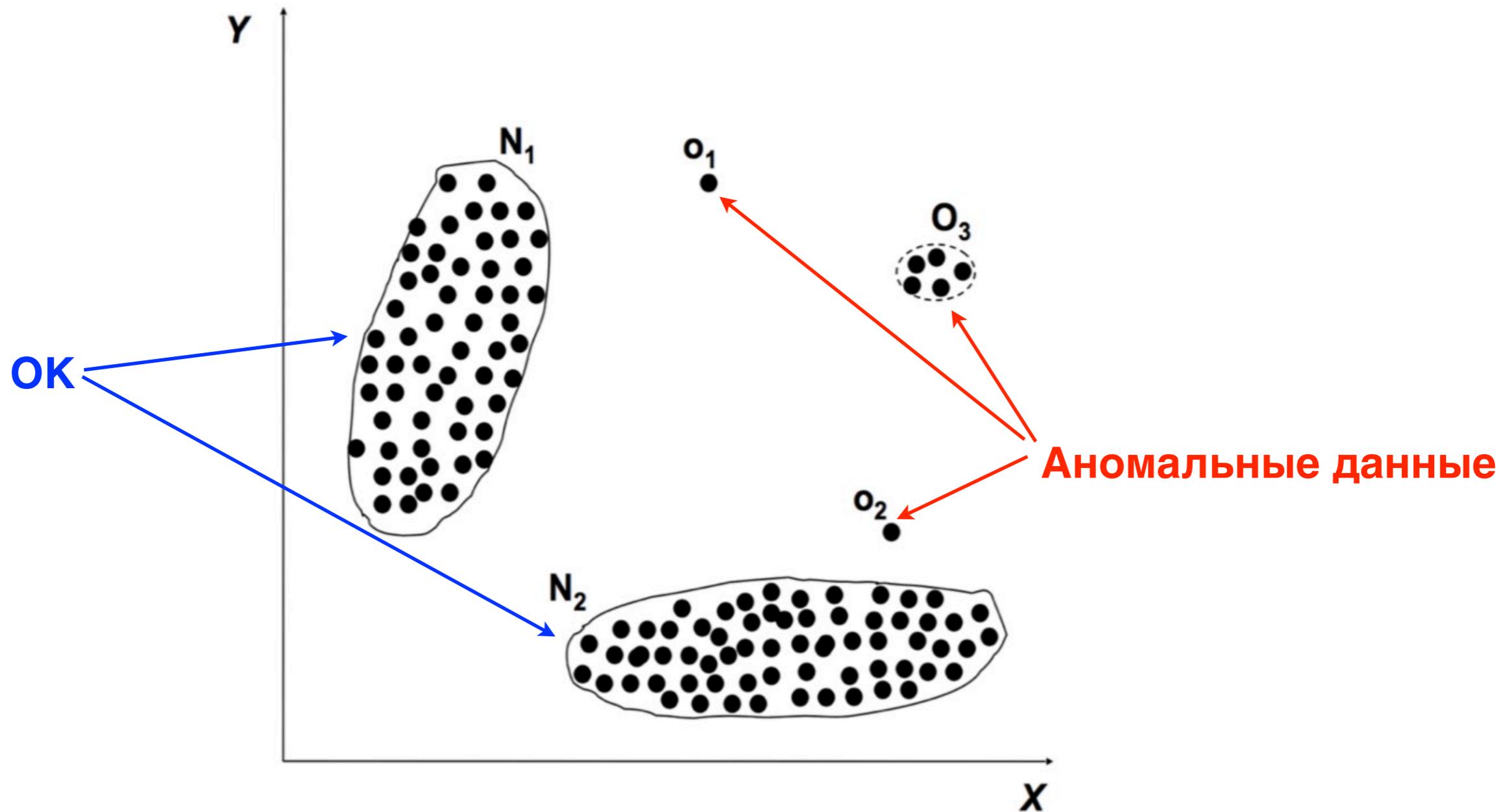
- сигнал: $\tau \rightarrow 3\mu$ МС arXiv:1912.08001
- фон: " τ " $\rightarrow 3\mu$ реальные данные вне массы τ
- контрольные данные: $D_s \rightarrow \phi\pi \rightarrow 3\mu$



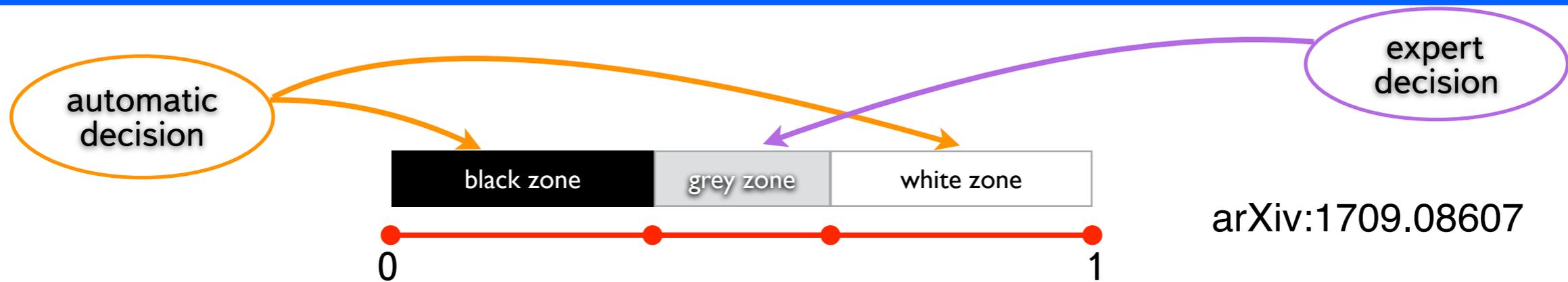
← без ДА

← с ДА

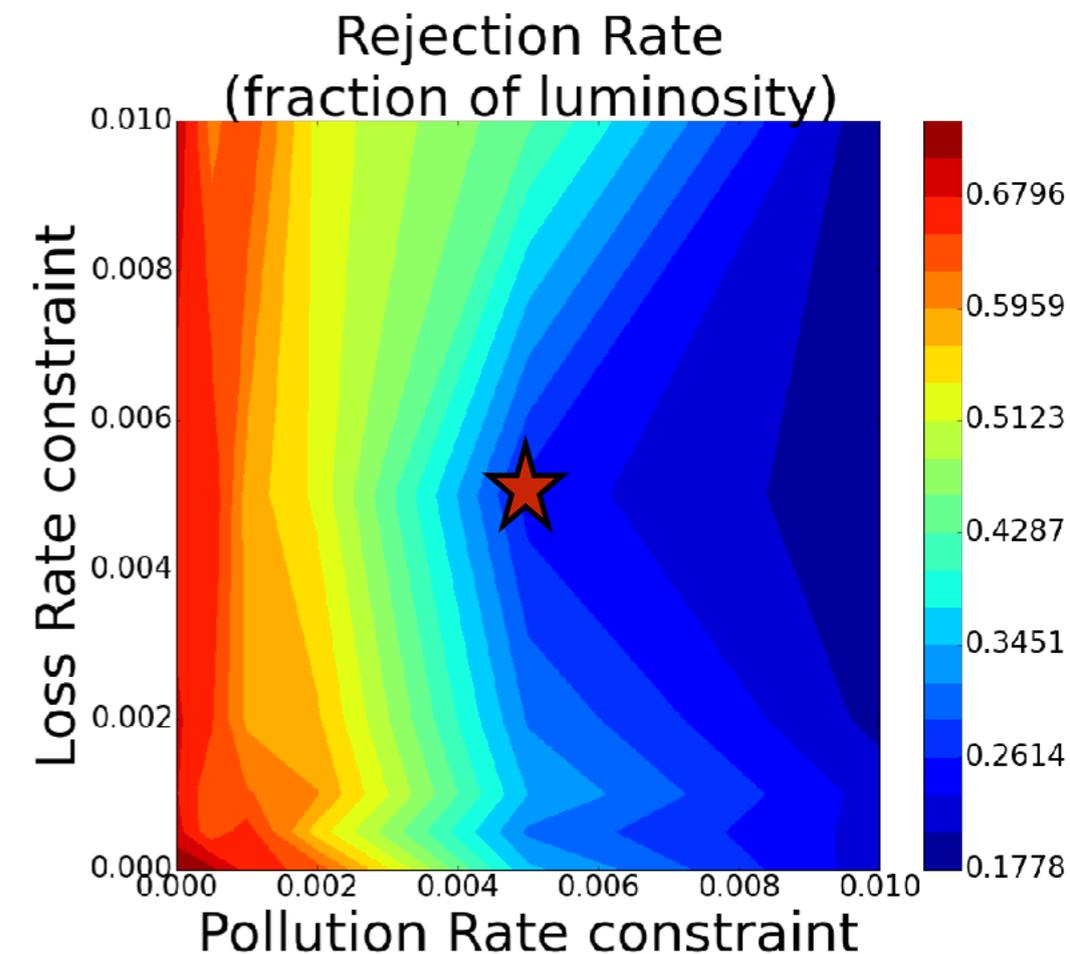
Поиск аномалий в данных



Поиск технических аномалий



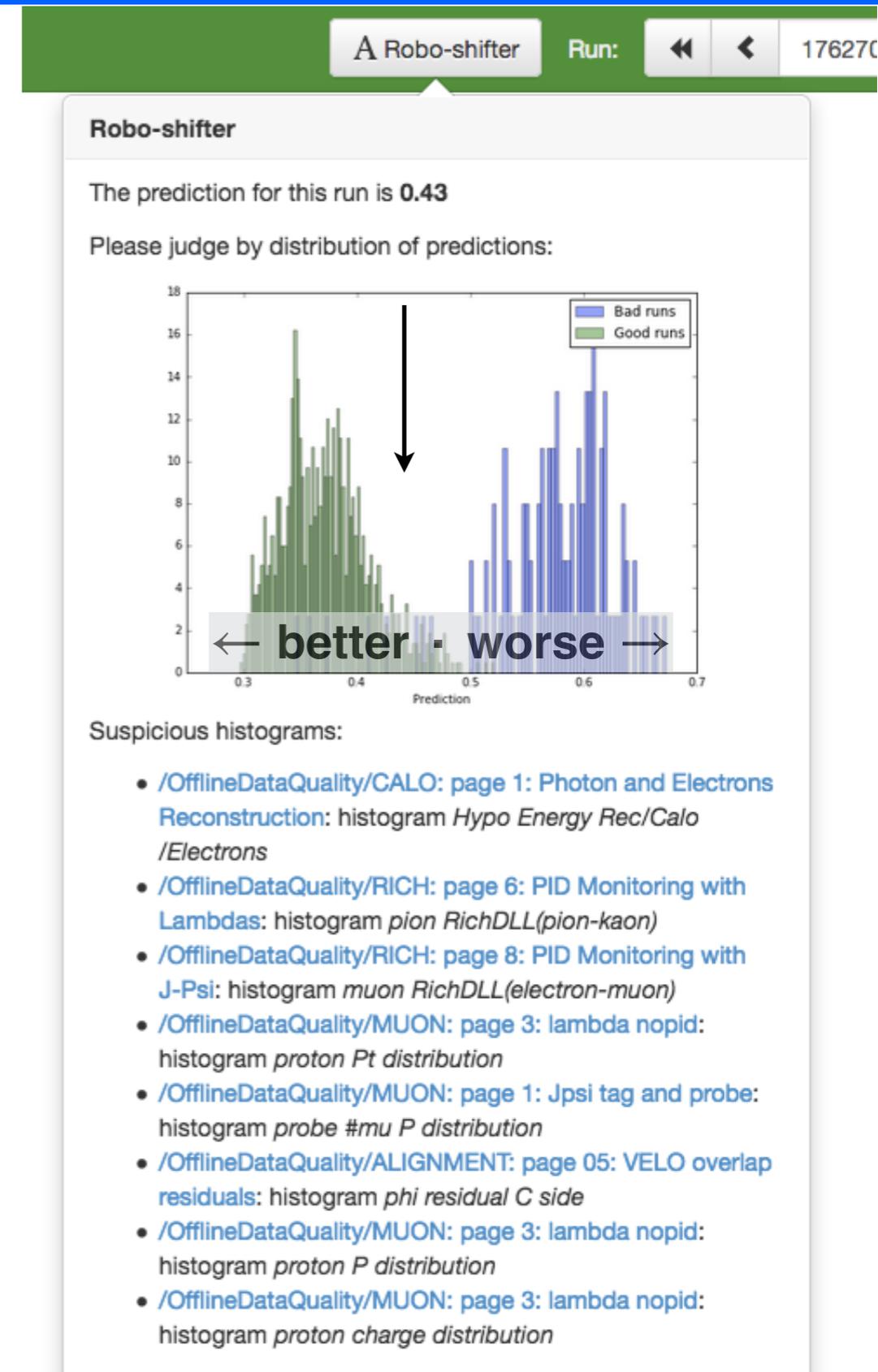
- Вход: кинематические распределения (процентили) для различных физических объектов в различных физических данных CMS
 - ~2500-D данные на входе
- Выход:
 - “данные ОК”
 - “данные BAD”
 - “отдать эксперту для принятия решения”
 - < 20 % работы



Робошифтер (LHCb, DUNE)

- Основанный на МО помощник оператора
- Анализирует данные текущие высокоуровневые данные, делает предсказание о качестве текущих данных
- Подсказывает возможные источники проблем

J.Phys.Conf.Ser. 898 (2017) no.9, 092027



МО для извлечения эффективных параметров модели

- Проблема inference: по наблюдаемым эксперимента x определить параметры модели θ^*
 - $\mathcal{L}(\theta | x) = p(x | \theta)$, $\theta^* = \arg \max_{\theta} p(x | \theta)$
- исследуемая модель задает распад на партонном уровне
 - $p(z_{parton} | \theta) = \frac{1}{\sigma(\theta)} \frac{d\sigma(\theta)}{dz_{parton}}$
 - z_{parton} - партонные распределения, θ - параметры модели
- затем: адронизация, прохождение через детектор, взаимодействие с детектором, дигитизация, реконструкция...
 - $p(z_{hadronization} | z_{parton})$, $p(z_{parton} | z_{detector})$ - распределения для адронизации партонов и различных откликов детектора
 - множество неконтролируемых промежуточных состояний
- В конце получаем наблюдаемые физического анализа
 - $p(z_{detector} | x)$

Стандартный подход

- Ответ на вопрос “почему в ФВЭ так важна симуляция?”

Стандартный подход

- Ответ на вопрос “почему в ФВЭ так важна симуляция?”
 - решаем обратную задачу:

$$\mathcal{L}(\theta | x) = p(x | \theta) = \int dz_{parton} \int dz_{hadronization} \int dz_{detector} \times$$

$$p(z_{parton} | \theta) p(z_{hadronization} | z_{parton}) p(z_{detector} | z_{hadronization}) p(x | z_{detector}) =$$

$$\int dz_{parton} \int dz \times p(z_{parton} | \theta) p(z | z_{parton}) p(x | z)$$

$$\theta^* = \arg \max_{\theta} \mathcal{L}(\theta | x)$$

- интеграл включает в себя многие-многие внутренние “латентные” величины z_i , ненаблюдаемые в эксперименте
 - θ^* не может быть вычислена аналитически
- решаем обратную задачу через итеративное решение прямой задачи:
 - симулируем наблюдаемые для разных θ
 - подбираем θ^* , чтобы симулируемые данные максимально соответствовали наблюдаемым в эксперименте

МО для извлечения эффективных параметров модели

$$\mathcal{L}(\theta | x) = p(x | \theta) = \int dz_{parton} \int dz_{hadronization} \int dz_{detector} \times \\ p(z_{parton} | \theta) p(z_{hadronization} | z_{parton}) p(z_{detector} | z_{hadronization}) p(x | z_{detector}) = \\ \int dz_{parton} \int dz \times p(z_{parton} | \theta) p(z | z_{parton}) p(x | z)$$

- **Отношение локальных правдоподобий двух гипотез**

$$\frac{p(x, z_{parton}, z | \theta_0)}{p(x, z_{parton}, z | \theta_1)} = \frac{p(z_{parton} | \theta_0) p(z | z_{parton}) p(x | z)}{p(z_{parton} | \theta_1) p(z | z_{parton}) p(x | z)} = \frac{p(z_{parton} | \theta_0)}{p(z_{parton} | \theta_1)}$$

- **не зависит от свойств детектора, реконструкции и анализа, но только от свойств партонных распределений для модели**

$$r(x, z_{parton} | \theta_0, \theta_1) \equiv \frac{p(x, z_{parton}, z | \theta_0)}{p(x, z_{parton}, z | \theta_1)} = \frac{p(z_{parton} | \theta_0)}{p(z_{parton} | \theta_1)}$$

- $t(x, z_{parton} | \theta_0) \equiv \nabla_{\theta} \log p(x, z_{parton} | \theta) \Big|_{\theta=\theta_0} = \frac{\nabla_{\theta} p(z_{parton} | \theta)}{p(z_{parton} | \theta)} \Big|_{\theta=\theta_0}$

МО для извлечения эффективных параметров модели

- Отношение локальных правдоподобий двух гипотез

$$r(x, z_{parton} | \theta_0, \theta_1) \equiv \frac{p(x, z_{parton}, z | \theta_0)}{p(x, z_{parton}, z | \theta_1)} = \frac{p(z_{parton} | \theta_0)}{p(z_{parton} | \theta_1)}$$

- Отношение полных правдоподобий двух гипотез

$$r(X, | \theta_0, \theta_1) \equiv \frac{\mathcal{L}(\theta_0 | X)}{\mathcal{L}(\theta_1 | X)} \equiv \frac{p(X | \theta_0)}{p(x | \theta_1)} = \mathbb{E}_{p(x, z_{parton} | \theta_0, \theta_1)} \frac{p(x, z_{parton} | \theta_0)}{p(x, z_{parton} | \theta_1)}$$

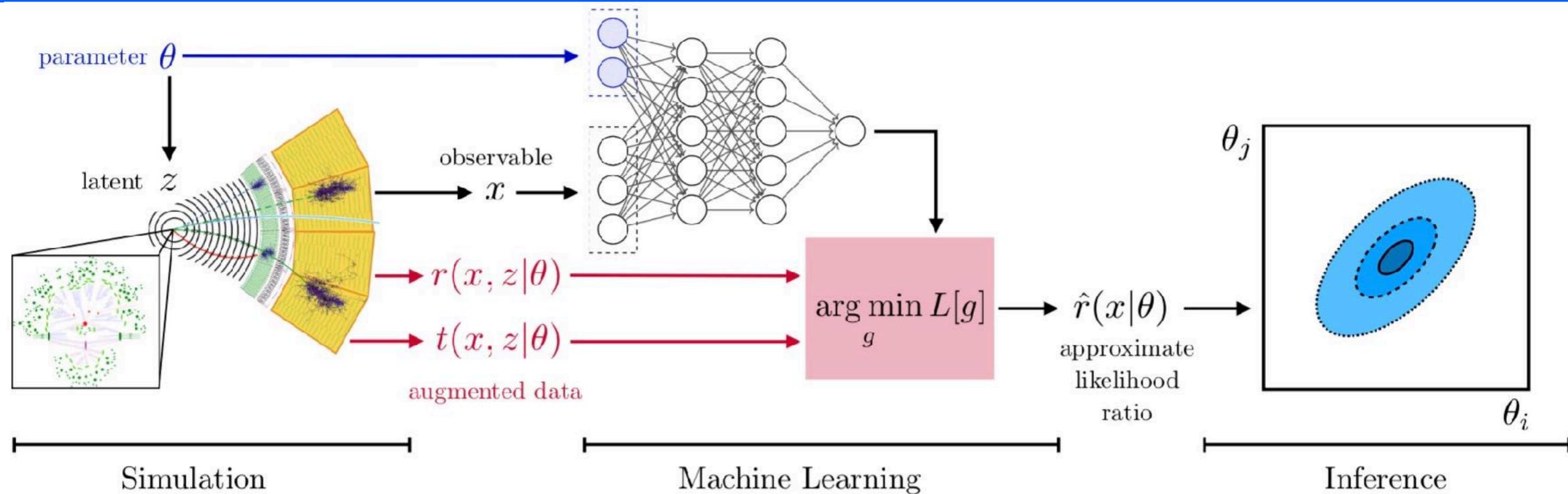
- Трюк: ML регрессор для построения суррогата $r(X, | \theta_0, \theta_1)$

$$r(x | \theta_0, \theta_1) = \arg \min_{\hat{r}} \mathbb{E}_{p(x, z_{parton} | \theta_0, \theta_1)} (r(x, z_{parton} | \theta_0, \theta_1) - \hat{r}(x))^2$$

- Аналогично для градиента $t(x, z_{parton} | \theta_0) = \frac{\nabla_{\theta} p(z_{parton} | \theta)}{p(z_{parton} | \theta)} \Big|_{\theta=\theta_0} :$

$$t(x, z_{parton} | \theta_0) = \arg \min_{\hat{t}} \mathbb{E}_{p(x, z_{parton} | \theta_0)} (t(x, z_{parton} | \theta_0) - \hat{t}(x))^2$$

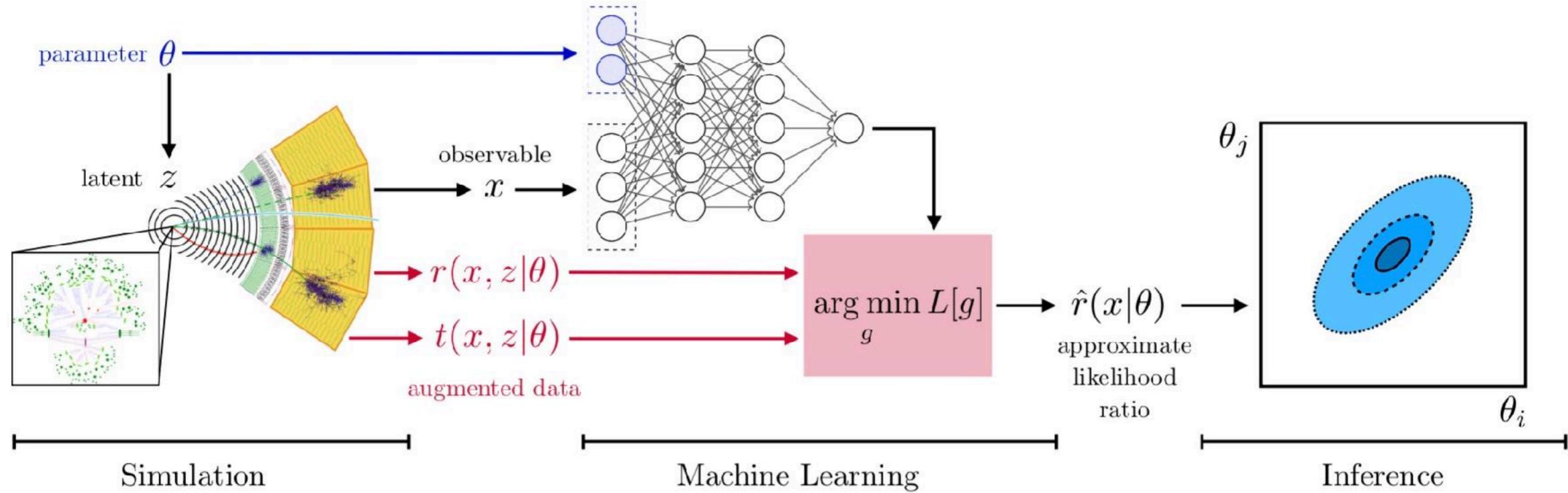
МО для извлечения эффективных параметров модели



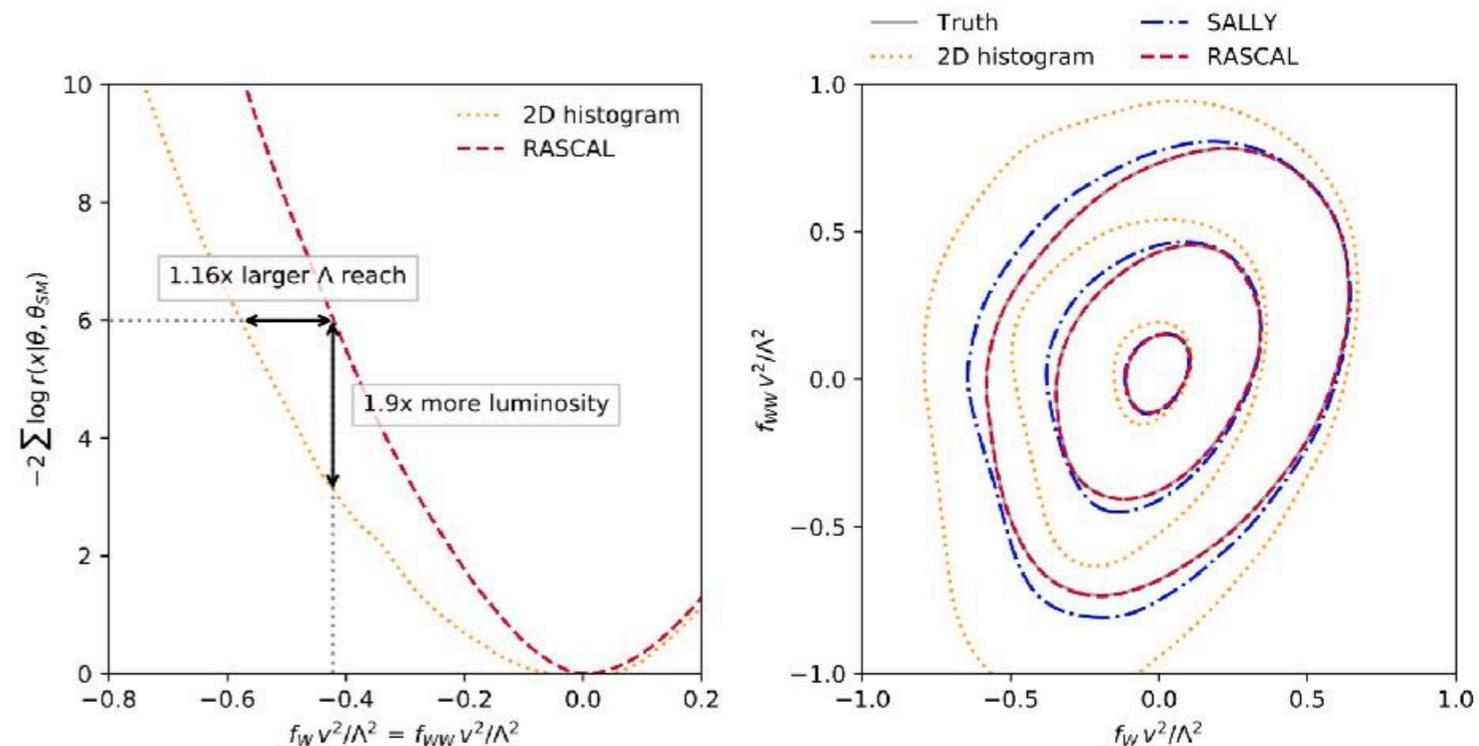
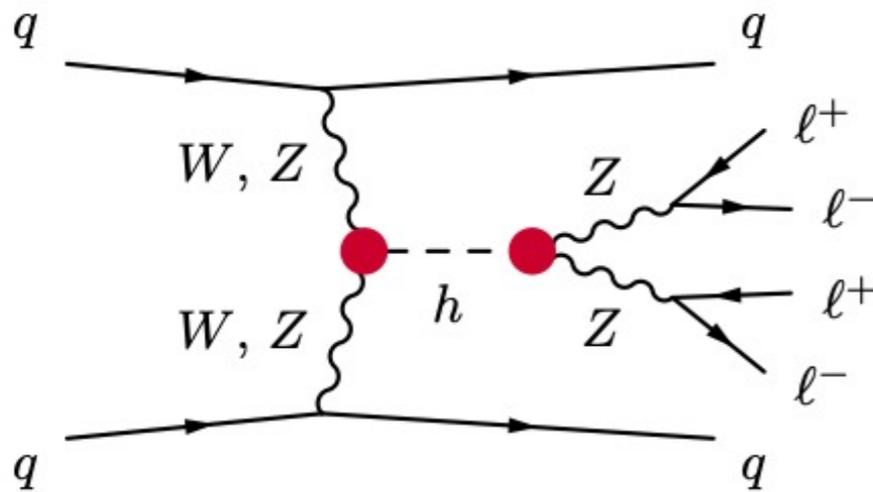
- симулируем данные $\{(x, z_{parton} | \theta_i)\}$
- конструируем $r(x, z_{parton} | \theta), t(x, z_{parton} | \theta)$
- обучаем регрессор, получаем аппроксимацию для отношения правдоподобий $\hat{r}(x | \theta_1, \theta_2)$
- используем $\hat{r}(x | \theta_1, \theta_2)$ для получения параметров модели

МО для извлечения эффективных параметров модели

arXiv:1805.00013



$$\mathcal{L} = \mathcal{L}_{\text{SM}} + \frac{f_W}{\Lambda^2} \frac{ig}{2} (D^\mu \phi)^\dagger \sigma^a D^\nu \phi W_{\mu\nu}^a - \frac{f_{WW}}{\Lambda^2} \frac{g^2}{4} (\phi^\dagger \phi) W_{\mu\nu}^a W^{\mu\nu a}$$



МО в цикле оптимизации детекторов

- МО предоставляет набор подходов и методов для эффективного фитирования многомерных функций преобразования данных к непараметрическим (универсальным) функциональным представлениям
 - позволяет быстрое переобучение при изменении параметров детекторов
 - устраняет ручную подгонку алгоритмов симуляции и реконструкции
 - создает дифференцируемые суррогатные модели для глобальной оптимизации

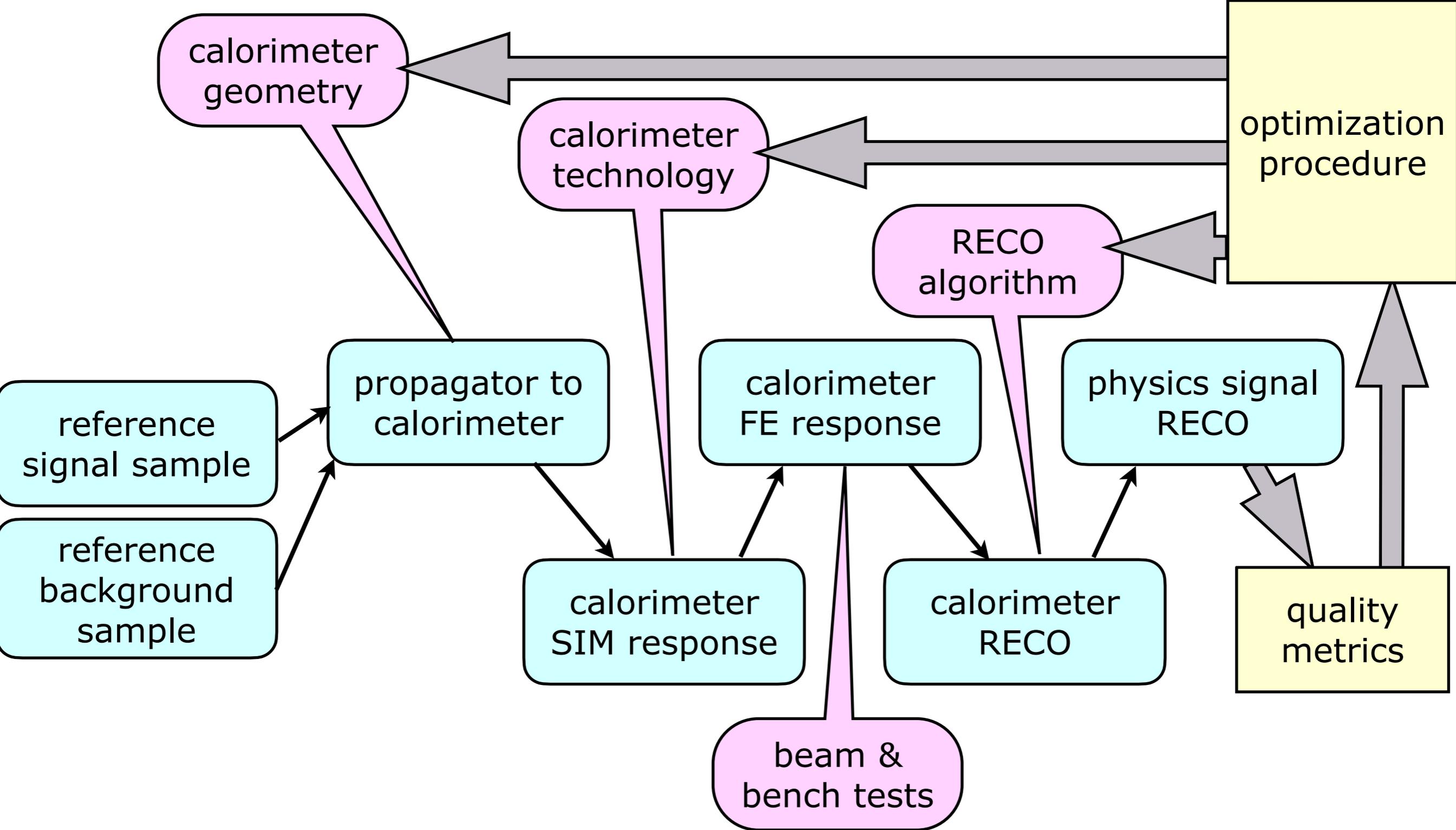
МО в цикле оптимизации детекторов

- Функции преобразования данных для типичного детектора
 - Кинематика частиц → SIM отклик детектора
 - SIM отклик детектора → RO отклик детектора (DIGI)
 - RO отклик детектора → RECO отклик детектора
 - RECO отклик детектора → объекты физической реконструкции
 - объекты физической реконструкции → метрика физической чувствительности детектора

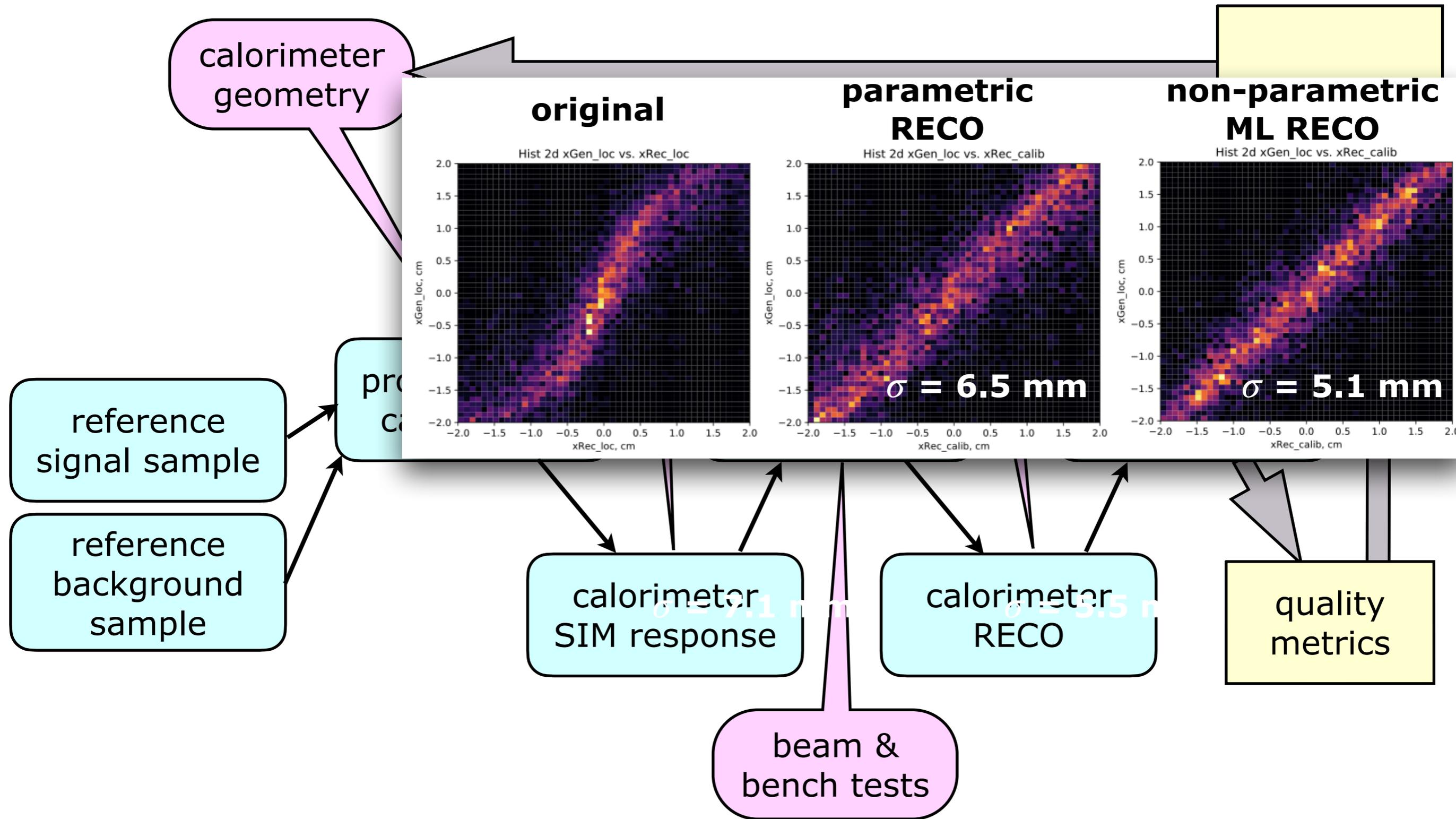
МО в цикле оптимизации детекторов

- МО предоставляет набор подходов и методов для эффективного фитирования многомерных функций преобразования данных к непараметрическим (универсальным) функциональным представлениям
- МО модель может быть несколько хуже, чем “наилучшее” возможное решение
 - тем не менее, такая модель способна определить и использовать основные закономерности в данных
 - такой оценки обычно достаточно для глобальной оценки получаемого физического качества детектора, что дает необходимую обратную связь для текущей R&D детектора
 - такую модель можно обучить быстро и автоматически

Использование МО для R&D детекторов

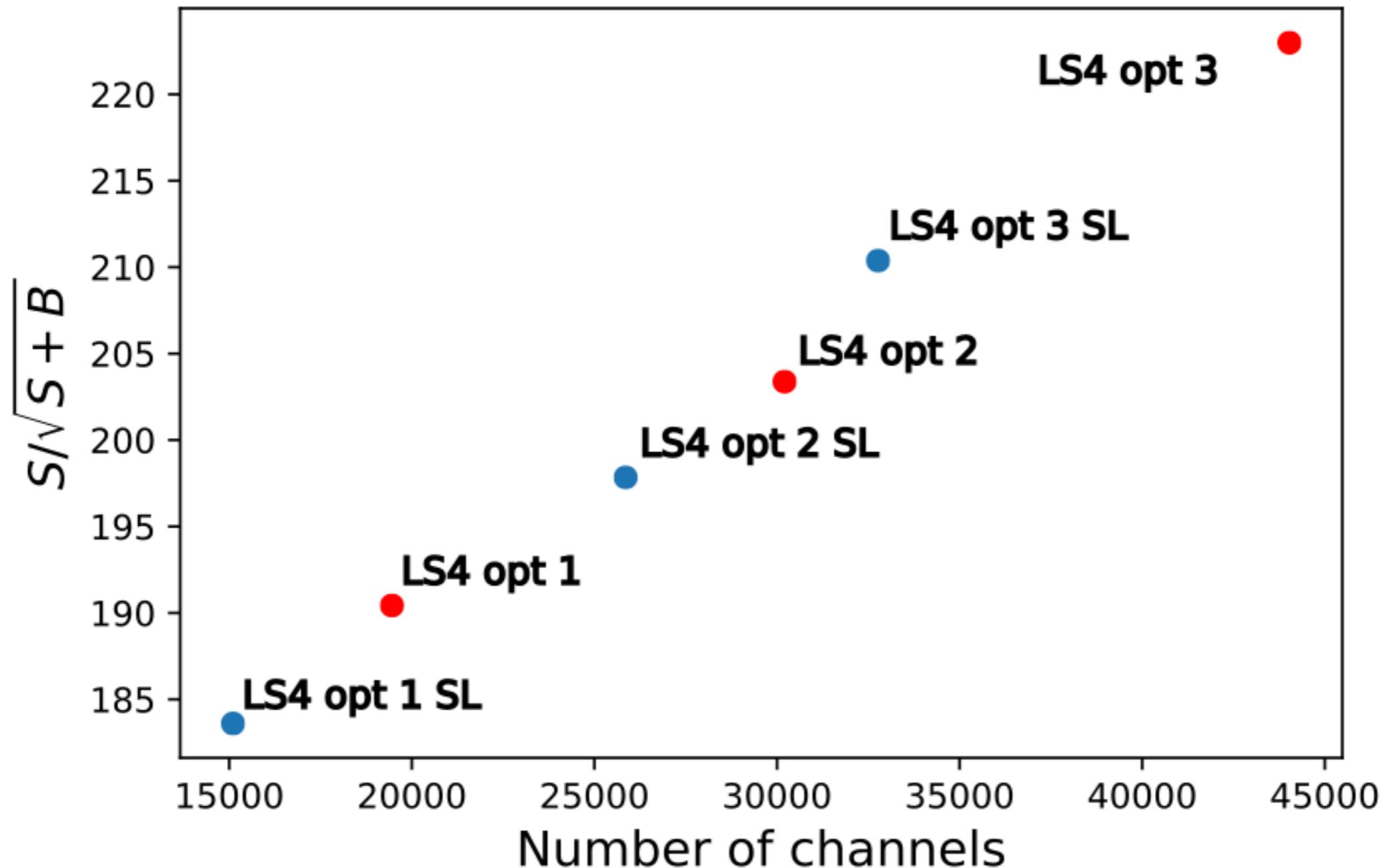


Использование МО для R&D детекторов



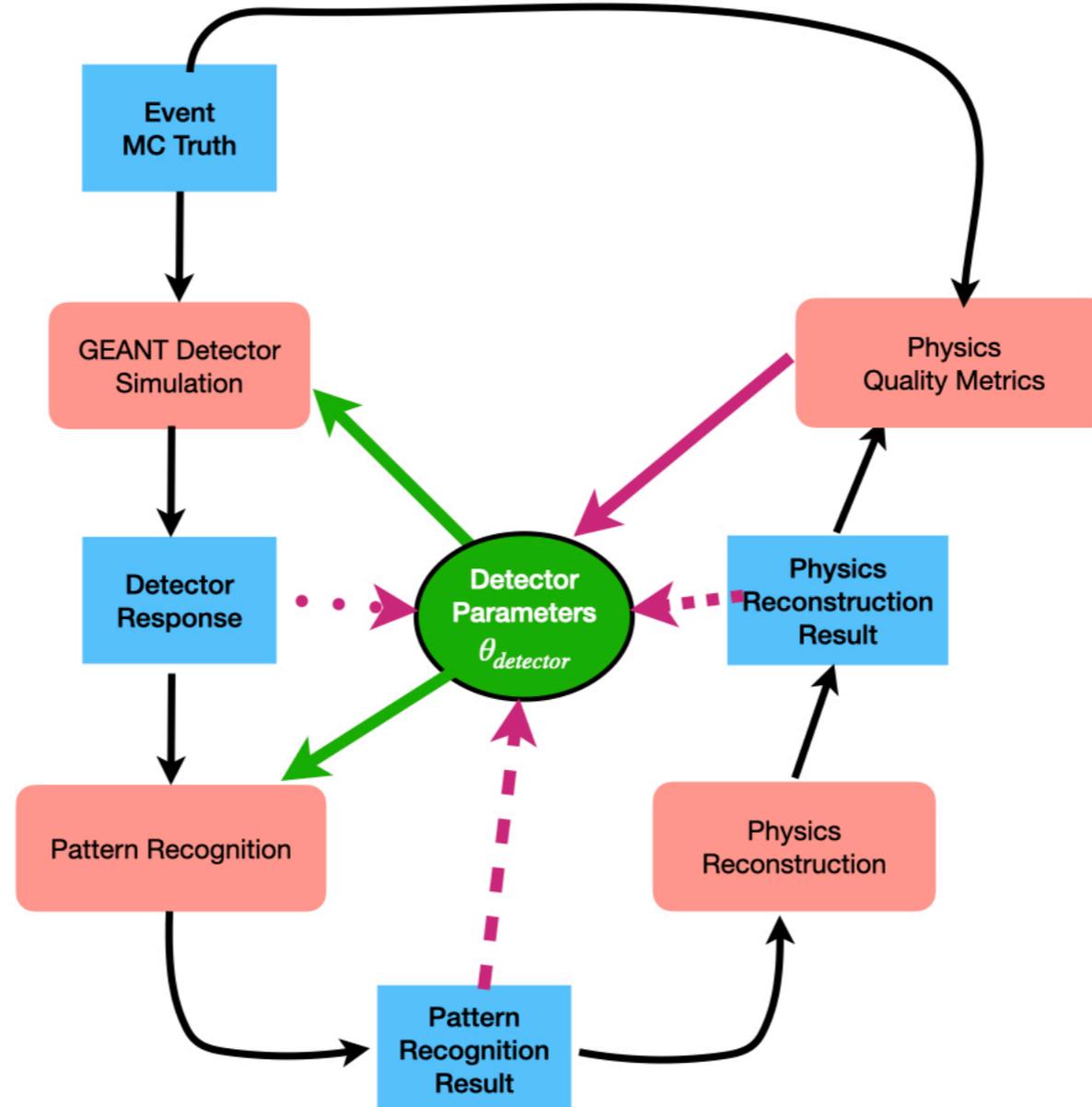
arXiv:2003.05118

Глобальная оптимизация детектора



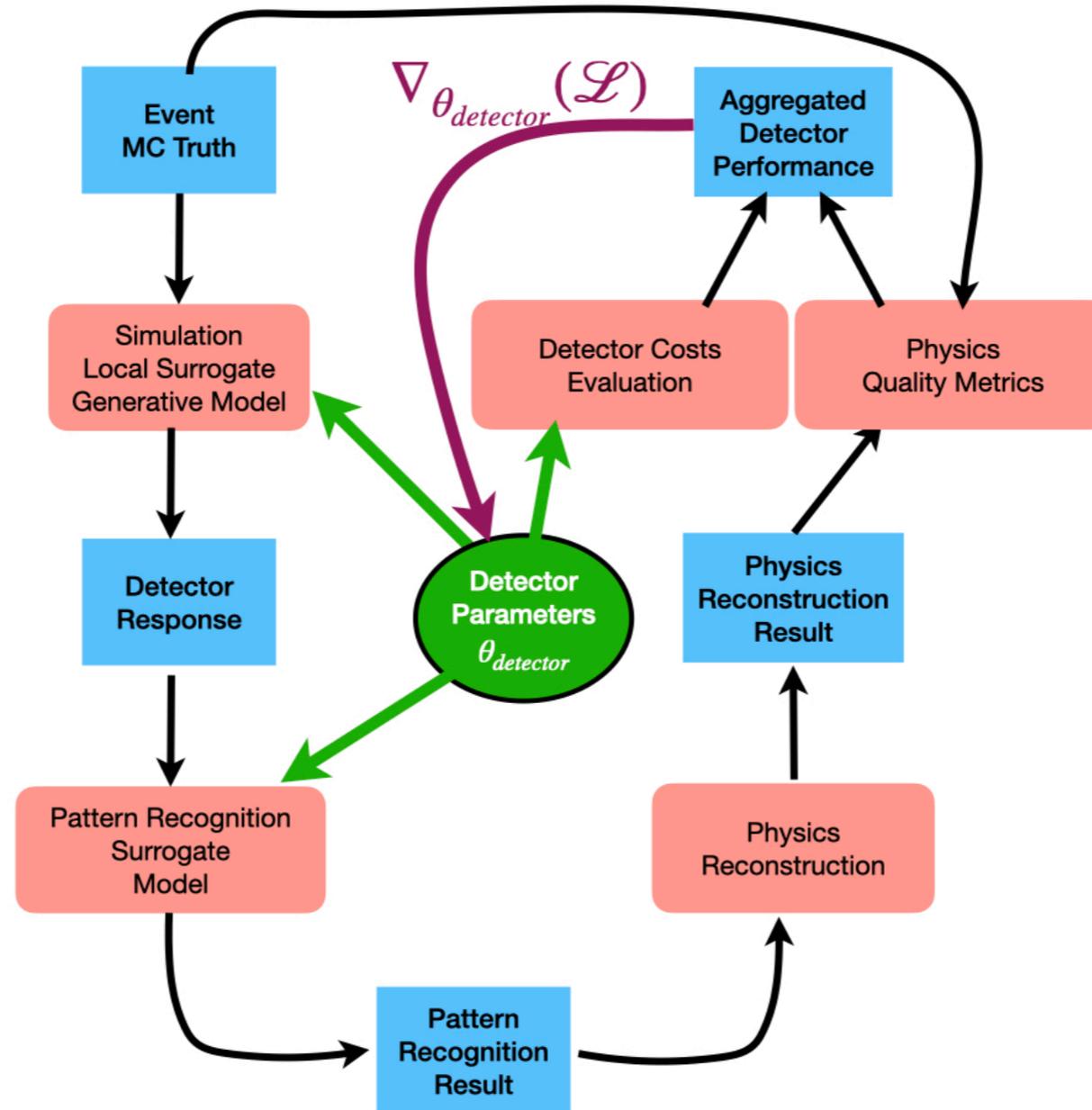
- В идеале хотим получить зависимость оценки физической чувствительности детектора от необходимых ресурсов

Оптимизационный цикл



- Стандартный цикл оптимизации использует автономную оптимизацию каждой компоненты по промежуточным метрикам
- это “жадный” алгоритм не гарантирующий глобального оптимума

Комплексная оптимизация



- одновременная оптимизация всей цепочки
- SIM → RO → RECO → PHYS объекты → физическая чувствительность

Заключение

- Применение машинного обучения, искусственный интеллект, анализ больших данных оказывают существенное влияние на развитие современной цивилизации
- Полученные наработки широко адаптируются и используются в современной ФВЭ

Физический анализ, оптимизация отделения сигнала от фона (MVA)
Быстрый отбор событий в триггере
Реконструкция в детекторах
Идентификация частиц в детекторе
Поиск аномалий

- технические аномалии: качество данных
- физические аномалии: поиск новой физики

Ускорение МС генераторов
Ускорение симуляции детекторов
Оптимизация детекторов

- Навык владения аппаратом МО стал существенным аспектом успеха физических программ современных экспериментов ФВЭ
- ... а также, использование МО для задач ФВЭ повышает мотивацию научной молодёжи, поскольку эти навыки высоко ценятся и имеют широчайшую область применения

Backup

Оптимизационный цикл

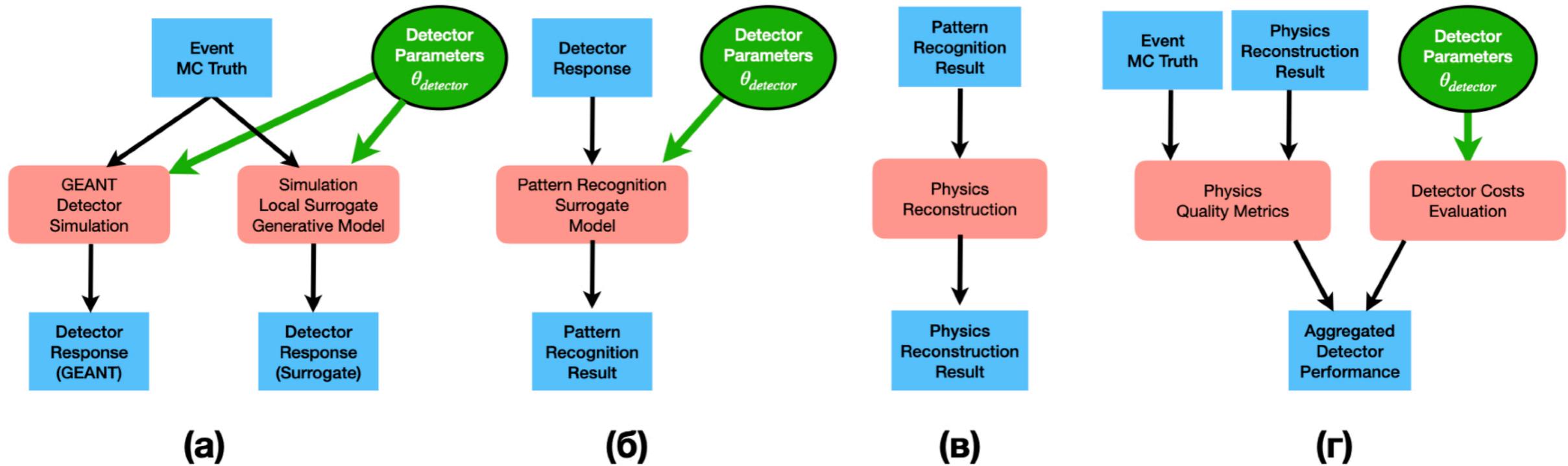


Рис.3 Четыре основных шага цикла оптимизации детектора

- (а)** - симуляция (генерация) отклика детектора для данного события
- (б)** - реконструкция отклика детектора
- (г)** - реконструкция физических параметров событий
- (д)** - вычисление физической метрики, оценка стоимости детектора, построение агрегированной метрики качества

Оптимизационный цикл

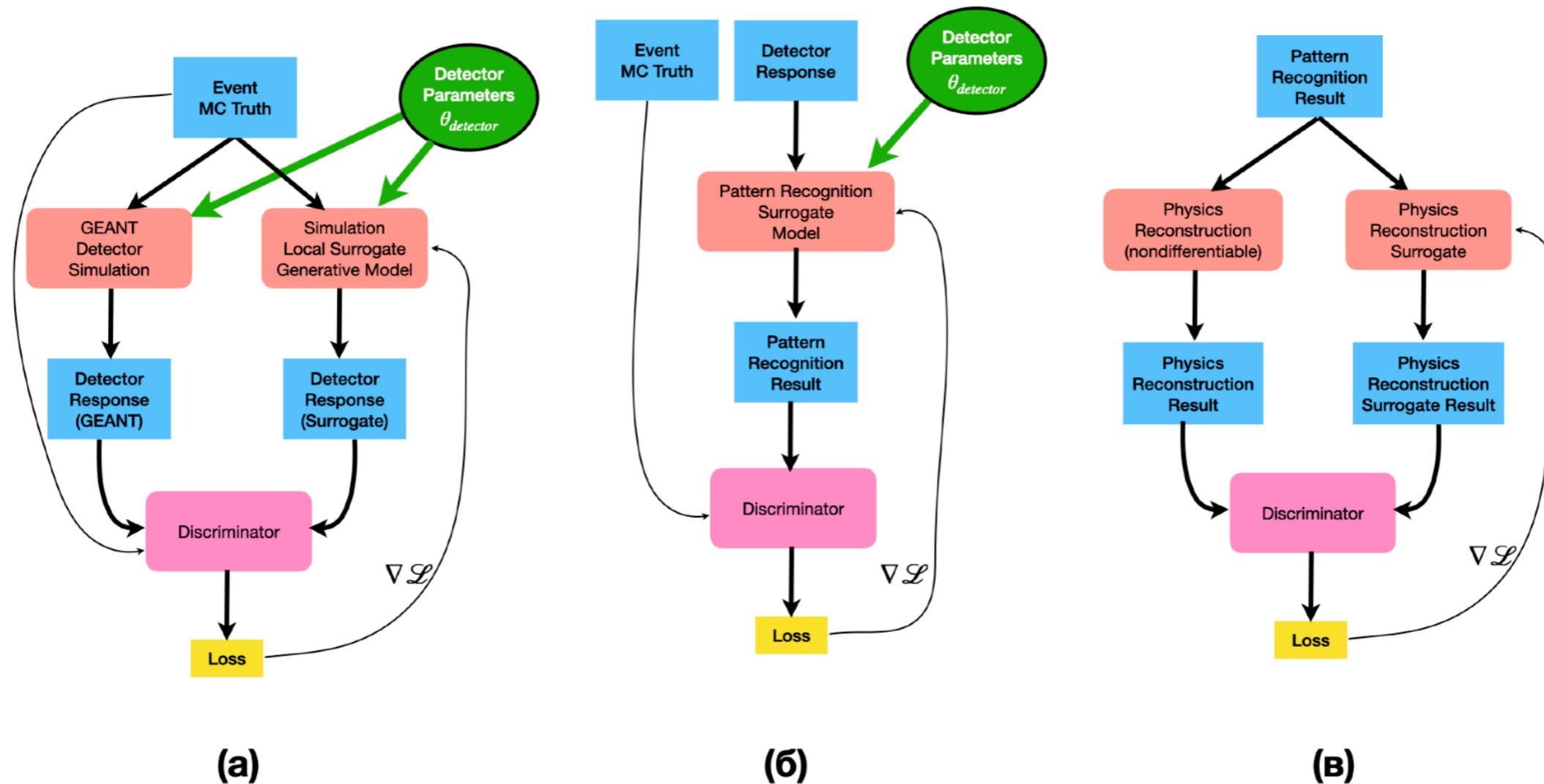


Рис.4 Схемы обучения отдельных суррогатных моделей

(а) - обучение порождающей модели для симуляции отклика детектора

(б) - обучение регрессионной модели для реконструкции в детекторе

(в) - обучение регрессионной модели для физической реконструкции, если алгоритм реконструкции не представляется в дифференцируемом виде