About Multi-Variable Analysis(MVA)

师晓东
IHEP



中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

2026.11.14-18
Belle II Winter school 2026, Southeast University



报告正体之前

- 参考: 2024 南开冬季学校 MVA 报告
- 此次报告:
  - 减少了MVA算法的介绍(不代表不重要，只是可以咨询Deekseek等LLM)
  - 侧重具体basf2的script (相比2024, 增加了存变量, 改变了优化脚本)
  - 侧重B介子全重建衰变中实际使用时train/validation的内容
    - 适用面、适用期有限(可能不适用于其他physics 组，可能light-27xx-xxx不适用)
    - 重点在“所以然”

机器学习/人工神经元网络——诺奖代言的工具

The Nobel Prize in Physics 2024





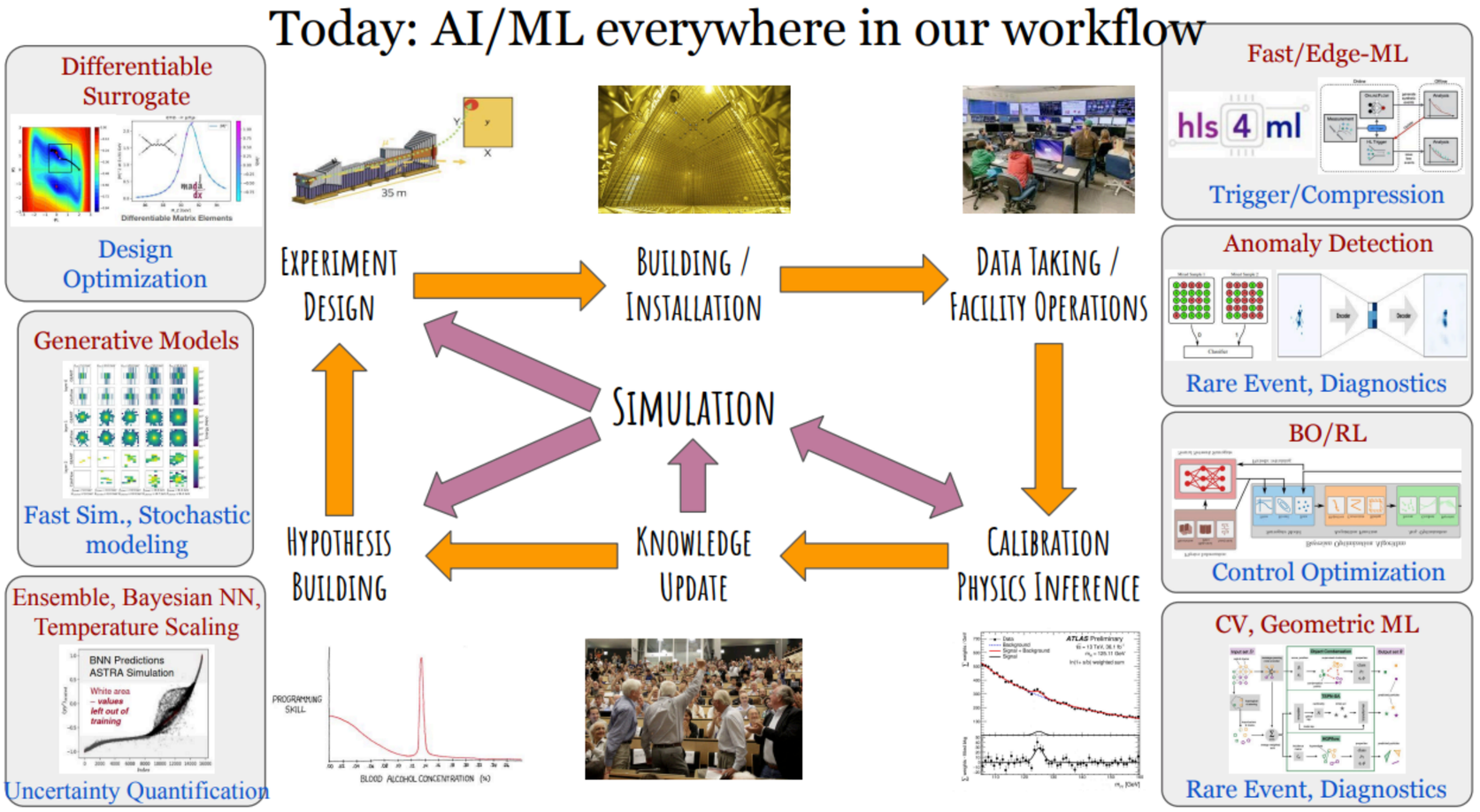
The Nobel Prize in Physics 2024 was awarded jointly to John J. Hopfield and Geoffrey E. Hinton "for foundational discoveries and inventions that enable **machine learning with artificial neural networks**"

- “工欲善其事，必先利其器”。
- “利其器”!=直接使用
- 否则，跳过新手教程后的你：



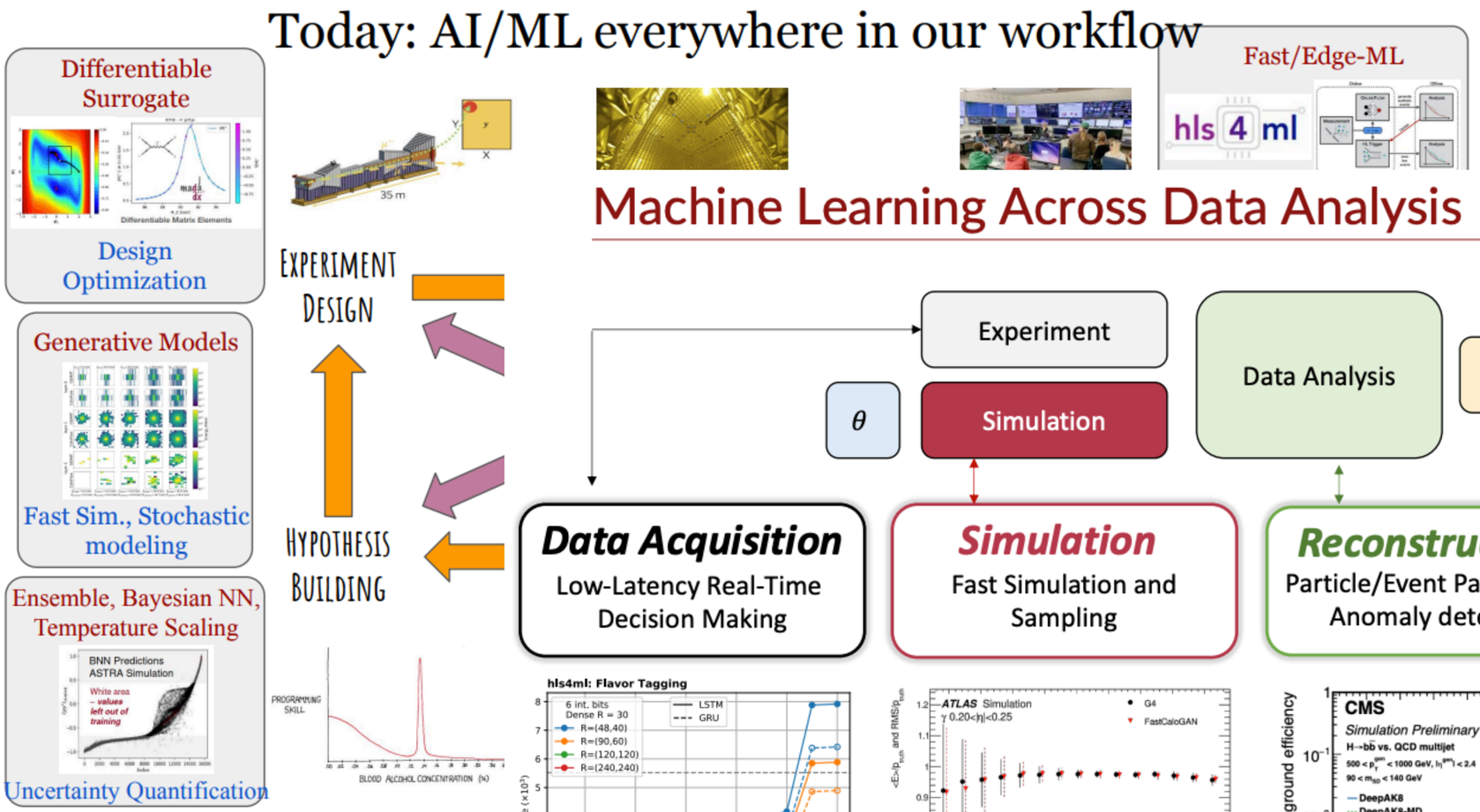


机器学习已经应用(“渗透”)在粒子物理中的各个层次



From Kazuhiro Terao

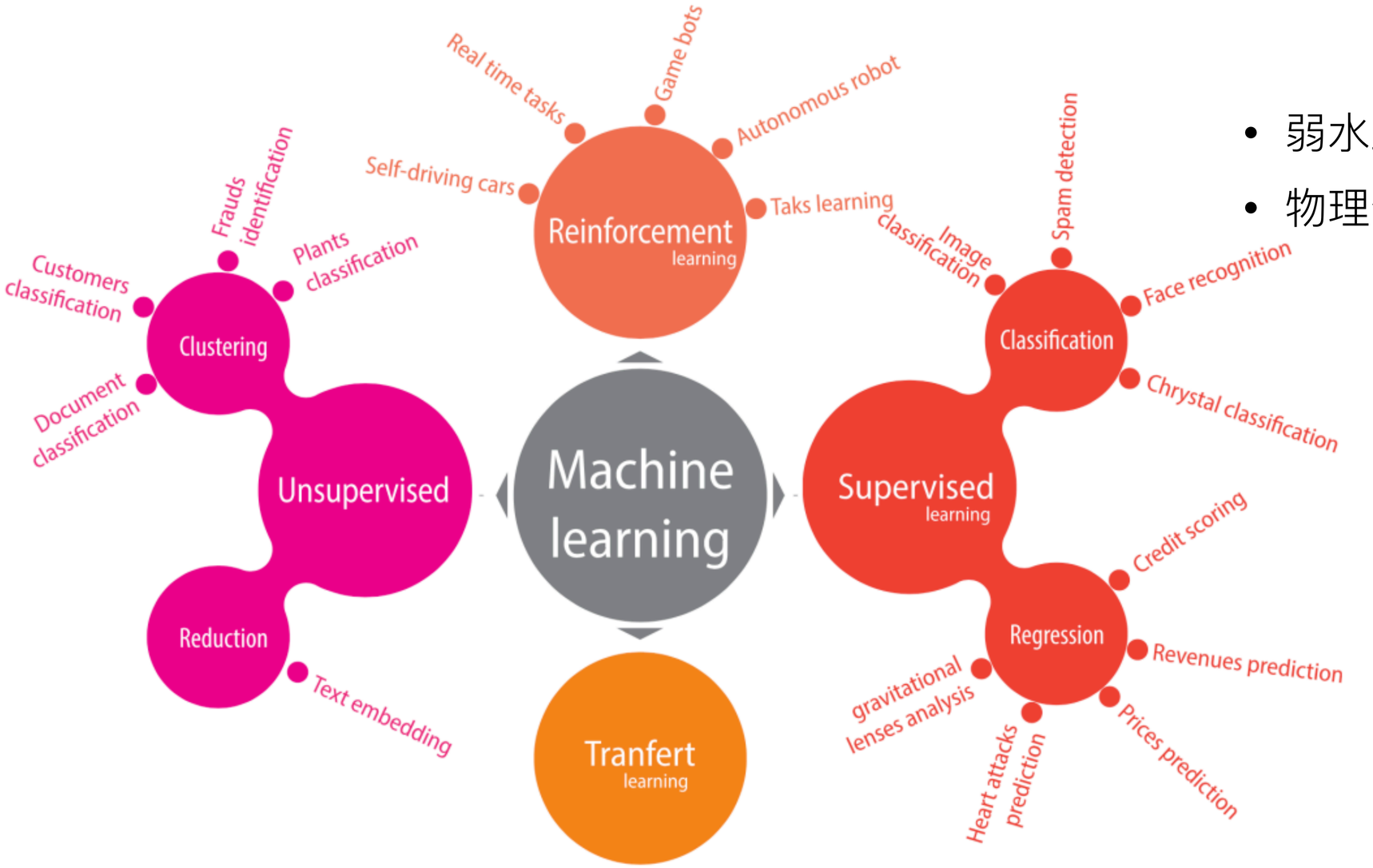
机器学习已经应用(“渗透”)在粒子物理中的各个层次



From Kazuhiro Terao

From Michael Kagan

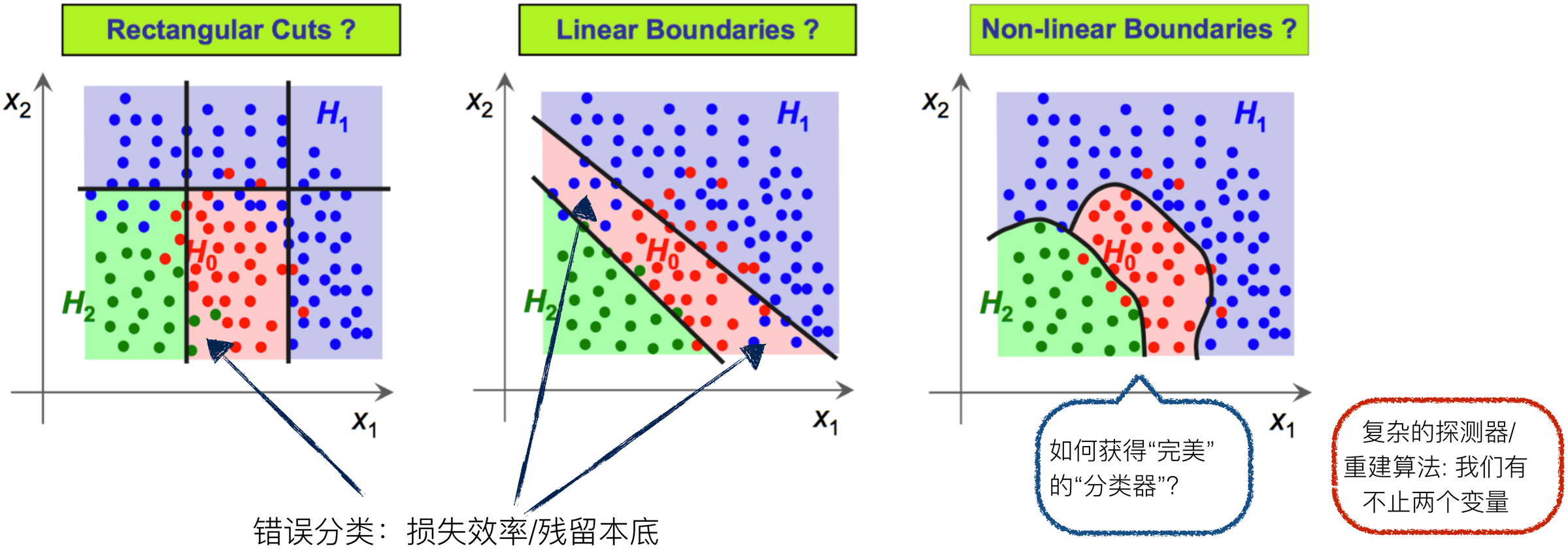
机器学习



- 弱水三千，只取一瓢—“分类器”
- 物理分析中最常用: 压低本底, “Cut”

分类器—区分信号/本底

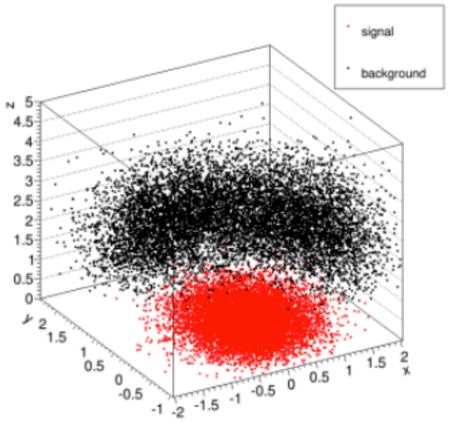
- Q:“分类”不同类别的事例，例如蓝点、绿点、红点
- A:“加cut”。(对新生:“看到一些分布，先理解/问为什么这样分布，再加cut/或做其他操作。)

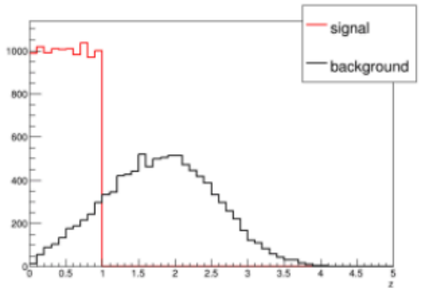


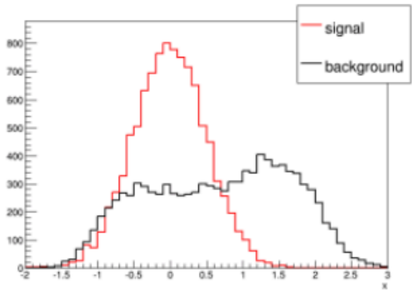
另一个例子：

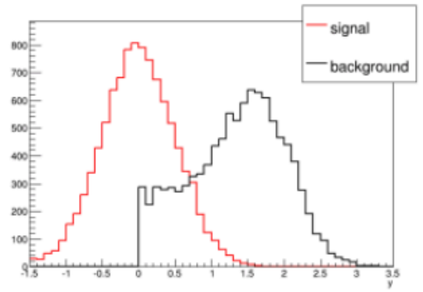
复杂的探测器/重建算法，意味着 我们有不止两个变量

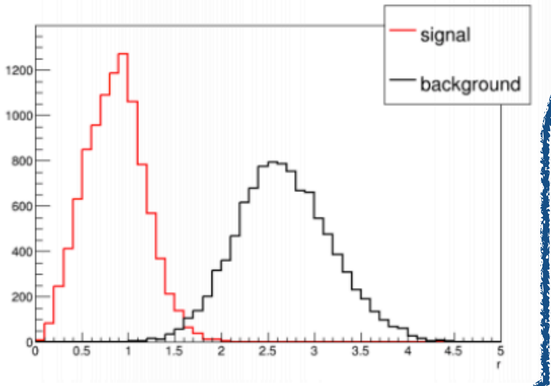
观测量： x, y, z











- **判别函数**
  $r=\sqrt{x^2+y^2+z^2}$

👨‍💻:通过理解x,y,z的含义，对比三维分布，r可能是不错的“分类器”

👨‍🏫:👍🏻

💻: 直接“暴力”破解，只要参数够多，我能变成AI SI

Boosted Decision Tree (BDT)

Why BDT, why not others?

- 客观、非物理原因: Belle II都在用，basf2自带，无需自己调包
- 专业原因:

开箱即用，调参负担较轻

BDT 在默认参数下往往就能达到接近最优的性能，对不想花大量时间做超参数搜索的使用者比较友好。有高能物理领域的实践者指出，BDT 在给定变量集下用很少的调参就能接近最优，而调优神经网络则要花费多得多的时间。同时，BDT 训练速度明显快于多层感知机，同样的时间内可以试几十组参数，而神经网络可能只够训一次。 9 10

对无关特征有较好的鲁棒性

BDT 在加入大量纯噪声维度时，性能下降相对平缓。一项研究对比发现，神经网络在加入少数几个无信息特征后性能就明显下滑，而 BDT 在加入 10 维噪声后下降很小，加入 30 到 50 维噪声后仍能保持相当的性能。这意味着在实际任务中，如果特征筛选做得不够精细，BDT 的容错空间更大。 3

不易过拟合，输出稳定

只要树的最大深度和叶节点最小样本数设置得当，BDT 通常不会因为树的数量多而严重过拟合。使用大量浅层小树（例如约 1000 棵）而非少量深树，既能保持鲁棒性，又能获得稳定的分类器。此外，某些场景下 BDT 对类别不平衡数据的稳定性也较好，在 25 个不平衡网络流量数据集上的实验中，BDT 被归为“相对稳定、不易受不平衡数据影响”的算法之一。 9

在特定任务中可与更强模型竞争

一项针对二值变量数字广告数据的研究中，作者提出的布尔决策表（BDT）在分类性能上优于随机森林和 CART，且接近 XGBoost，同时由于结果更可解释，在广告行业场景中比 XGBoost 更具实用价值。不过需要注意，这里的 BDT 是布尔决策表，不是提升决策树，不能直接等同于通用 BDT。 2

Why BDT, why not others?

- 客观、非物理原因: Belle II都在用，basf2自带，无需自己调包
- 专业原因:

开箱即用，调参负担较轻

BDT 在默认参数下往往就能达到接近最优的性能，对不想花大量时间做超参数搜索的使用者比较友好。有高能物理领域的实践者指出，BDT 在给定变量集下用很少的调参就能接近最优，而调优神经网络则要花费多得多的时间。同时，BDT 训练速度明显快于多层感知机，同样的时间内可以试几十组参数，而神经网络可能只够训一次。 9 10

对无关特征有较好的鲁棒性

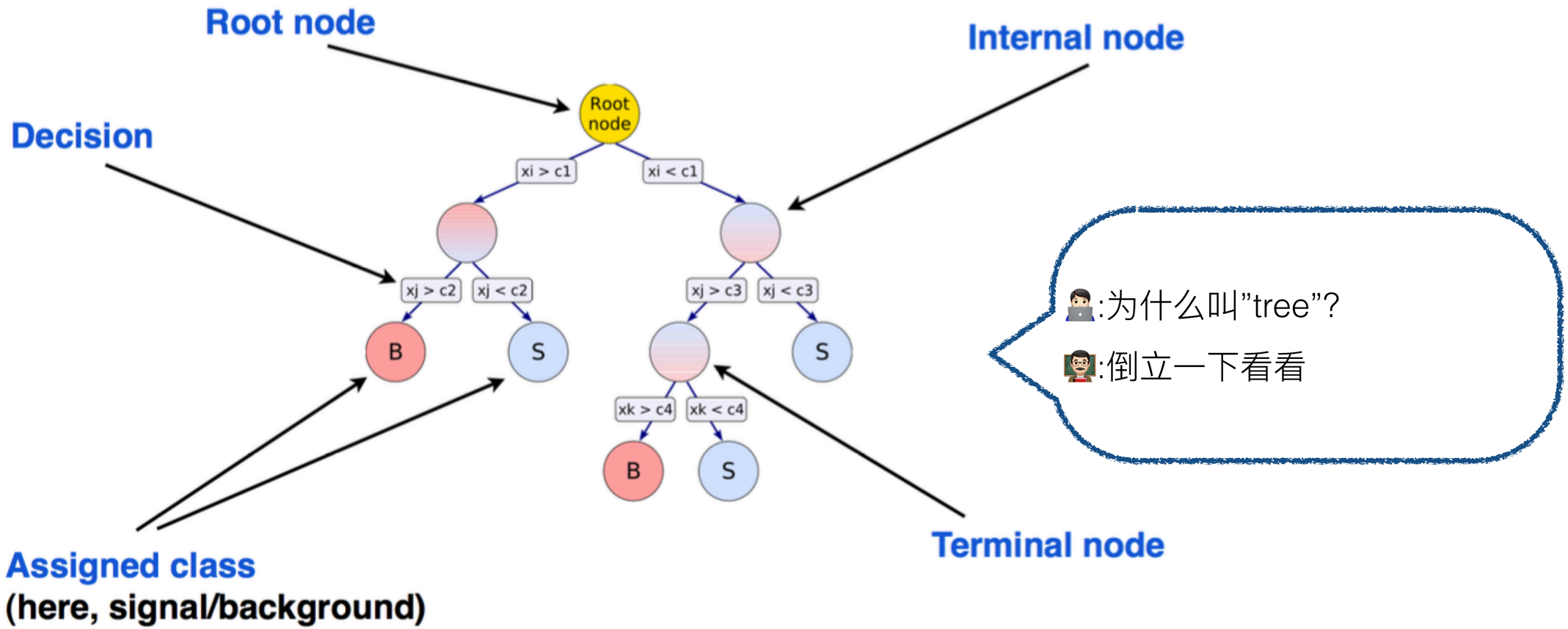
不等于“禁止尝试其他算法”
一切为了更优物理结果。最后有非BDT的例子

BDT 对类别不平衡数据的稳定性也较好，在 25 个不平衡网络流量数据集上的实验中，BDT 被归为“相对稳定、不易受不平衡数据影响”的算法之一。 9

在特定任务中可与更强模型竞争

一项针对二值变量数字广告数据的研究中，作者提出的布尔决策表（BDT）在分类性能上优于随机森林和 CART，且接近 XGBoost，同时由于结果更可解释，在广告行业场景中比 XGBoost 更具实用价值。不过需要注意，这里的 BDT 是布尔决策表，不是提升决策树，不能直接等同于通用 BDT。 2

Decision Tree(DT)

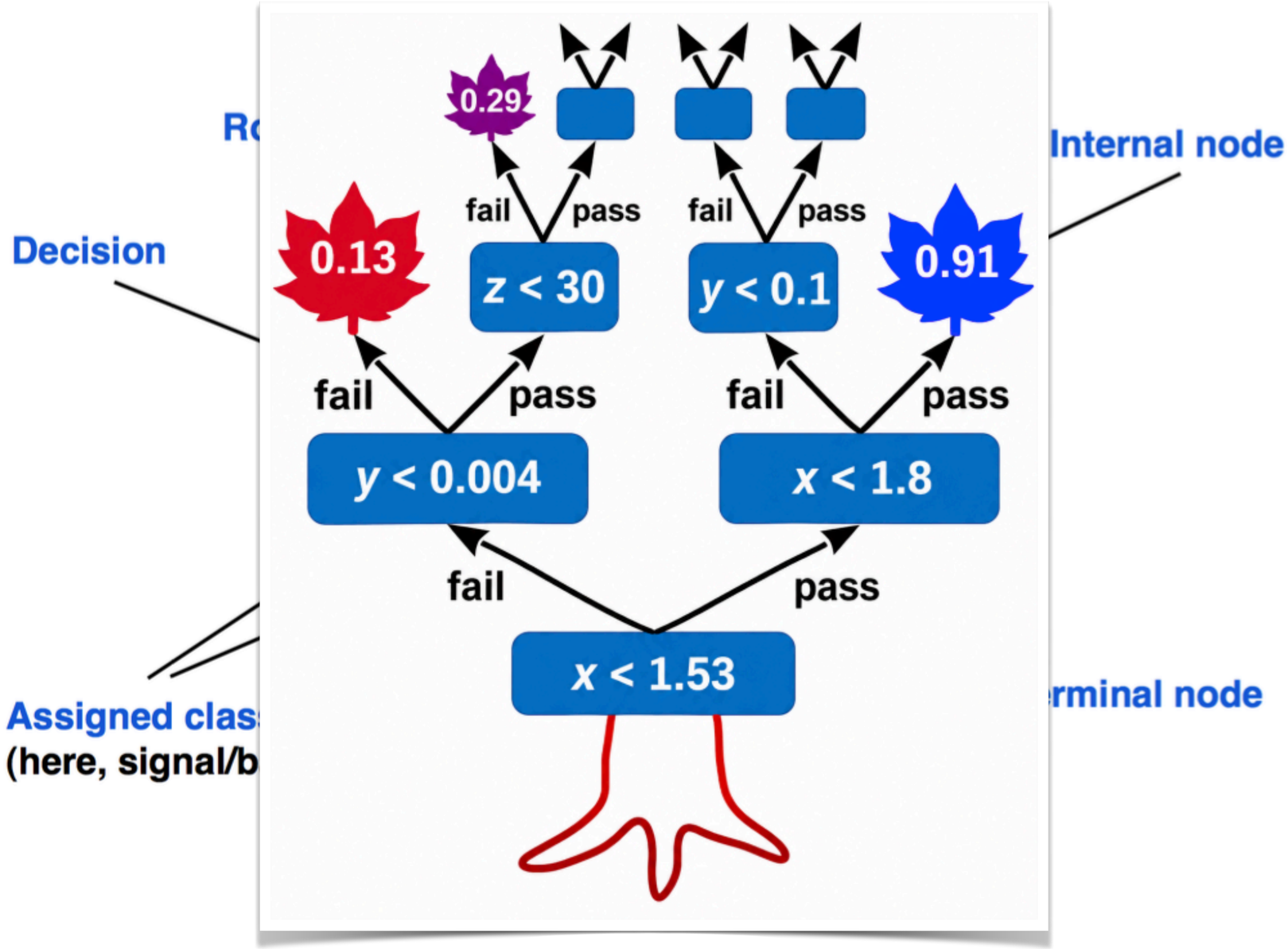


A Simple Decision Tree

👨‍💻:为什么叫”tree”?

👩‍🏫:倒立一下看看

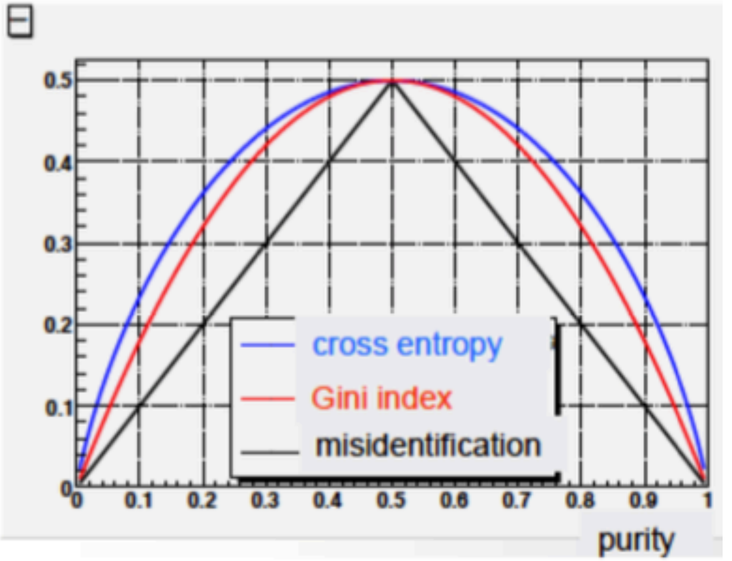
Decision Tree(DT)



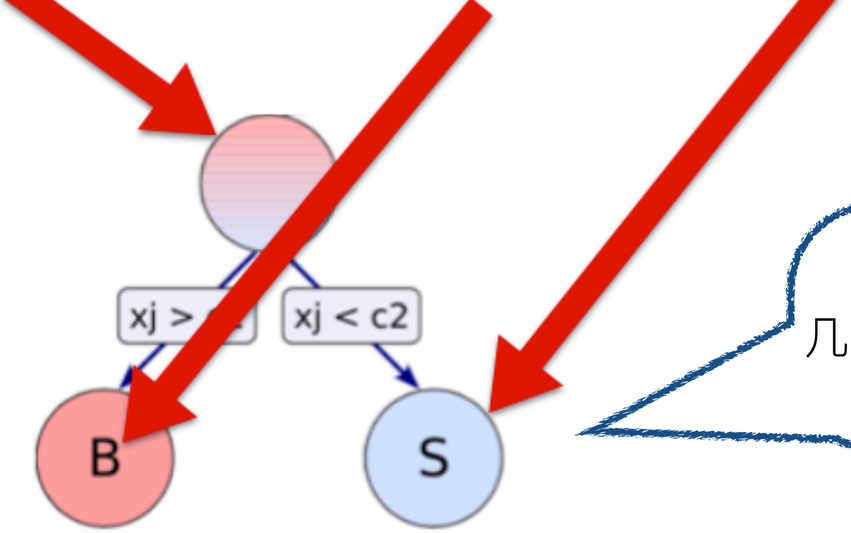
Decision Tree(DT)

- Training process:

- *Gini Index* [default], defined by $p \cdot (1-p)$;
- *Cross entropy*, defined by $-p \cdot \ln(p) - (1-p) \cdot \ln(1-p)$;
- *Misclassification error*, defined by $1 - \max(p, 1-p)$;
- *Statistical significance*, defined by $S/\sqrt{S+B}$;



MAXIMUM $N_{parent} \cdot Gini_{parent} - N_{left} \cdot Gini_{left} - N_{right} \cdot Gini_{right}$



训练即优化：设置一个量，告诉💻，让其取极值。

此处用Gini构造: purity越高, Gini越小。

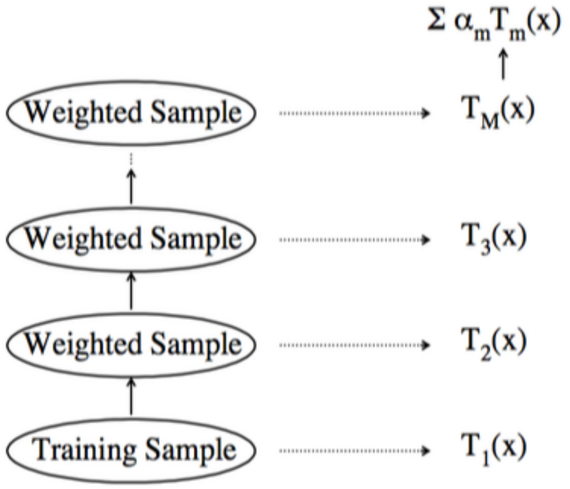
分到何时不再分了?

几个提前设置的条件: purity足够高 或 N过少 或层数过深

Decision Tree(DT) -> BDT

- DT的问题/局限：不稳定，因为训练样本的涨落可能会训练出完全不一样的tree；很容易overtrainging。
  - ✳ 对策: **Boosting**, bagging, random forests, pruning等等。

- Boosting 算法最早由Schapire在1990年给出。AdaBoost 由Freund, Schapire 在1996年给出。2005年首次在高能物理中应用(MiniBooNe实验在PID中)。
- 什么是 **Boosting**?
  - ✳ 很难训练出一个完美的DT(诸葛亮),但训练不完美有误判的DTs(一群臭皮匠)就容易多了;然后把众多DTs整合起来以减少误判。(注:Boosting不仅局限于DT算法)



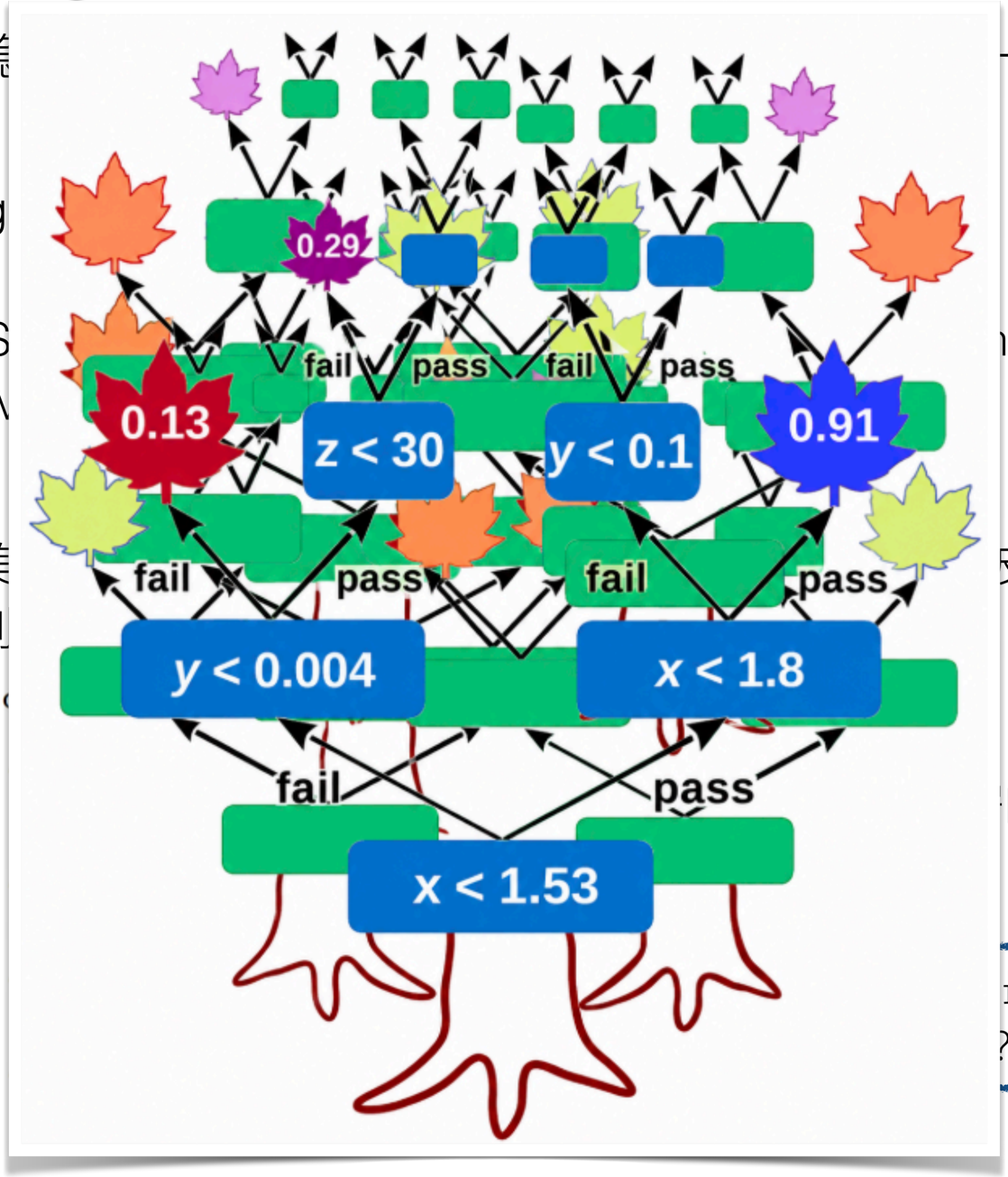
- 基于reweighted样本训练一个DT,根据DT对训练样本赋权重
- 重复第一步得到若干DTs
- 根据全部的DTs给出输出

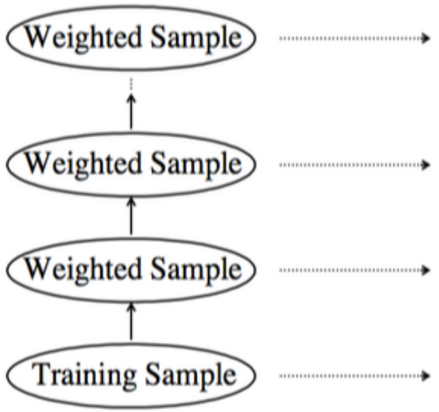
如何赋权重?

如何给最后结果?

Decision Tree(DT) -> BDT

- DT的问题/局限：不稳 一样的tree；很容易overtrainging。
  - 对策: **Boosting**, bag
- Boosting 算法最早由S napire 在1996年给出。2005年首次在高能物理中应用(M
- 什么是 **Boosting**?
  - 很难训练出一个完美 支匠)就容易多了;然后把众多DTs整合起来以减少误判





DT对训练样本赋权重

如何赋权重?

最
?

AdaBoostDT

- 例如:有个训练样本,总数为N,有两类事例(信号为1，本底为-1)
- 区分信号本底—也即贴标签,$T_m(x_i)$为第m个DT对x_i事例给出的标签

错误率:

$$err_m = \frac{\sum_{i=1}^{N} w_i I(y_i \neq T_m(x_i))}{\sum_{i=1}^{N} w_i}$$

- x为true时,I(x)=1;
- x为false时,I(x)=0

$$\alpha_m = \beta \times \ln((1 - err_m)/err_m)$$

- Beta默认为1,可调(超参数)。

$$w_i \rightarrow w_i \times e^{\alpha_m I(y_i \neq T_m(x_i))}$$

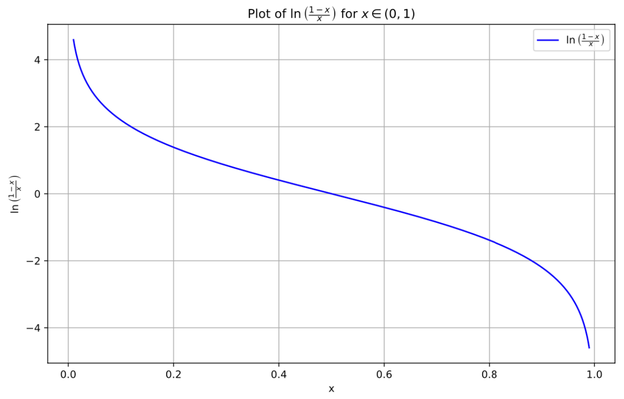
- 只对误标的事例改变权重。

$$w_i \rightarrow w_i / \sum_{i=1}^{N} w_i$$

- 并重归一化

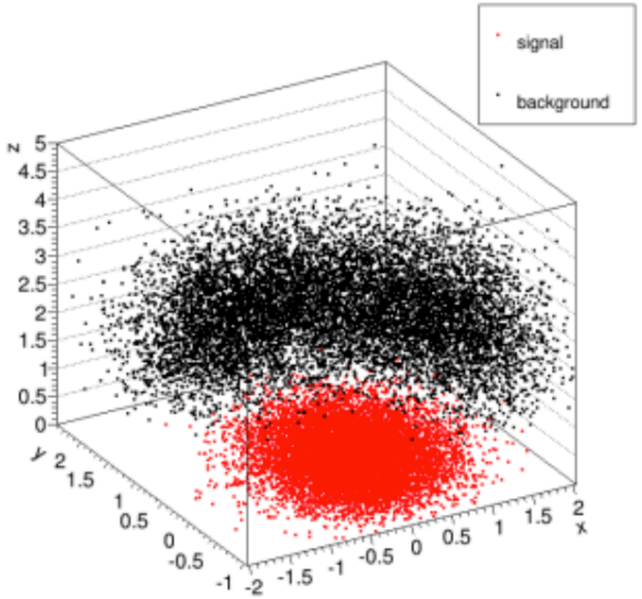
- 最终BDT的输出:

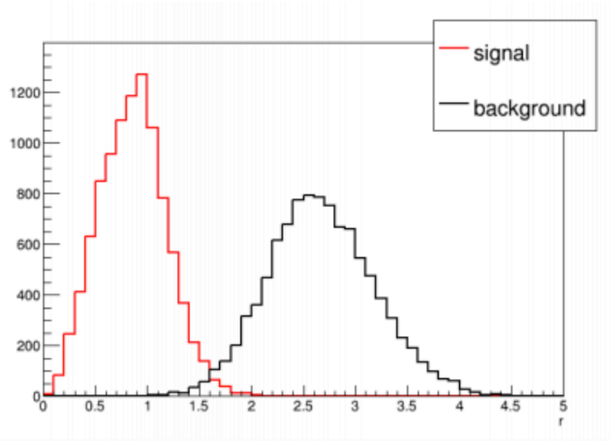
$$T(x) = \sum_{m=1}^{N_{tree}} \alpha_m T_m(x)$$



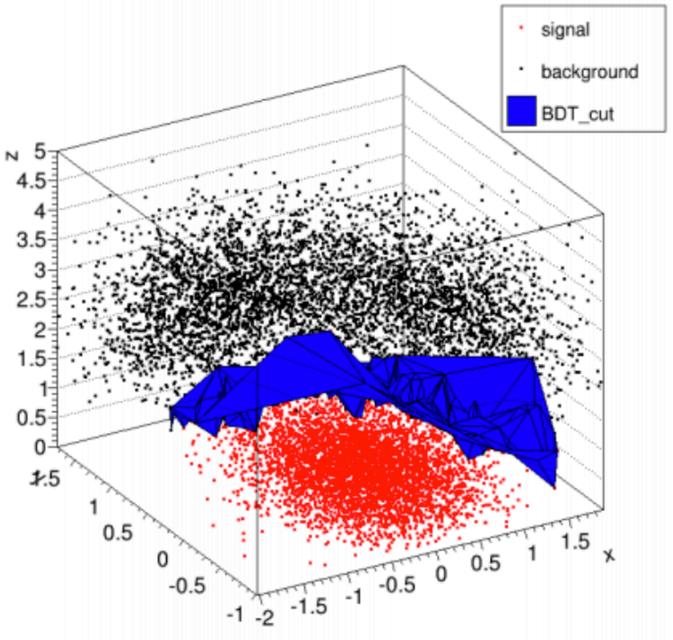
alpha随err的函数

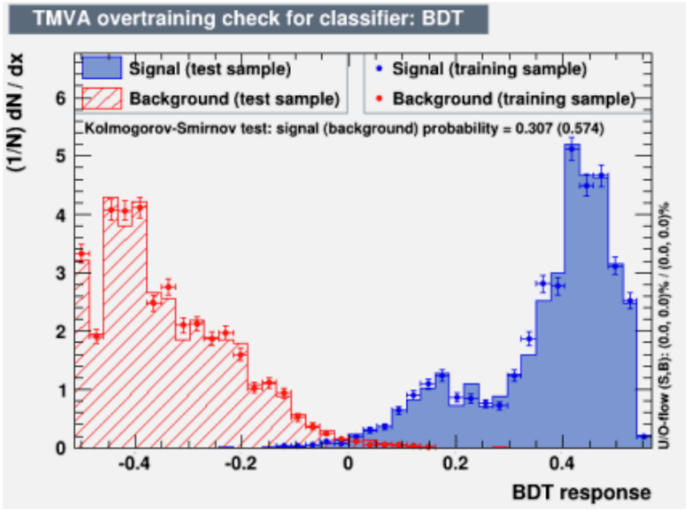
BDT的效果(例子)





- 判别函数 $r=\sqrt{x^2+y^2+z^2}$





👩‍💻👨‍💻:区分效果更好了，但不规则的cut的物理意义是?

💻:I don't care, do you?

👩‍💻👨‍💻:Sometimes we do

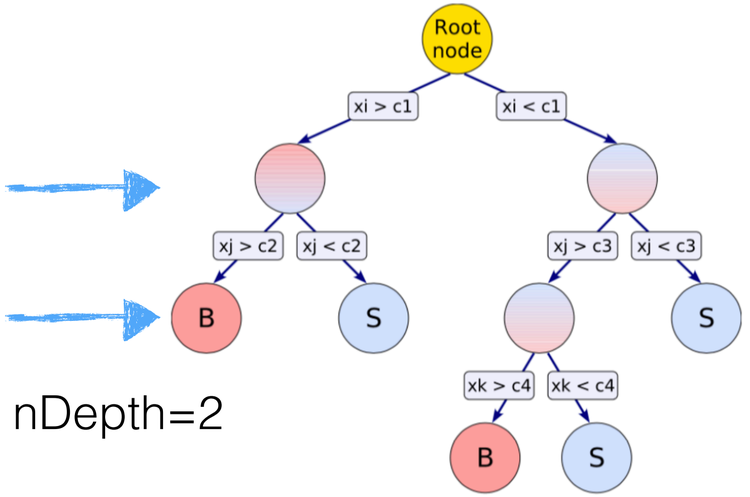
实际使用时的“这这那那”

实际使用MVA时的一些经典问题

- 我需要调整超参数么?
- 用哪些变量作为input?
- 如何判断是否overtraining?
- 如何使用output?
- 什么时候用MVA?
- 如何筛选input variables?
- 报告/note里放啥信息?
- …
- More questions?

实际使用MVA时的一些经典问题—个人答案

- 我需要调整超参数么?
  - 对应BDT要训练的参数个数/量级。过多可能train sample不够导致overtraining; 过少 可能导致结果不好
  - 可以尝试调整。有些MVA算法应避免一些超参数值,例如BDT的nDepth 不应过大(例如3)。
- 用哪些变量作为input?
- 如何判断是否overtraining?
- 如何使用output?
- 什么时候用MVA?
- 如何筛选input variables?
- 报告/note里放啥信息?
- …
- More questions?



nDepth=2

实际使用MVA时的一些经典问题—个人答案

- 我需要调整超参数么?
- 用哪些变量作为input?
  - 这些变量在定义和分布上本身能够“或多或少”区分信号和**一种**本底
    - 一般情况下区分signal与**一种**本底(例如qqbar)
    - 虽然BDT等算法支持区分三类(两种本底A, B)，但**我个人觉得**两个先后的BDT(一个区分signal与A，一个区分signal与B)要好于直接区分三类
- 如何判断是否overtraining?
- 如何使用output?
- 什么时候用MVA?
- 如何筛选input variables?
- 报告/note里放啥信息?
- …
- More questions?



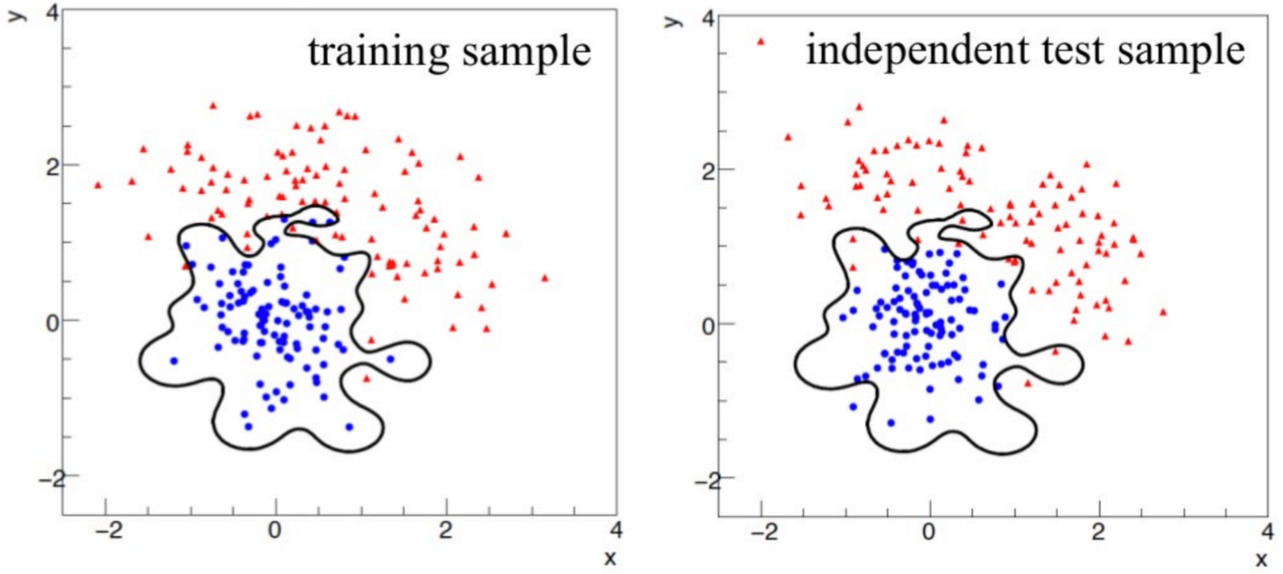


From Yann Coadou

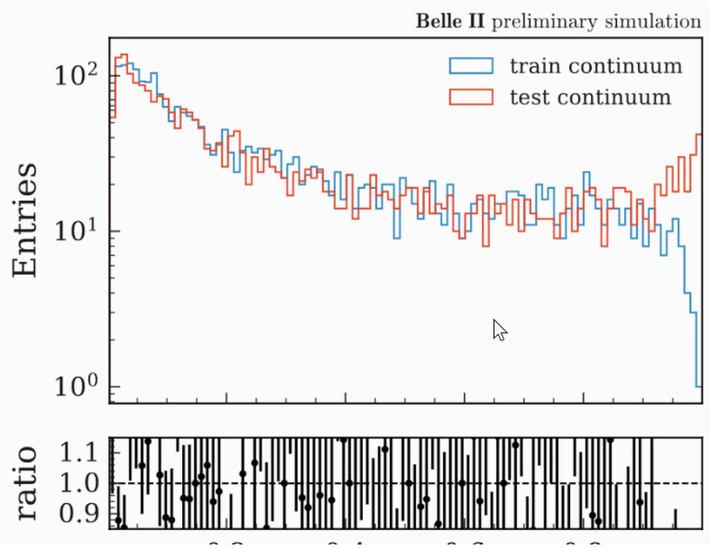
实际使用MVA时的一些经典问题—个人答案

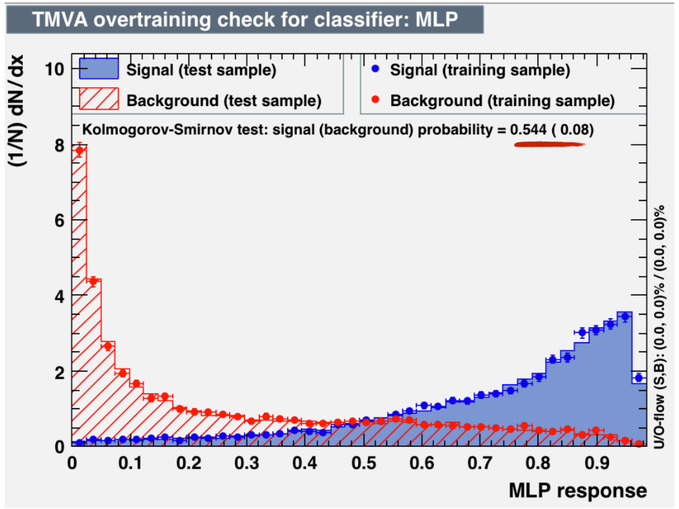
- 我需要调整超参数么?
- 用哪些变量作为input?
- 如何判断是否overtraining?
  - 使用独立的test sample。
- 如何使用output?
- 什么时候用MVA?
- 如何筛选input variables?
- 报告/note里放啥信息?
- …
- More questions?

overtraining 的例子



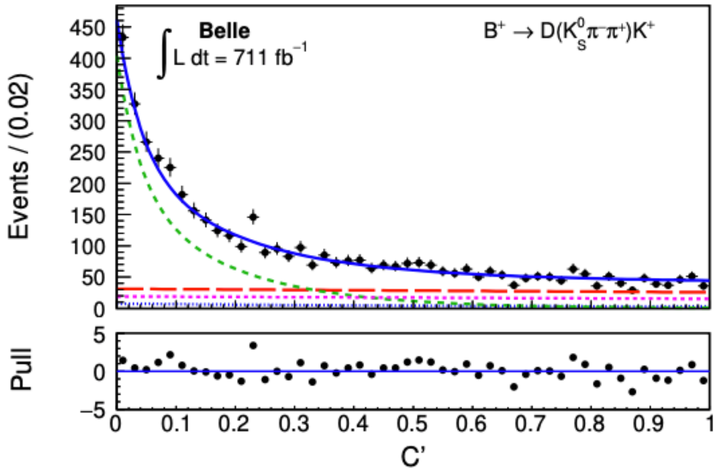
G. Cowan

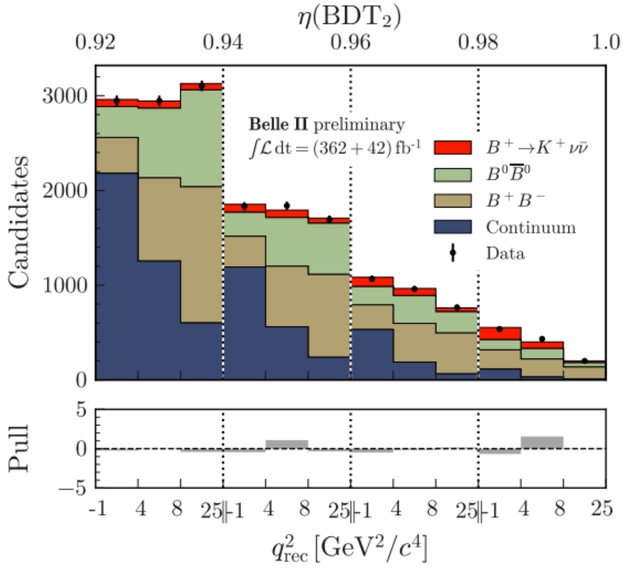




实际使用MVA时的一些经典问题—个人答案

- 我需要调整超参数么?
- 用哪些变量作为input?
- 如何判断是否overtraining?
- 如何使用output?
  - 一般情况,Cut, 来提高信噪比
  - 某些情况，也可以用来做fit变量
- 什么时候用MVA?
- 如何筛选input variables?
- 报告/note里放啥信息?
- …
- More questions?

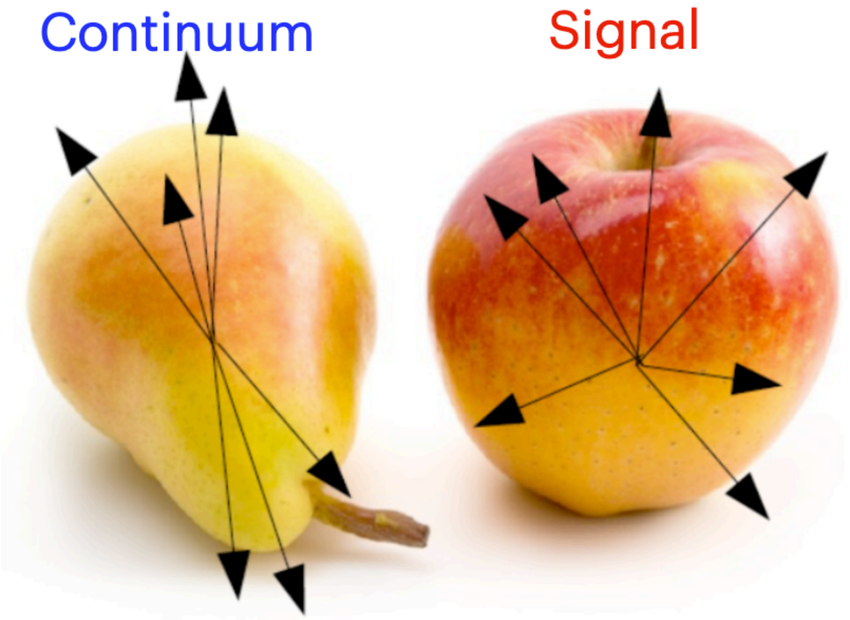




$K^+\nu\bar{\nu}$

实际使用MVA时的一些经典问题—个人答案

- 我需要调整超参数么?
- 用哪些变量作为input?
- 如何判断是否overtraining?
- 如何使用output?
- 什么时候用MVA?
  - 当你研究分类问题，有多个变量能提供区分度。
  - 比如qqbar v.s. BBbar, many event shape variables
- 如何筛选input variables?
- 报告/note里放啥信息?
- …
- More questions?



实际使用MVA时的一些经典问题—个人答案

- 我需要调整超参数么?
- 用哪些变量作为input?
- 如何判断是否overtraining?
- 如何使用output?
- 什么时候用MVA?
- 如何筛选input variables ?
  - 理想情况下(MC模拟无限精确, training sample无限大, BDT参数无限多): 韩信点兵
  - 实际情况下, 保持**BDT性能不变**下精简input variables:
    - 算法可以直接给出变量重要度(参考); 先去掉不重要变量
    - 通过control sample获得data/MC对比. 避免MC不准确带来的“虚假”优性能。也能减少BDT的系统误差。
- 报告/note里放啥信息?
- …
- More questions?

实际使用MVA时的一些经典问题—个人答案

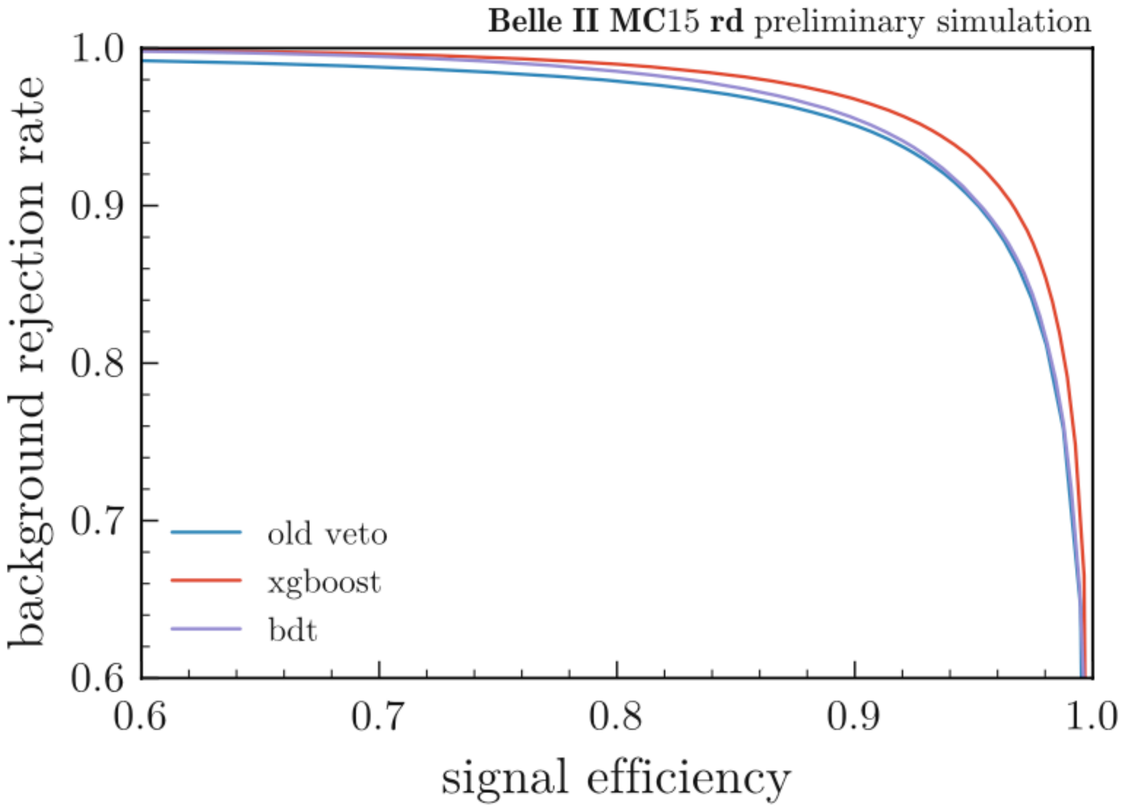
- 我需要调整超参数么?
- 用哪些变量作为input?
- 如何判断是否overtraining?
- 如何使用output?
- 什么时候用MVA?
- 如何筛选input variables?
- 报告/note里放啥信息?
  - Basic: N(training), N(test); final input variables list; input variables plots (signal v.s. background); output plots (train v.s. test); correlation of input variables;
- …
- More questions?

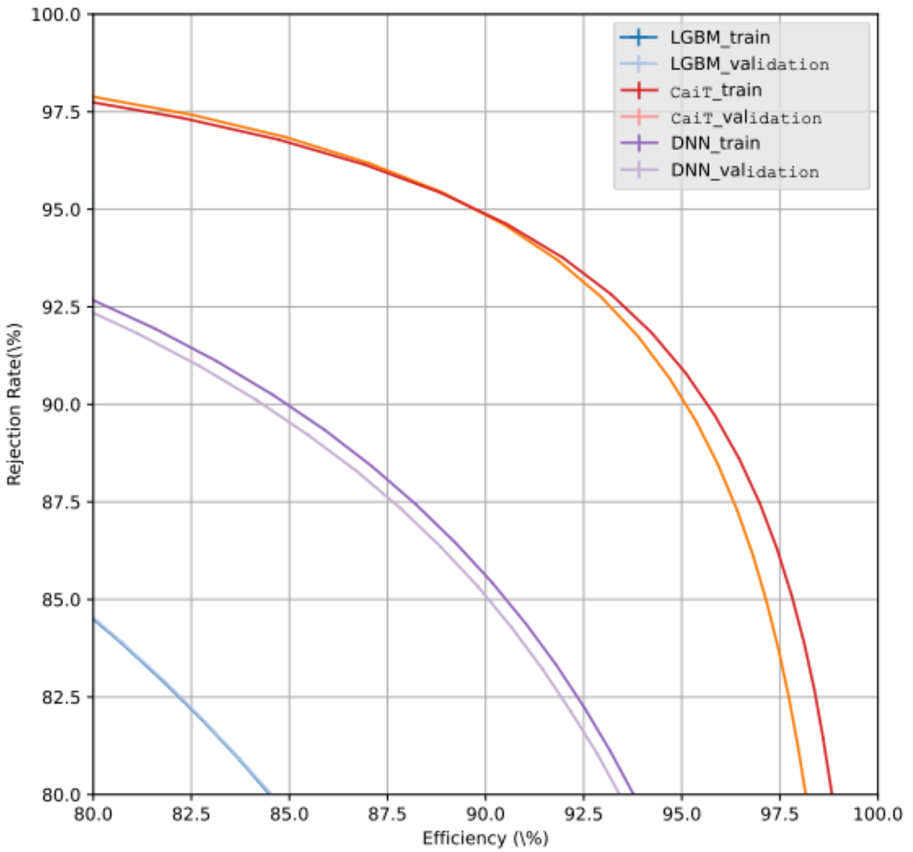
实际使用MVA时的一些经典问题—个人答案

- 我需要调整超参数么?
- 用哪些变量作为input?
- 如何判断是否overtraining?
- 如何使用output?
- 什么时候用MVA?
- 如何筛选input variables?
- 报告/note里放啥信息?
- Best Candidate selection (BCS) v.s. BDT, 谁先谁后
  - Always, after all selections, do best candidate selection:
    - 可能加完BDT cut后，多重数就不多了，不用做BCS了
    - 如果先做BCS, 正确的组合可能会被错误的淘汰掉，但错误的可能会被BDT先cut掉
- …
- More questions?

Rather than BDT

有时候 其他算法能给出更好的结果





- XGBoost v.s. BDT

- Transformer v.s. DNN v.s. LGBM
- 变形金刚完胜！

Scripts

“渔”:Belle II 官方教程

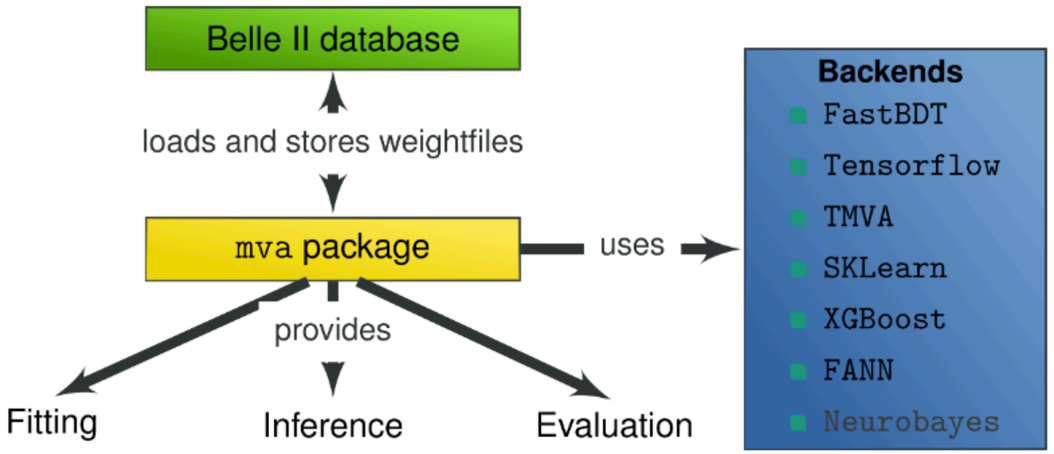
8. B2BII
9. Background Package
10. Calibration
11. Decay Files
12. The Belle II Event Display
13. Event Generators
14. Tools for Validation of the SoftwareTrigger
15. KLM (\(K_{L}^0\) and Muon Detector)
16. Belle II File Format
17. MVA package
18. PXD
19. Reconstruction
20. Simulation
21. Skims
22. SVD
23. Tracking
24. TRG
25. Tools for Physics Validation of the Software
26. Fitting training
27. Software development
28. How to document your code with Sphinx
29. Software publications

17. MVA package

Tip

For a hands-on introductory lesson with the MVA package, see Continuum suppression using Boosted Decision Trees.

17.1. Overview



17.1.1. Main goals

The mva package was introduced to provide:

link

我个人用的“鱼”

- https://gitlab.desy.de/belle2/analyses/b2chinaschool

Beyond MVA — AI-agent/LLM

Anthropic's Claude Opus 5.5 suggested a likely variant that gives a smooth interpolation with link holonomies that equal precisely the corresponding lattice link variables and such that the interpolated gauge field in a hypercube depend only on the link variables on the boundary of that hypercube. Pick coordinates on $\mathbb{R}^4$ so that the lattice consists of the

There are many possible variants of this construction, some with possibly desirable properties. For example, Opus 5.5 also described a variant in which the inductive step is made by a sort of linear interpolation rather than by solving a Dirichlet problem.

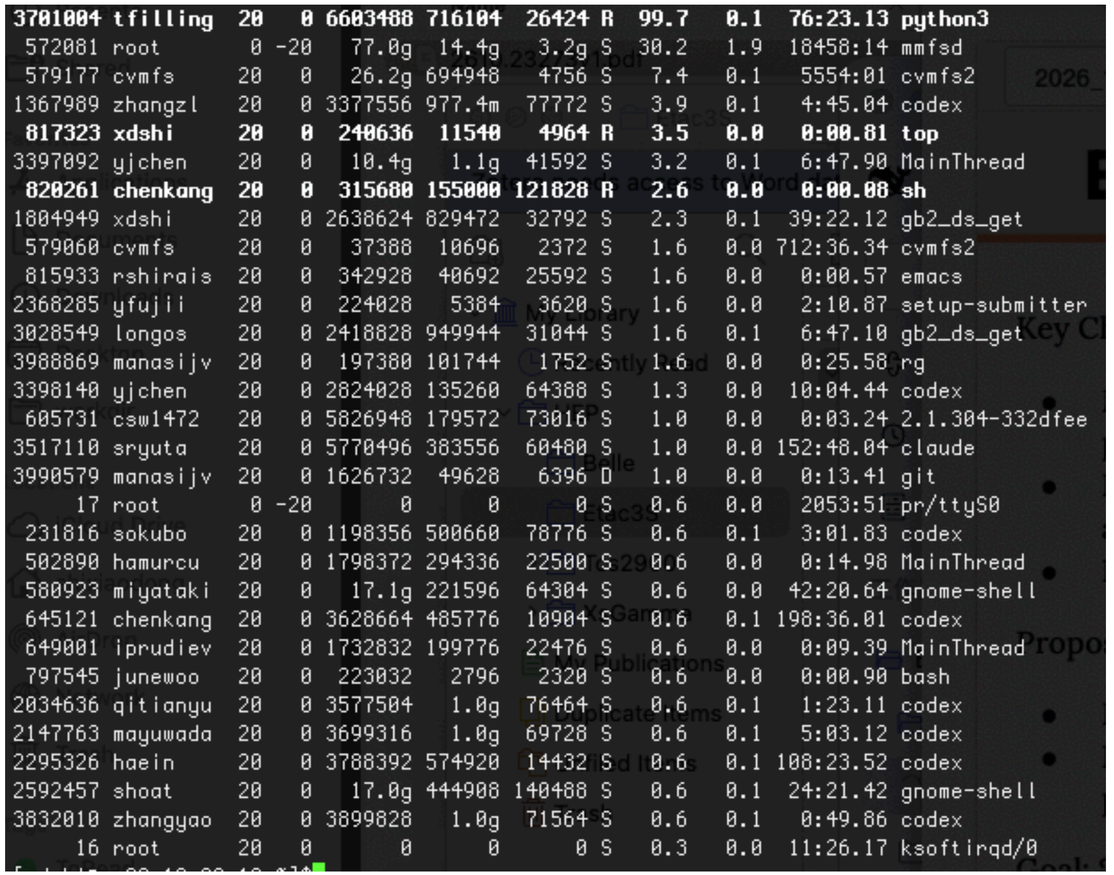
2026 年 11 月 3 日，著名理论物理学家、菲尔兹奖得主 Edward Witten 在 arXiv 发布新论文《A Note On Lattice Definition of Gauge Theories With Chiral Fermions》，研究如何为包含手征费米子的规范理论建立非微扰的格点定义。

Witten 在论文致谢中明确表示，Claude Opus 5.5 为研究中的许多技术细节提供了协助。

在论文第三节讨论格点规范场向连续时空的光滑插值时，Witten 特别提到，Claude Opus 5.5 提出了一种可能的改进构造方案。这意味着 AI 的参与可能已经不局限于文献检索、语言润色或常规计算，而是开始涉及具体数学构造的探索。

论文：Edward Witten, A Note On Lattice Definition of Gauge Theories With Chiral Fermions, arXiv:2611.05051 (2026).

Witten with Calude



3701004	tfilling	20	0	6603488	716104	26424	R	99.7	0.1	76:23.13	python3
572081	root	0	-20	77.0g	14.4g	3.2g	S	30.2	1.9	18458:14	mmfsd
579177	cvmfs	20	0	26.2g	694948	4756	S	7.4	0.1	5554:01	cvmfs2
1367989	zhangzl	20	0	3377556	977.4m	77772	S	3.9	0.1	4:45.04	codex
817323	xdshi	20	0	240636	11540	4964	R	3.5	0.0	0:00.81	top
3397092	yjchen	20	0	10.4g	1.1g	41592	S	3.2	0.1	6:47.90	MainThread
820261	chenkang	20	0	315680	155000	121828	R	2.6	0.0	0:00.08	sh
1804949	xdshi	20	0	2638624	829472	32792	S	2.3	0.1	39:22.12	gb2_ds_get
579060	cvmfs	20	0	37388	10696	2372	S	1.6	0.0	712:36.34	cvmfs2
815933	rshirais	20	0	342928	40692	25592	S	1.6	0.0	0:00.57	emacs
2368285	yfujii	20	0	224028	5384	3620	S	1.6	0.0	2:10.87	setup-submitter
3028549	longos	20	0	2418828	949944	31044	S	1.6	0.1	6:47.10	gb2_ds_get
3988869	manasijv	20	0	197380	101744	1752	S	1.6	0.0	0:25.58	rg
3398140	yjchen	20	0	2824028	135260	64388	S	1.3	0.0	10:04.44	codex
605731	csw1472	20	0	5826948	179572	73016	S	1.0	0.0	0:03.24	2.1.304-332dfee
3517110	sryuta	20	0	5770496	383556	60480	S	1.0	0.0	152:48.04	claude
3990579	manasijv	20	0	1626732	49628	6396	D	1.0	0.0	0:13.41	git
17	root	0	-20	0	0	0	S	0.6	0.0	2053:51	pr/ttyS0
231816	sokubo	20	0	1198356	500660	78776	S	0.6	0.1	3:01.83	codex
502890	hamurcu	20	0	1798372	294336	22500	S	0.6	0.0	0:14.98	MainThread
580923	miyataki	20	0	17.1g	221596	64304	S	0.6	0.0	42:20.64	gnome-shell
645121	chenkang	20	0	3628664	485776	10904	S	0.6	0.1	198:36.01	codex
649001	iprudiev	20	0	1732832	199776	22476	S	0.6	0.0	0:09.39	MainThread
797545	junewoo	20	0	223032	2796	2320	S	0.6	0.0	0:00.90	bash
2034636	qitianyu	20	0	3577504	1.0g	76464	S	0.6	0.1	1:23.11	codex
2147763	mayuwada	20	0	3699316	1.0g	69728	S	0.6	0.1	5:03.12	codex
2295326	haein	20	0	3788392	574920	14432	S	0.6	0.1	108:23.52	codex
2592457	shoat	20	0	17.0g	444908	140488	S	0.6	0.1	24:21.42	gnome-shell
3832010	zhangyao	20	0	3899828	1.0g	71564	S	0.6	0.1	0:49.86	codex
16	root	20	0	0	0	0	S	0.3	0.0	11:26.17	ksoftirqd/0

One day on kekcc, many codex/claude…

11 月 2 日，麻省理工学院（MIT）计算机科学与人工智能实验室（CSAIL）的一项研究在技术圈引发热议。

研究团队让 GPT-6 Astra 直接分析并修改另一个 AI 模型 Qwen3.5-0.8B 的神经网络权重。在没有使用传统训练方法的情况下，Astra 仅调整了 0.016% 的参数，就让 Qwen 的数学测试成绩提高了约 6 倍。

更令人意外的是，Astra 不仅能够修改已有模型，还能通过分析权重判断小型神经网络的用途。它甚至从零设计并写出一套神经网络参数，让它在小游戏中击败经过 1 亿步强化学习训练的模型。

研究团队将这种能力称为“权重空间编程”（Weight-Space Programming）。相关帖子在 X 平台现已获得超过 700 万次浏览，迅速成为刷屏 AI 技术社区的热门话题。

- LLM 甚至可以直接改权重。。
- “黑箱子”不再不可见。。
- 很难想象/预料 未来with LLM。。

> 很多核心业务有险可守。开始大家以为 AI 对搜索影响很大，又觉得 Google 好像顶得住；到今年再看，浪真的已经杀过来。其他行业、内容产业也许觉得安全，但从今天的分享能看到浪越来越近。不动起来，就会慌。远处浪潮已经吞没一个个山丘，你不能还坐着欣赏。要着急、站起来，准备新打法，找到高地，搬物资、上船，从陆战转海战。已有的陆军部队里，谁能上船、能不能打海战，得提前准备。坦率说现在都有点晚，不能再悠闲地觉得安全。

- 也许粒子物理不用过于焦虑