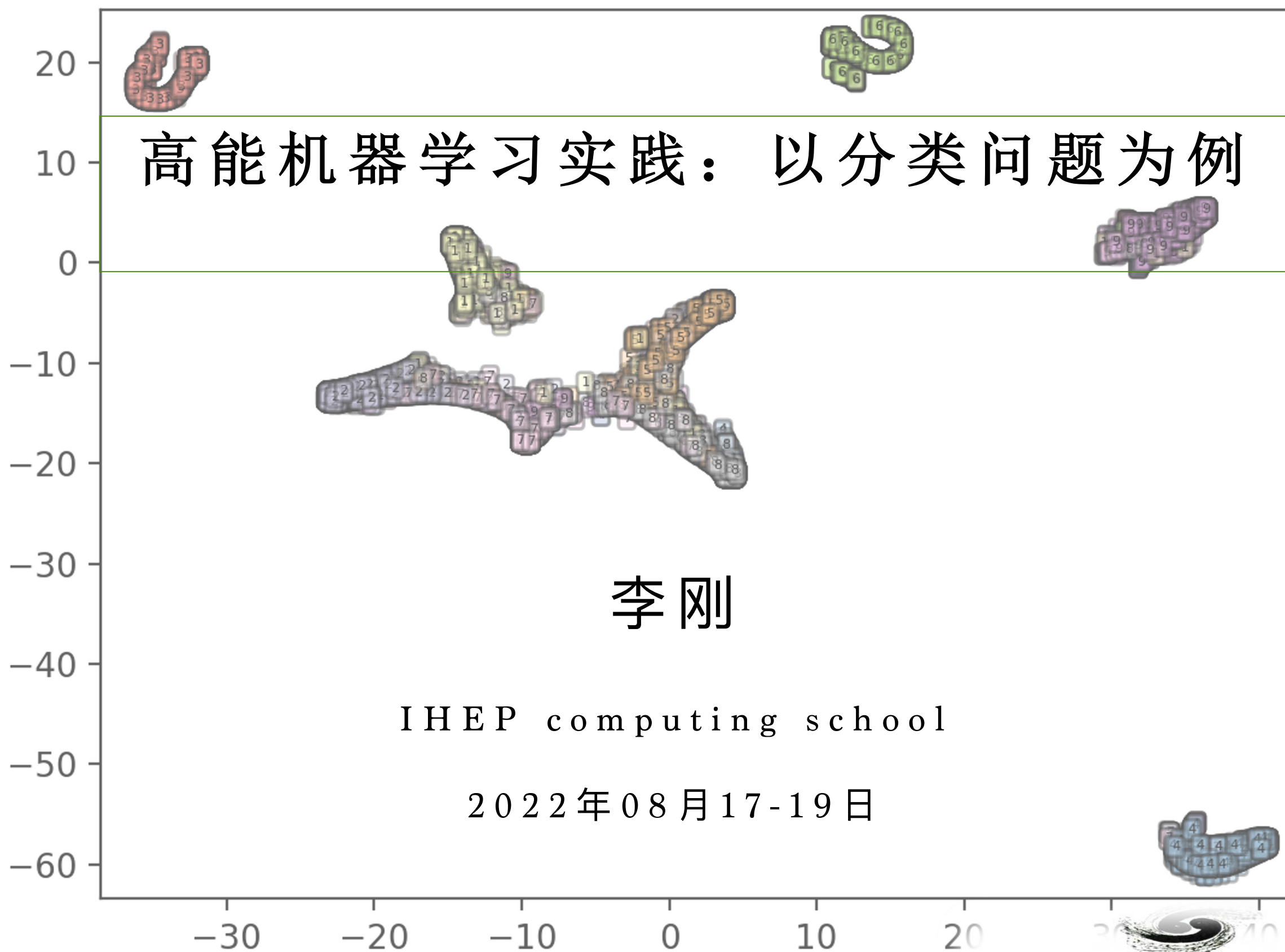


t-SNE (7.3e+02 sec)



高能机器学习实践：以分类问题为例

李刚

I H E P computing school

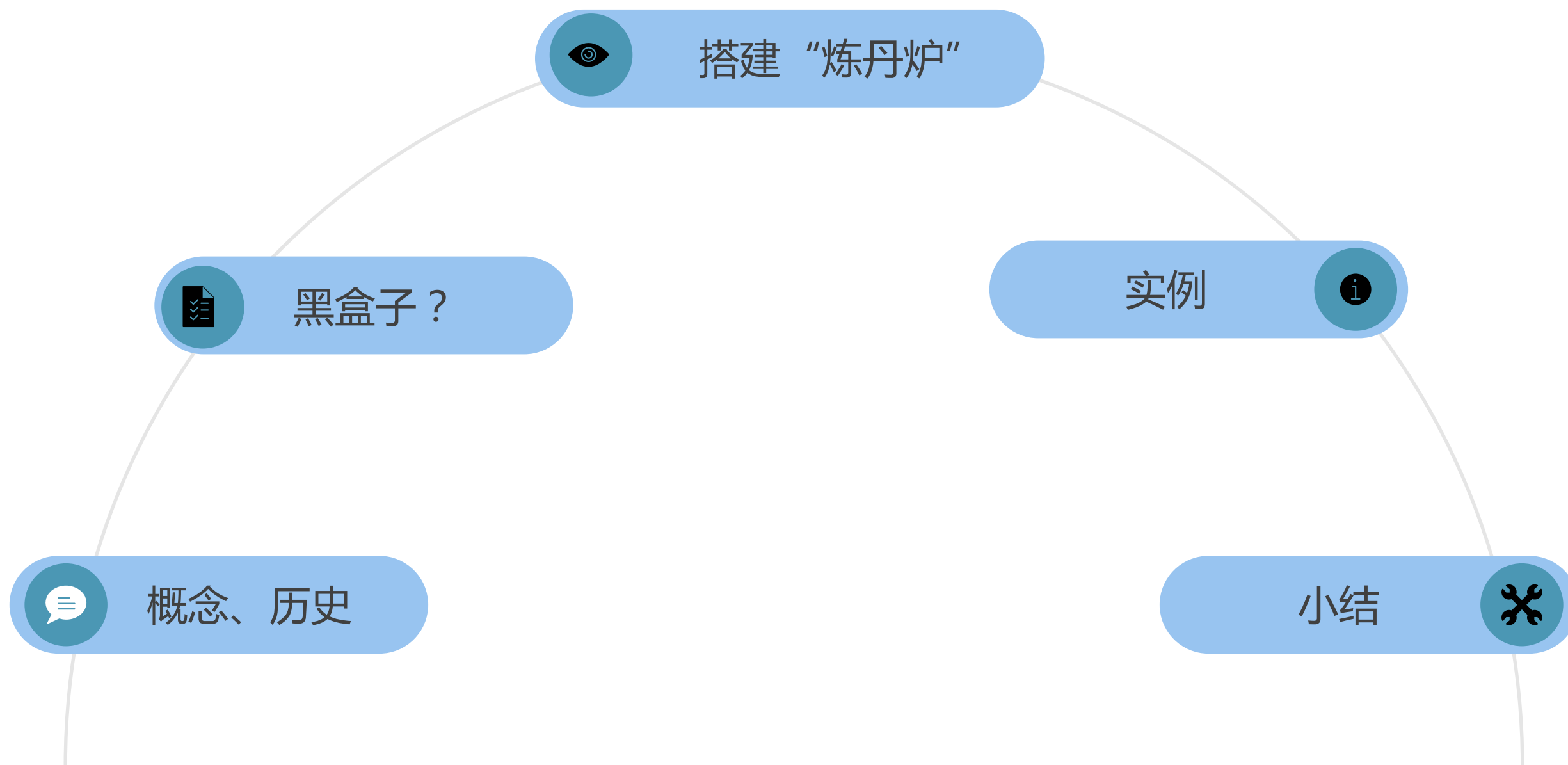
2022年08月17-19日

刘北江老师的报告是很好的 review，非常全面，
大家能得到比较系统的认识。
而我将会从比较具体的点出发。

我个人对机器学习的理解和用机器学习做分类的实例

- 机器学习一般介绍
- 环境搭建
- 跑例子
- 如果大家能复现别人的结果，我的目的就达到了。

目录



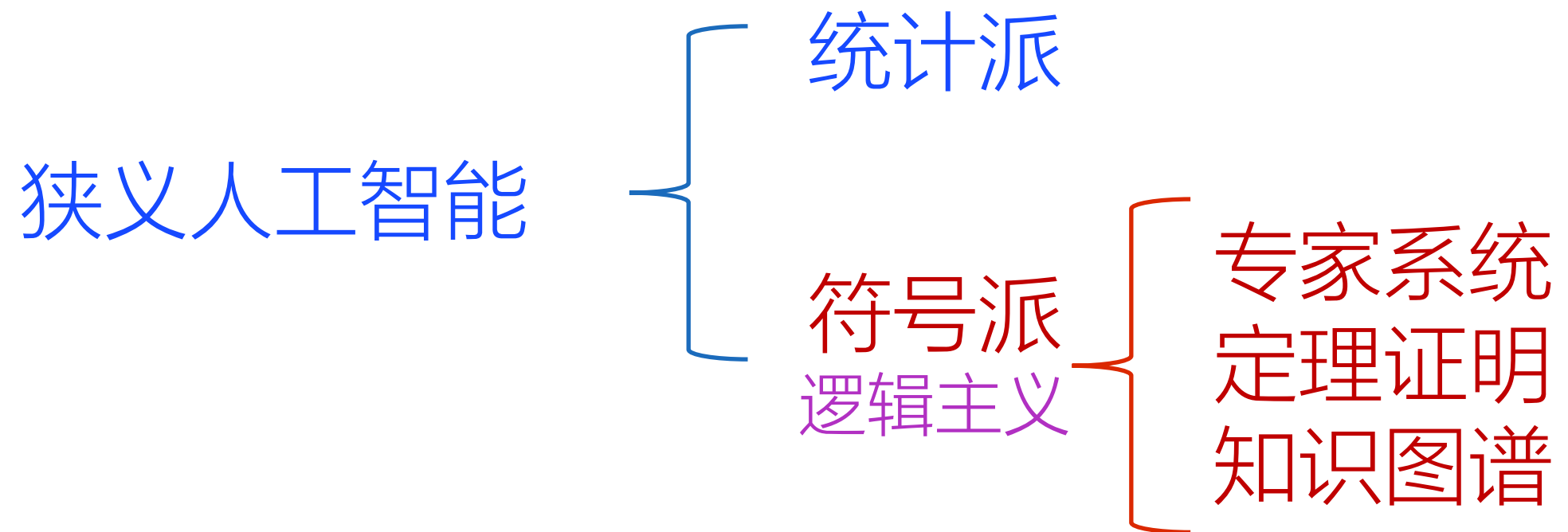
人工智能 (AI)

像人一样行动：图灵测试

像人一样思考：认知建模

合理的思考：逻辑主义

合理地行动：为了实现最佳期望而行动

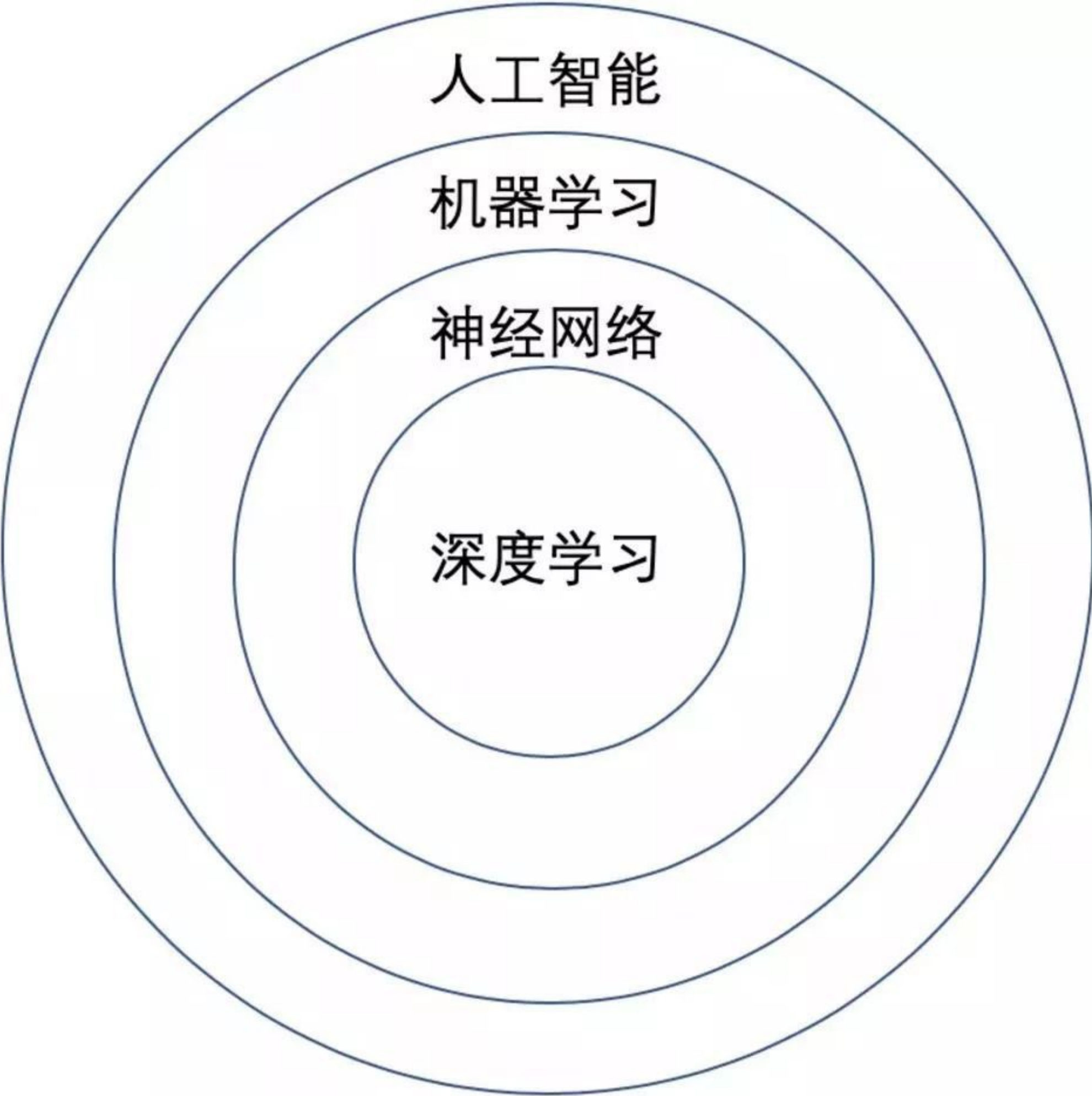


机器学习？实现人工智能的一种途径

推理 → 知识 → 学习

神经网络和深度学习？

是一种以人工神经网络为架构，
对资料进行表征学习的算法



A Venn diagram consisting of four concentric circles. The innermost circle is labeled '深度学习' (Deep Learning). The next circle out is labeled '神经网络' (Neural Networks). The third circle is labeled '机器学习' (Machine Learning). The outermost circle is labeled '人工智能' (Artificial Intelligence). This illustrates that Deep Learning is a subset of Neural Networks, which is a subset of Machine Learning, which is a subset of Artificial Intelligence.

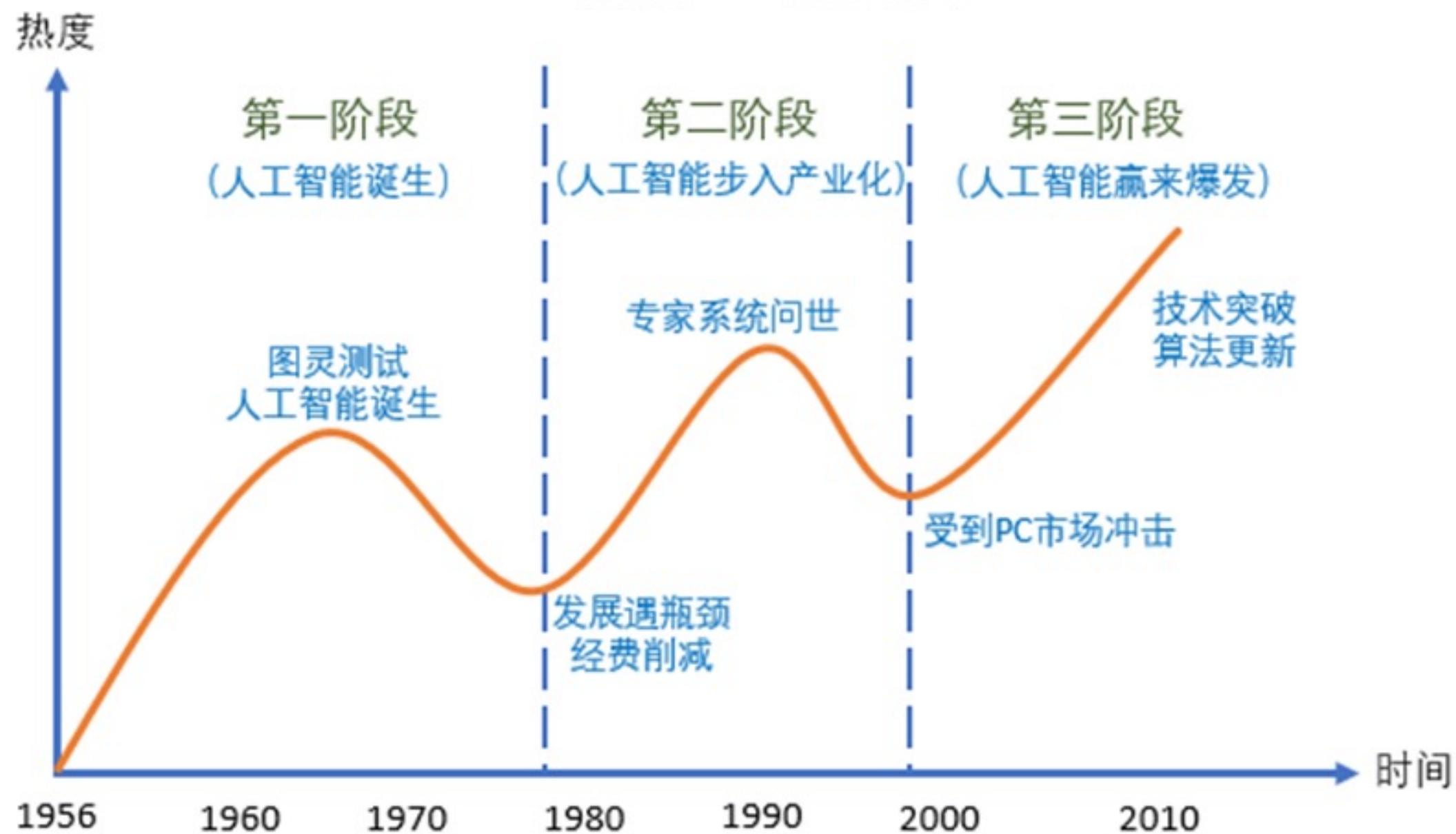
人工智能

机器学习

神经网络

深度学习

人工智能产业发展历程



神经网络

- McCulloch & Pitts (1943)
- Rosenblatt (1962)
- Paul Werbos(1974)
- Hopfield (1982)
- Hinton(2006)
-

神经网络的开山祖师“扫地僧”

Walter Pitts
1923-1969

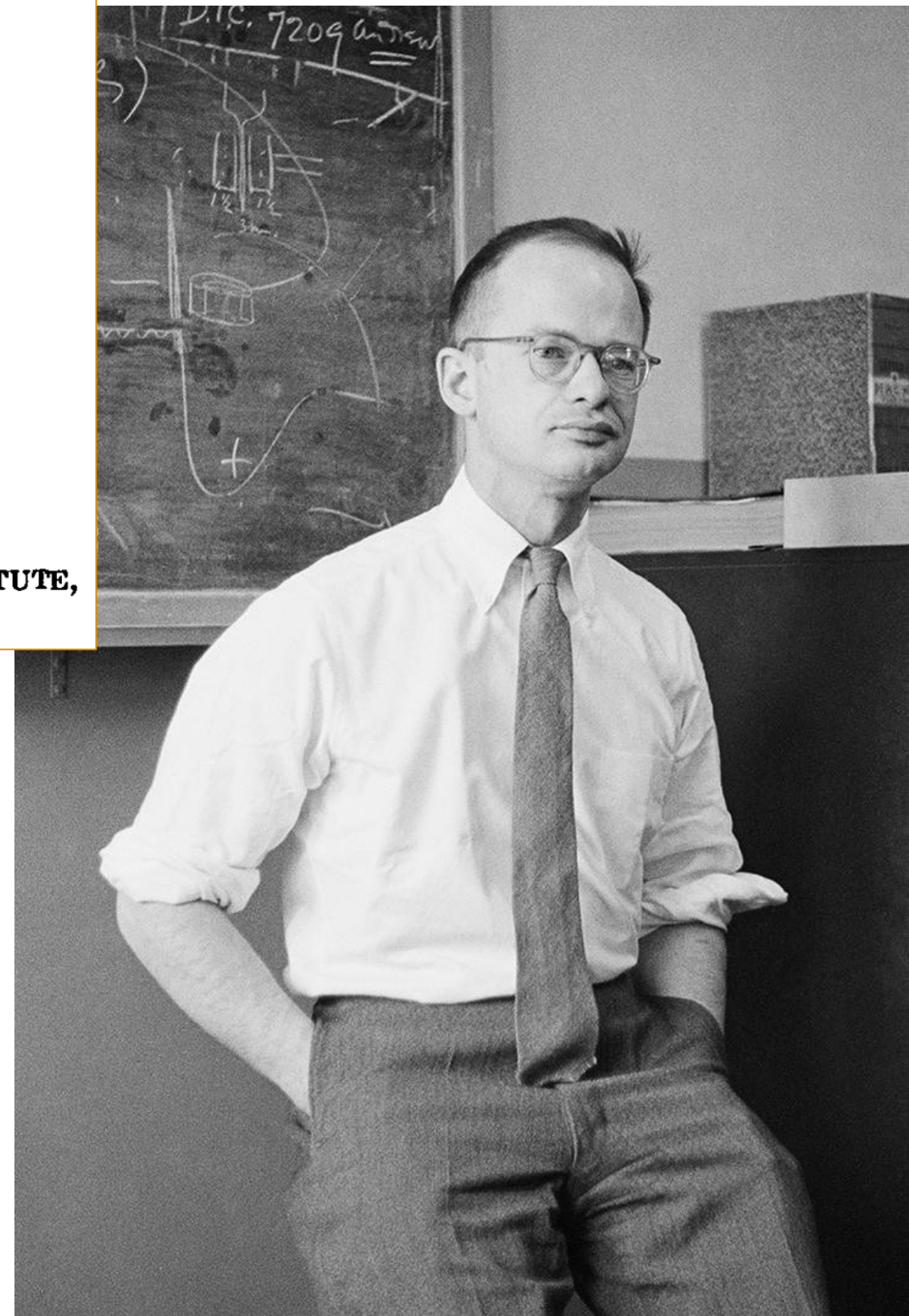
BULLETIN OF
MATHEMATICAL BIOPHYSICS
VOLUME 5, 1943

A LOGICAL CALCULUS OF THE IDEAS IMMANENT IN NERVOUS ACTIVITY

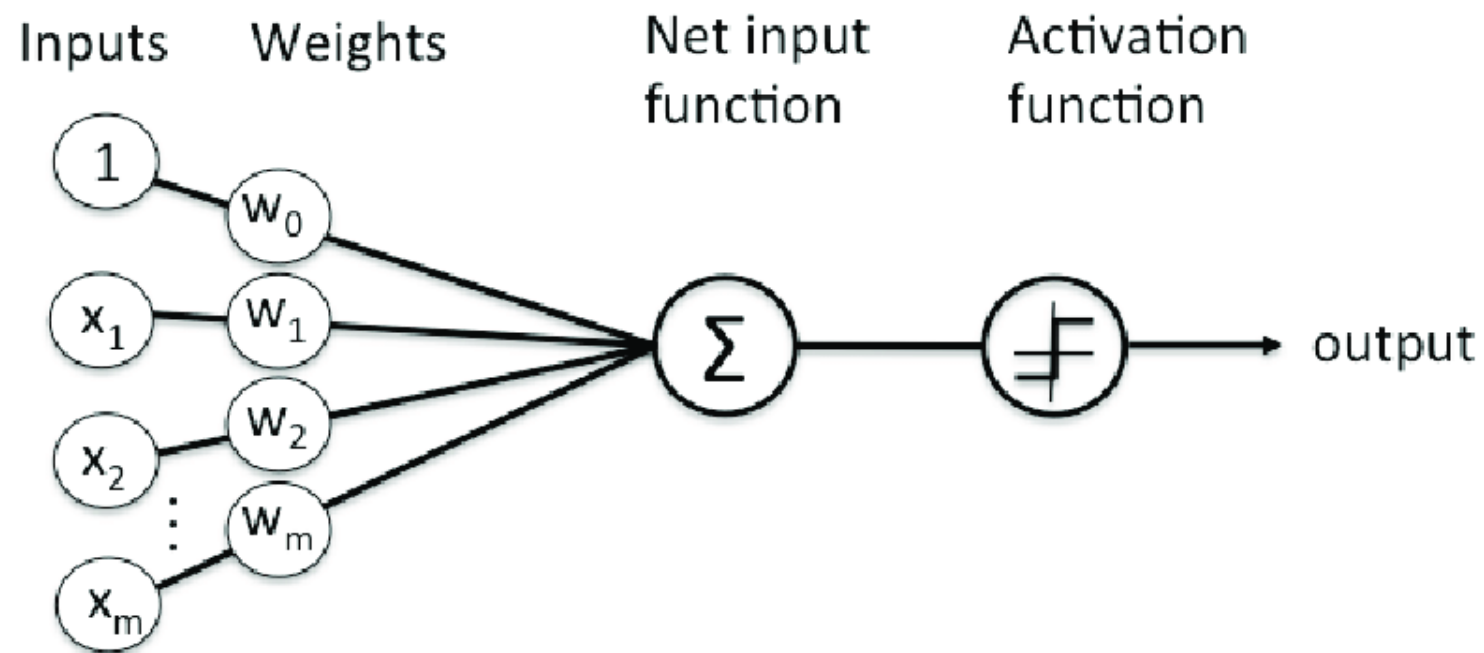
WARREN S. MCCULLOCH AND WALTER PITTS

FROM THE UNIVERSITY OF ILLINOIS, COLLEGE OF MEDICINE,
DEPARTMENT OF PSYCHIATRY AT THE ILLINOIS NEUROPSYCHIATRIC INSTITUTE,
AND THE UNIVERSITY OF CHICAGO

- 数学天才15岁离家出走
- 在芝加哥大学扫地
- 维纳的“特别”学生
- 1943：20岁与McCulloch合作发表 ANN 奠基之作
- 拒绝芝加哥大学的学位
- 46 岁去世

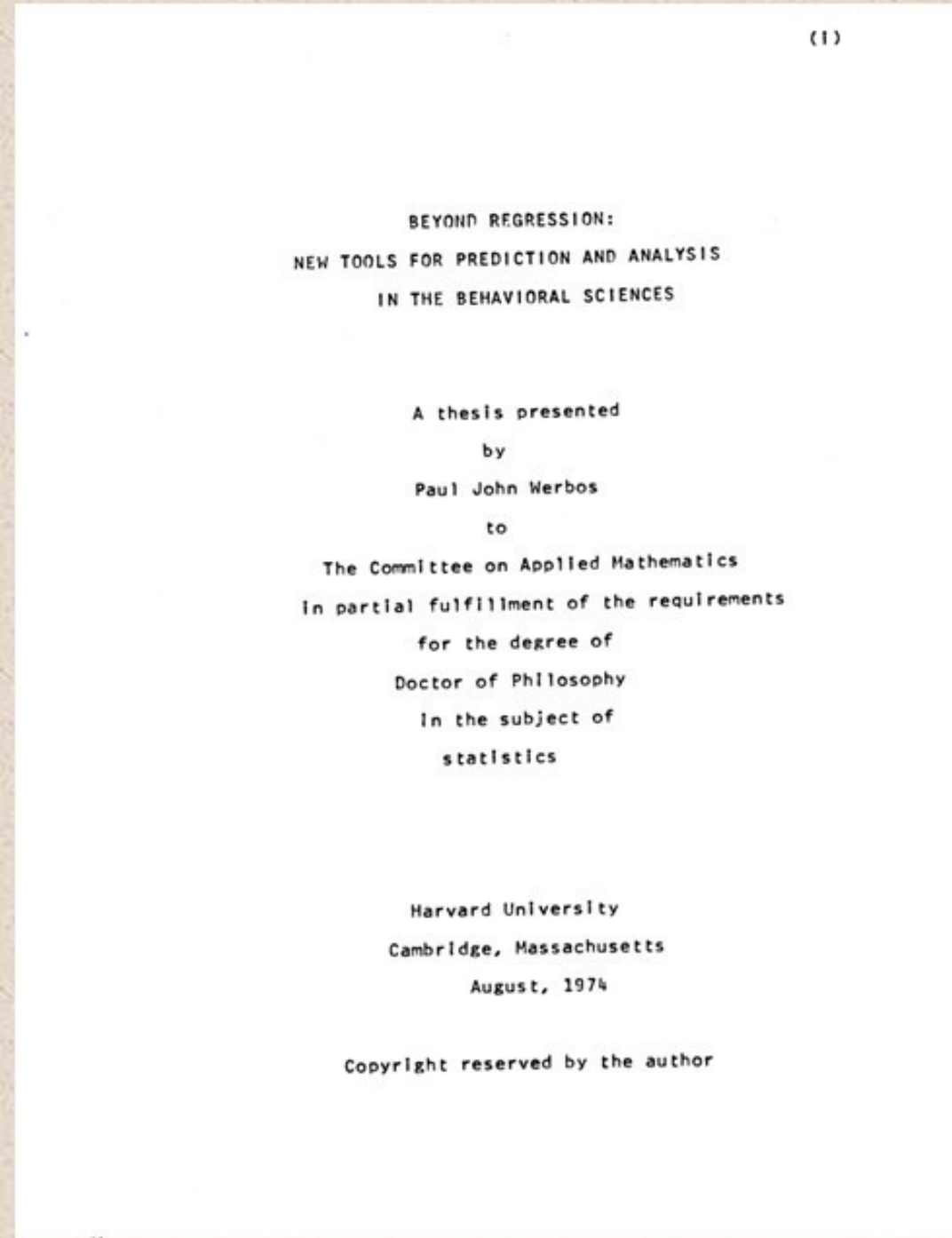


过犹不及：Rosenblatt



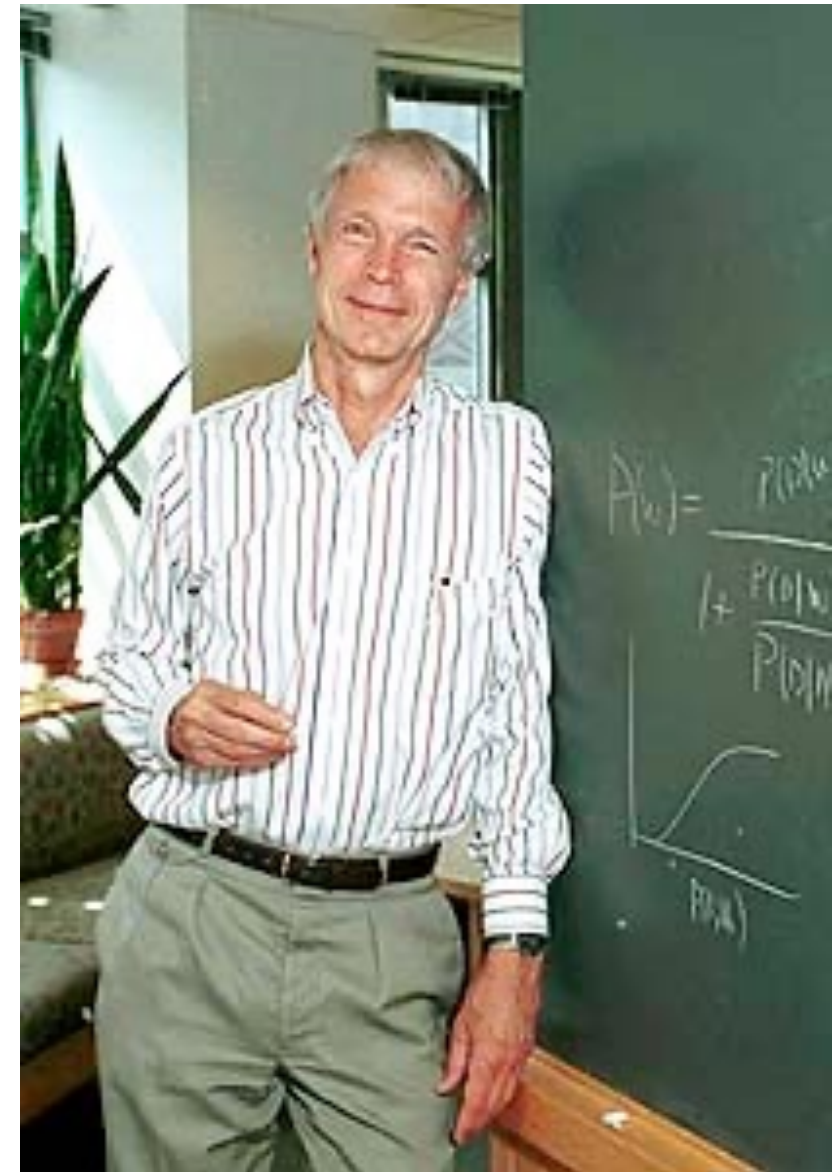
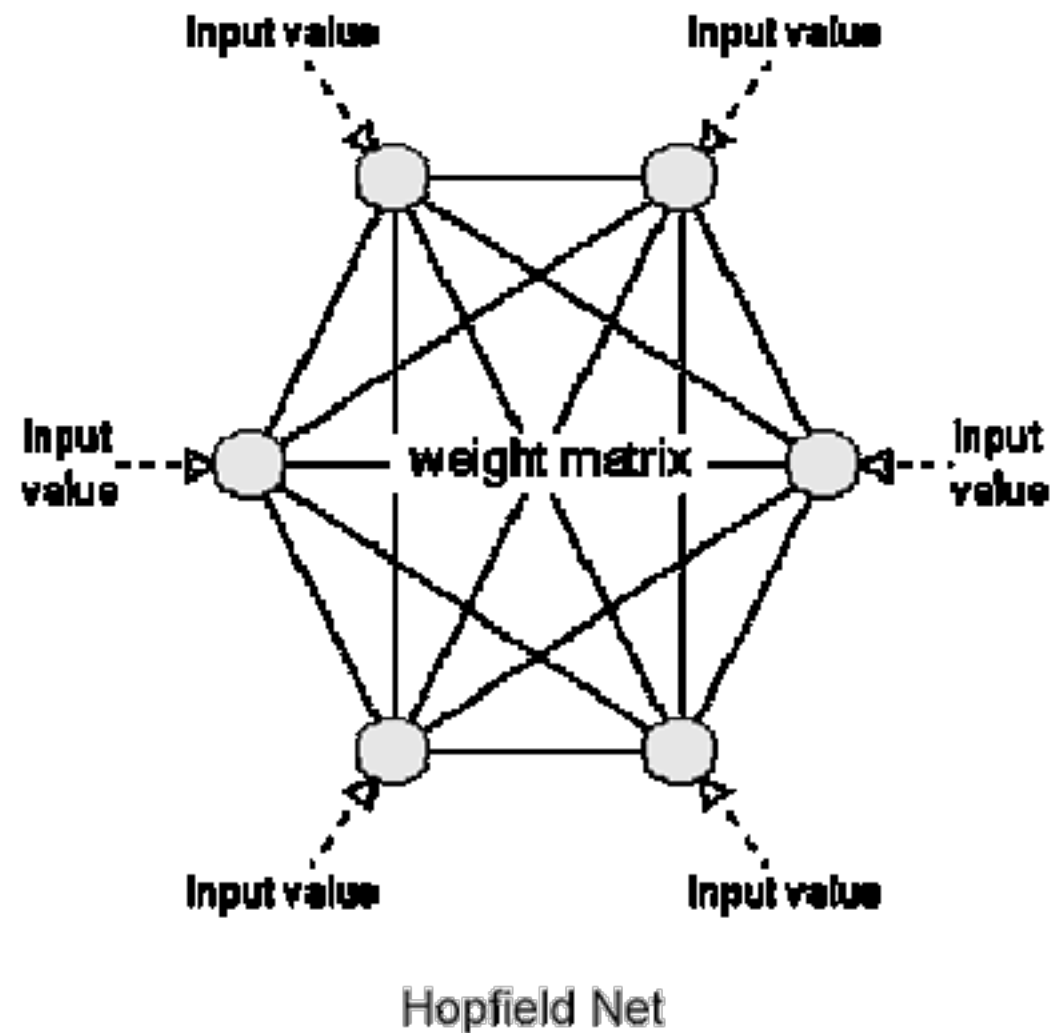
- 感知机 Perceptron
- 能完成一些简单的视觉处理任务
- 在理论上证明了单层神经网络在处理线性可分的模式识别问题时，可以收敛
- (他的高中校友)明斯基：无法处理异或（XOR）问题。 去世于43 岁生日当天。
- Bronx 中学毕业（8 个诺奖，6 个普利策奖，1 个图灵奖）

好的研究也得赶上好时机 (Werbos)



- 发明了反向传播(BACK PROPAGATION), 可以解决XOR问题。
- 但在1974年并未引起重视。
- 神经网络低谷时期, 文章不合时宜。

回归物理：Hopfield



Ising model of a neural network as a memory model is first proposed in 1974, acknowledged by Hopfield in his 1982 paper.

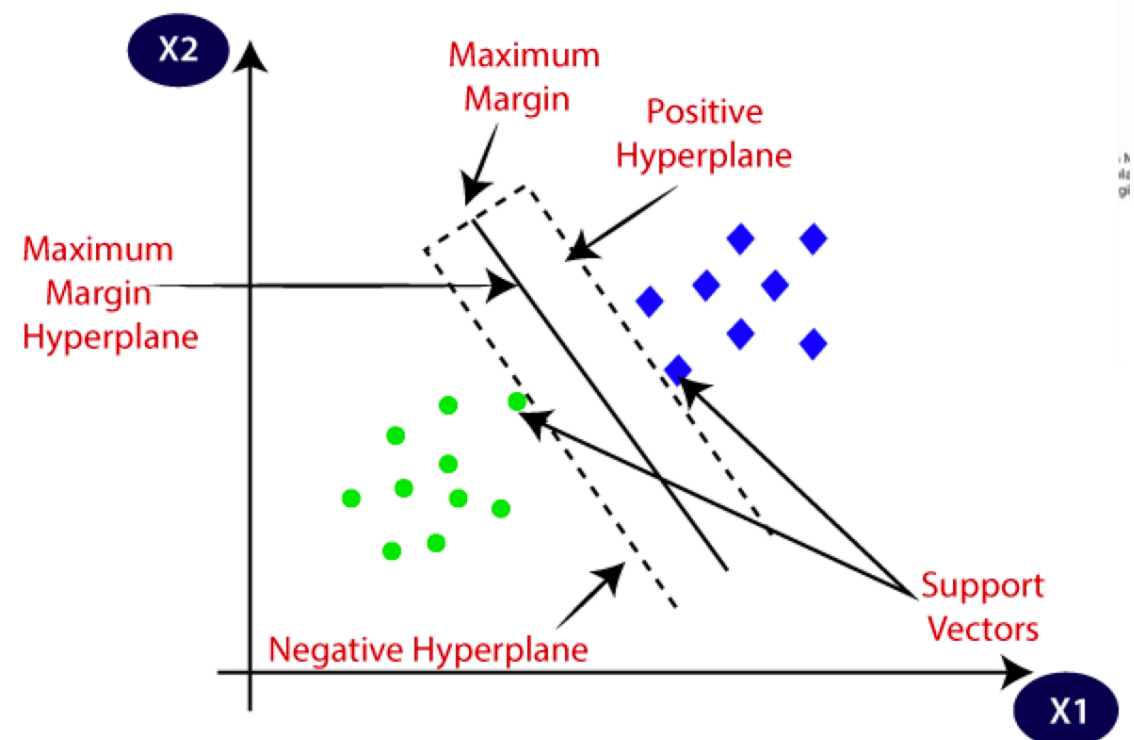
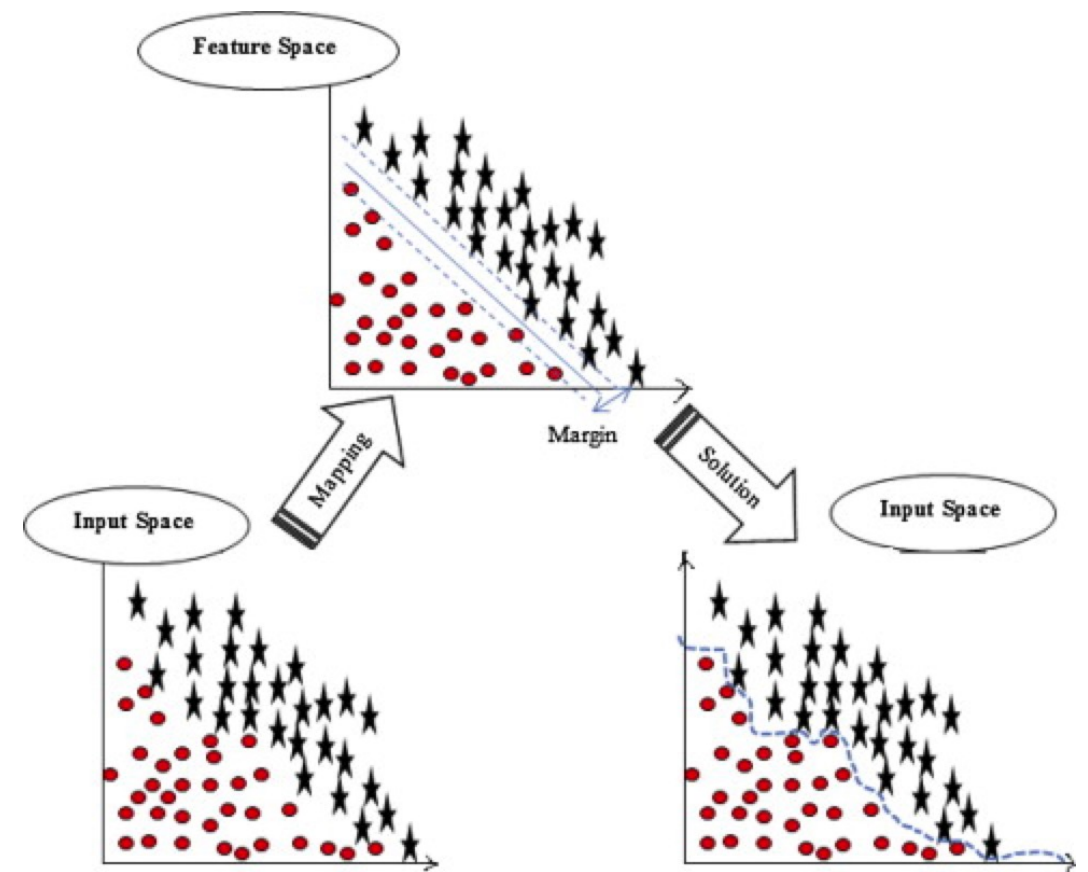
可以解决一大类模式识别和组合优化问题，但是
风头被互联网的兴起盖过。

90年代的“潜伏”

互联网掩盖了霍老带来的神经网络热度，却也为“王者回归”积累了“大数据”

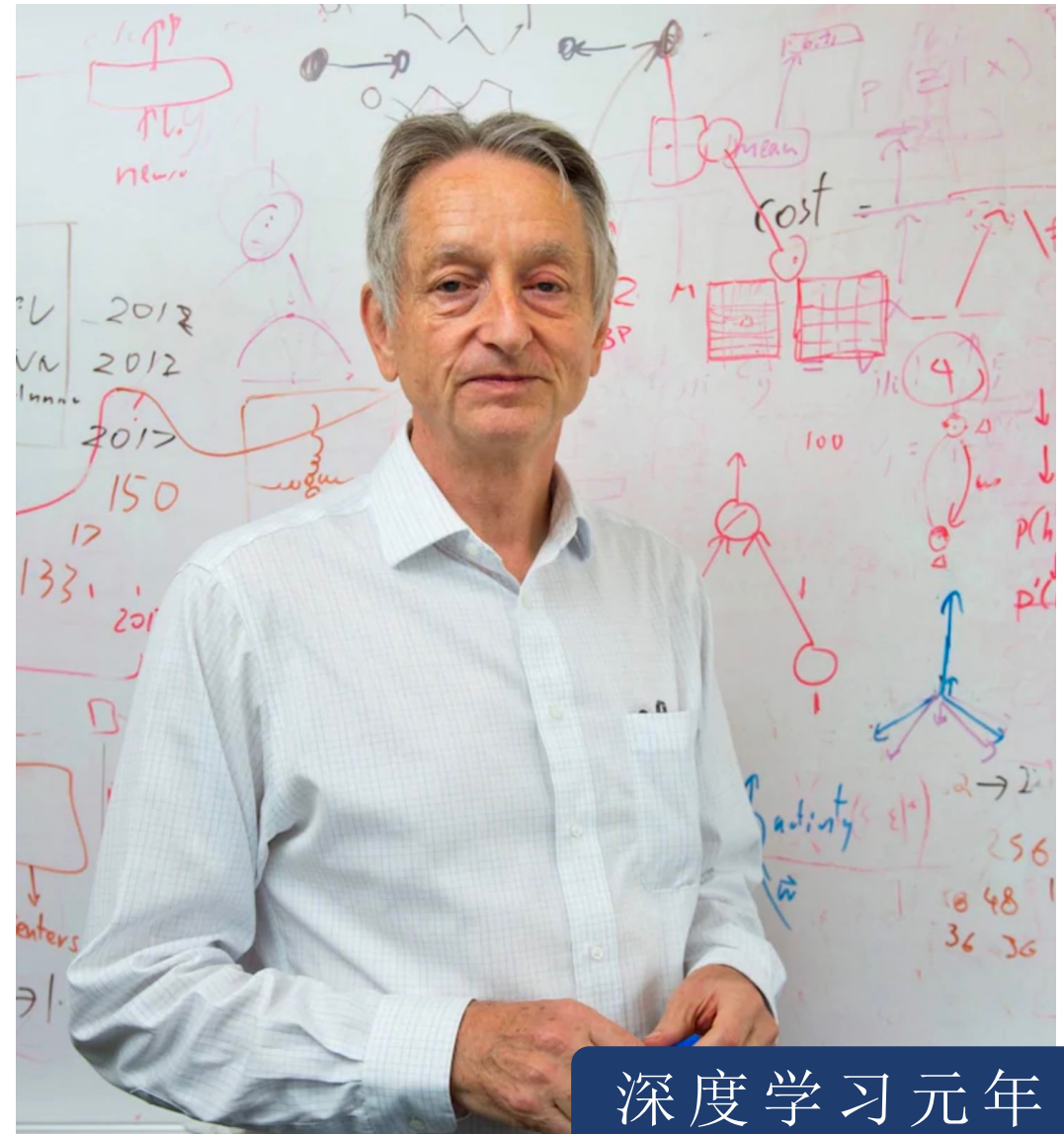
8、90年代出现了一些新的算法

- 比如：Vladimir Vapnik et al., 支持矢量机 (Support Vector Machine, SVM), 1992
- 大家觉得这些算法也挺好，比如 SVM，BDT，kNN，k-Means
- 不一定要用 NN，当时 NN 训练困难、性能无明显优势
- 进入低潮期，但是一些人还在坚持
- 新的契机：大数据和硬件的发展

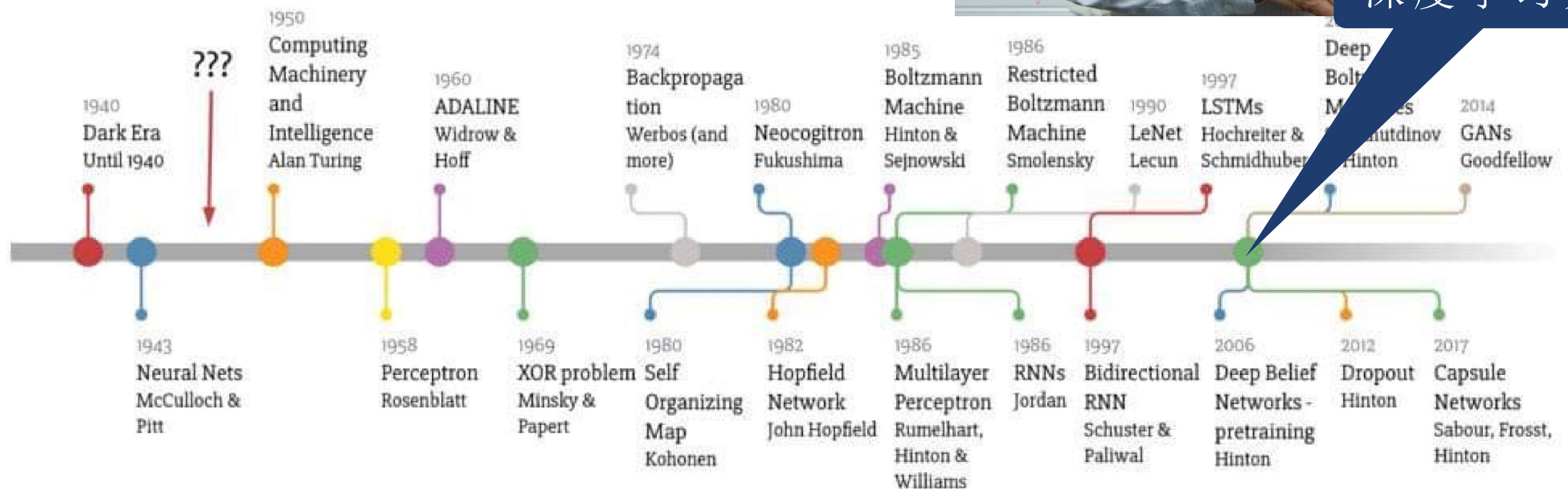


大数据和硬件的加持：(Hinton)

- 深度学习：降维、逐层预训练
- 2012年ILSVRC大赛领先第二名10%以上：错误率26%→15%
- 关键词“加拿大”



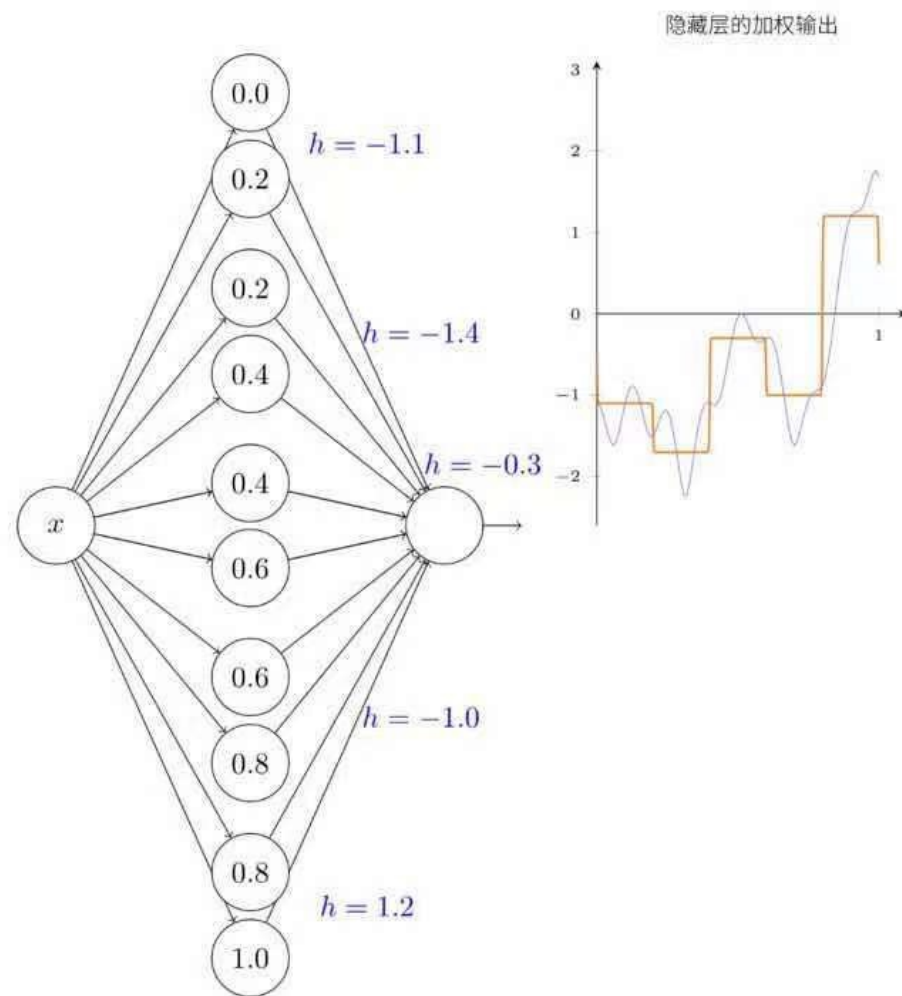
深度学习元年



机器学习不是黑盒子，但也不是万能的！

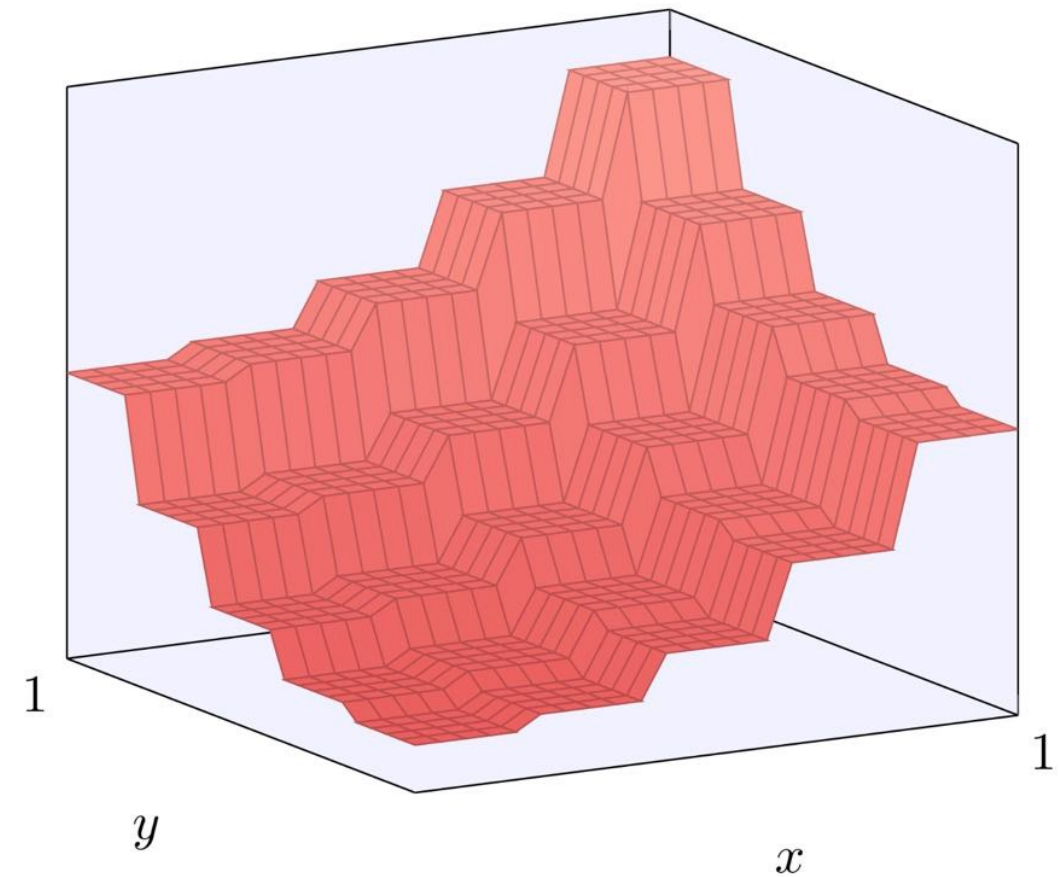
神经网络：万能函数展开器

Universal function approximator



一维

Many towers

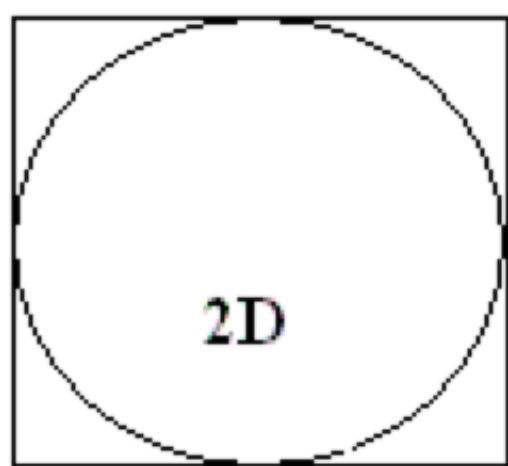


二维

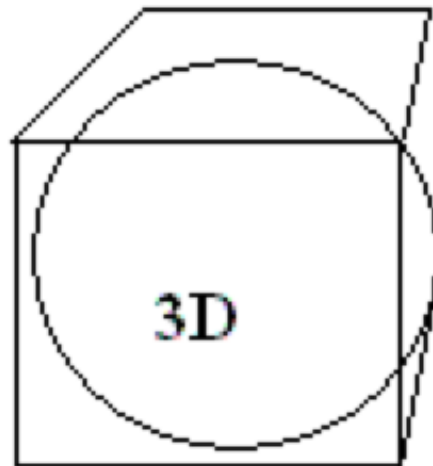
神经网络的一个显著事实：
可以以任意精度近似一个连续函数，如果隐藏层的神经元足够多的话 ...

维度诅咒

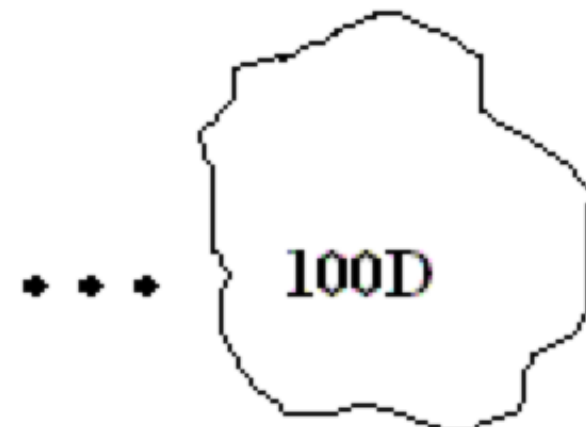
Curse of dimensionality



ratio: $4/\pi = 1.27$

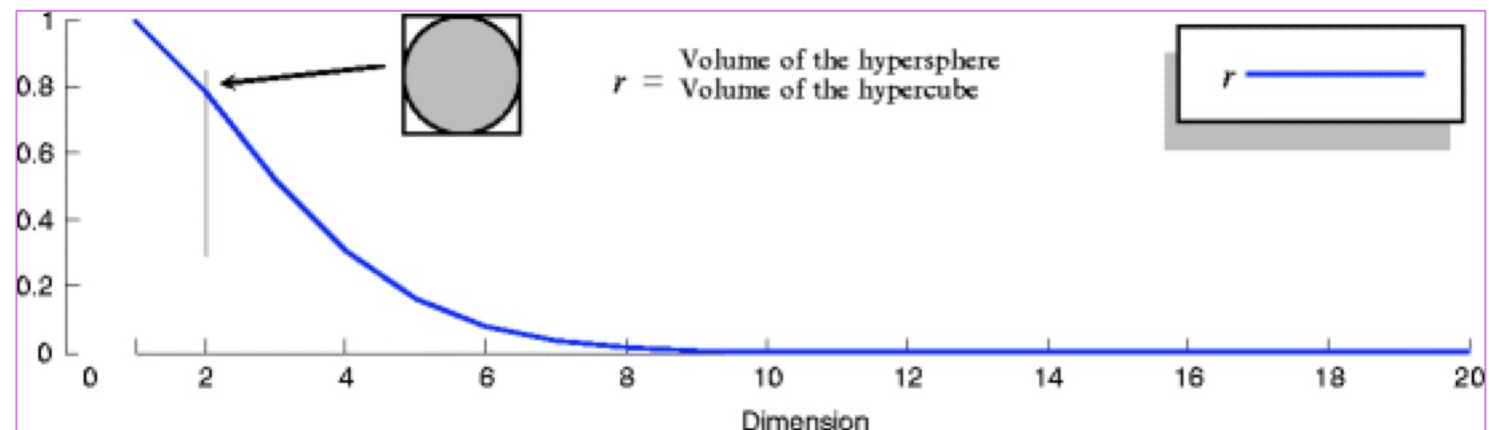


ratio: $6/\pi = 1.91$



ratio: $4.2 \cdot 10^{39}$

$$\frac{A_{\text{circle}}}{A_{\text{square}}} = \frac{\pi}{4} \text{ for } d = 2$$
$$\frac{V_{\text{sphere}}}{V_{\text{cube}}} = \frac{\pi}{6} \text{ for } d = 3$$
$$\frac{V_{\text{hypersphere}}}{V_{\text{hypercube}}} = \frac{\pi^{d/2}}{d 2^{d-1} \Gamma(d/2)} \rightarrow 0 \text{ as } d \rightarrow \infty$$



- D=1, 100个平均分布的点能把一个单位区间以距离不超过0.01采样;
- D=10, 则需要 10^{20} 个采样点才能达到同样的采样率。

机器学习的主要对象是高维空间的数据，
而高维空间没有中心，都是“角落”！

降维，
且大多具体问题可以降维！

模型没有好坏，但对具体问题有“偏好”

No Free Lunch Theorem

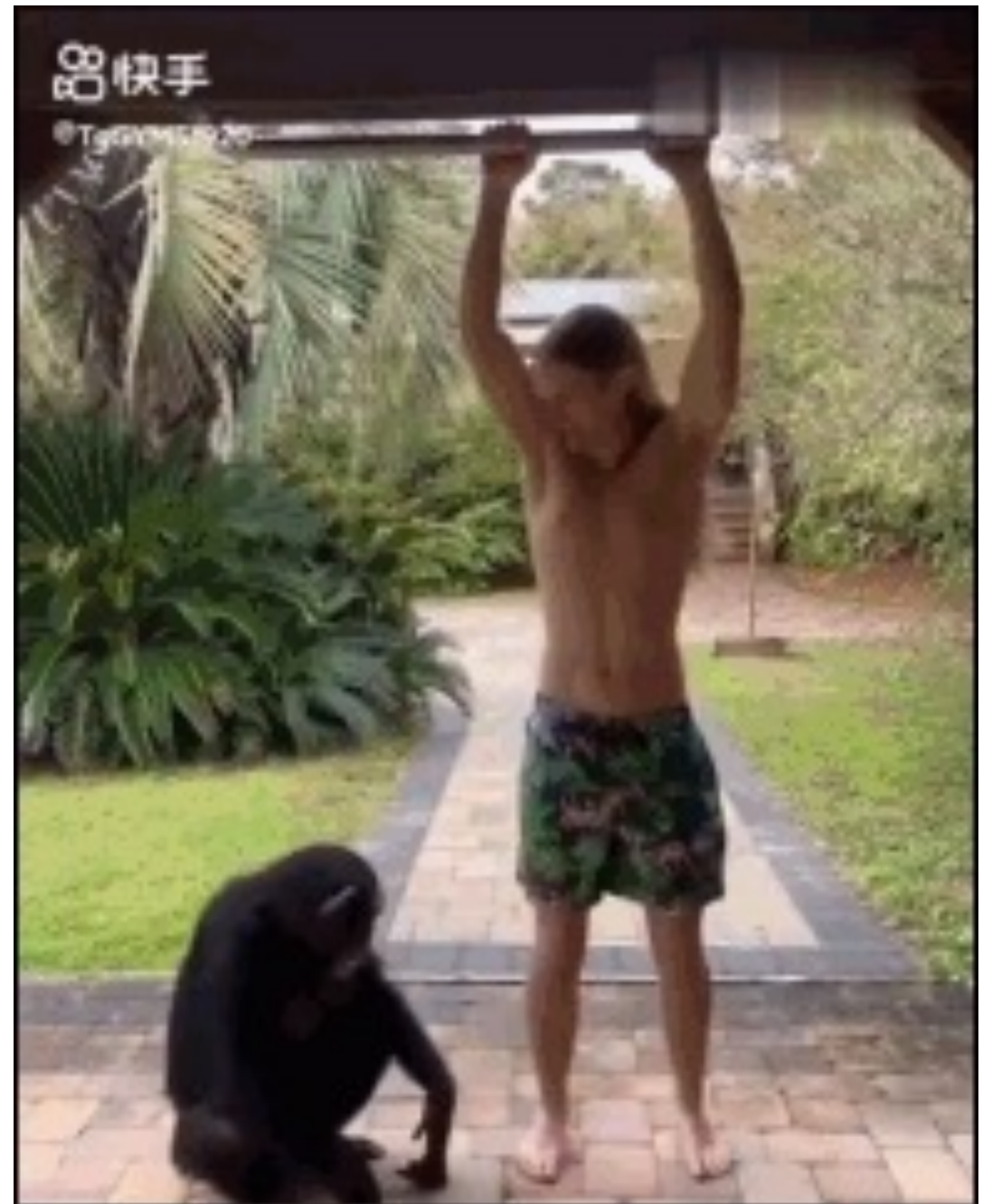
What does dinner cost? David Wolpert

- 这是个严肃的数学定理
- 可以严格证明（假设：
数据集是全集）

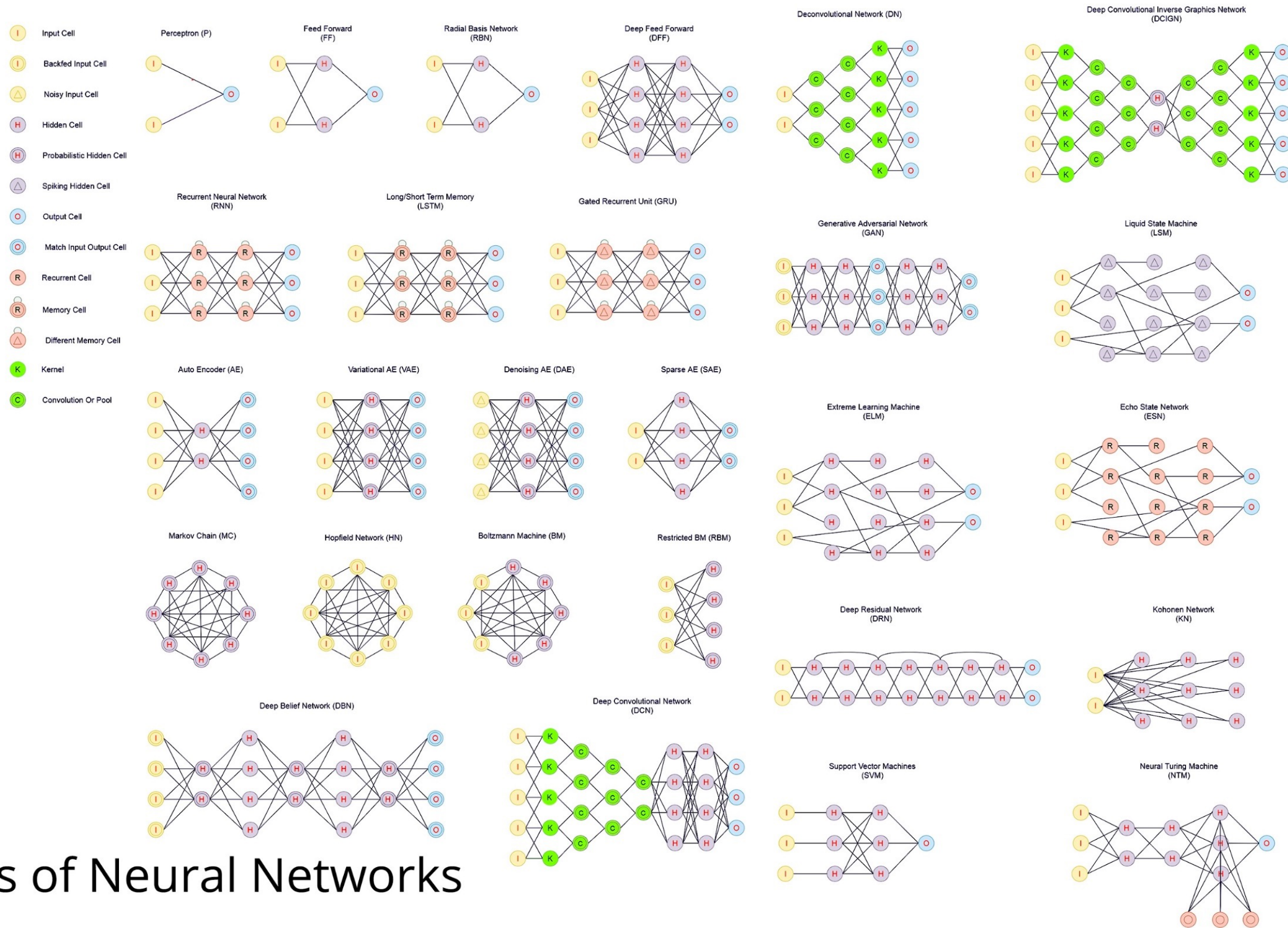
$$\sum_f P(d_m^Y | f, m, a) = \sum_f P(d_m^Y | f, m, a')$$

$$\forall a, a', d_m$$

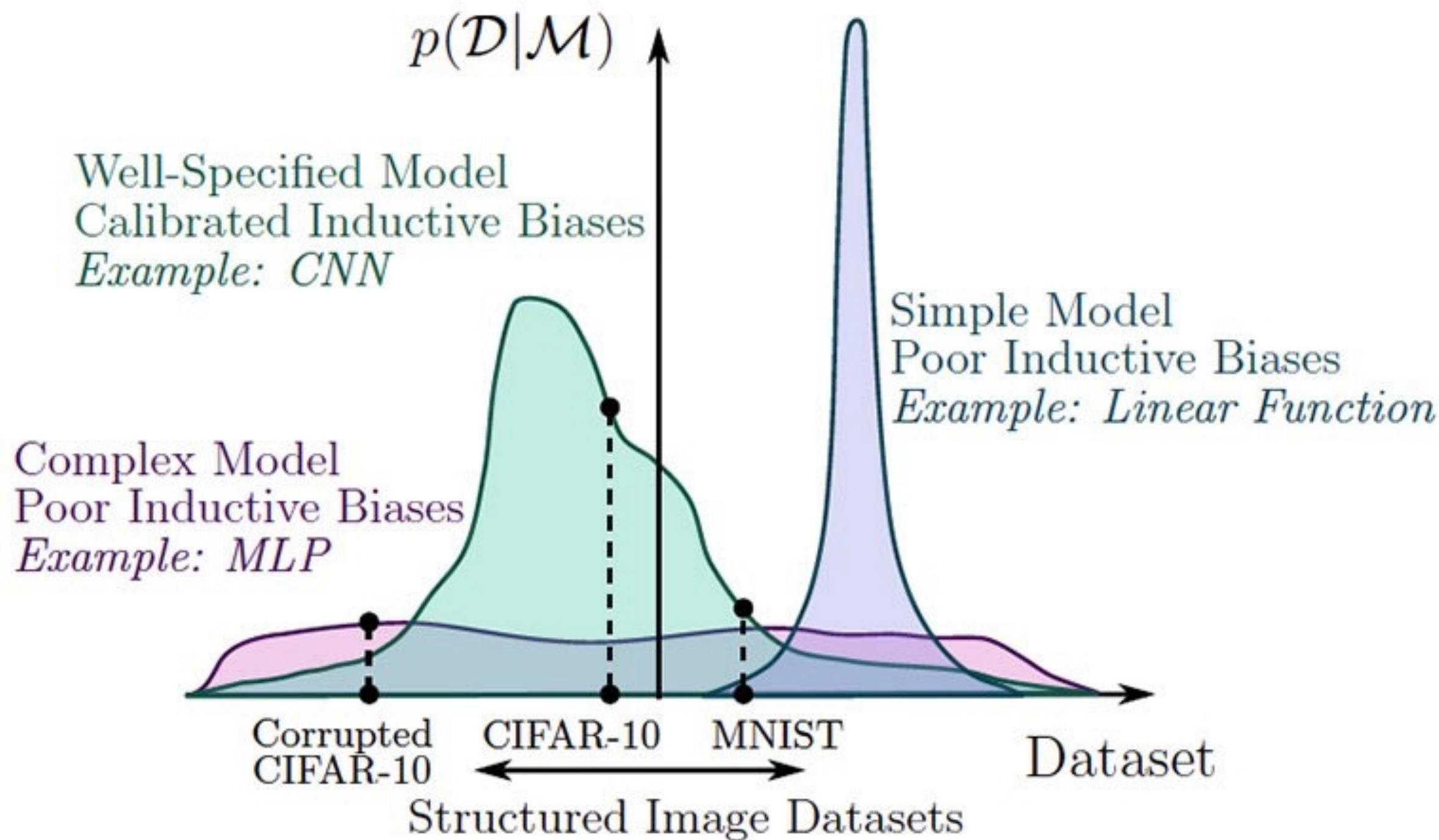
- 但具体问题的数据一个子集
- 关键：合适的算法



我们只找合适的算法

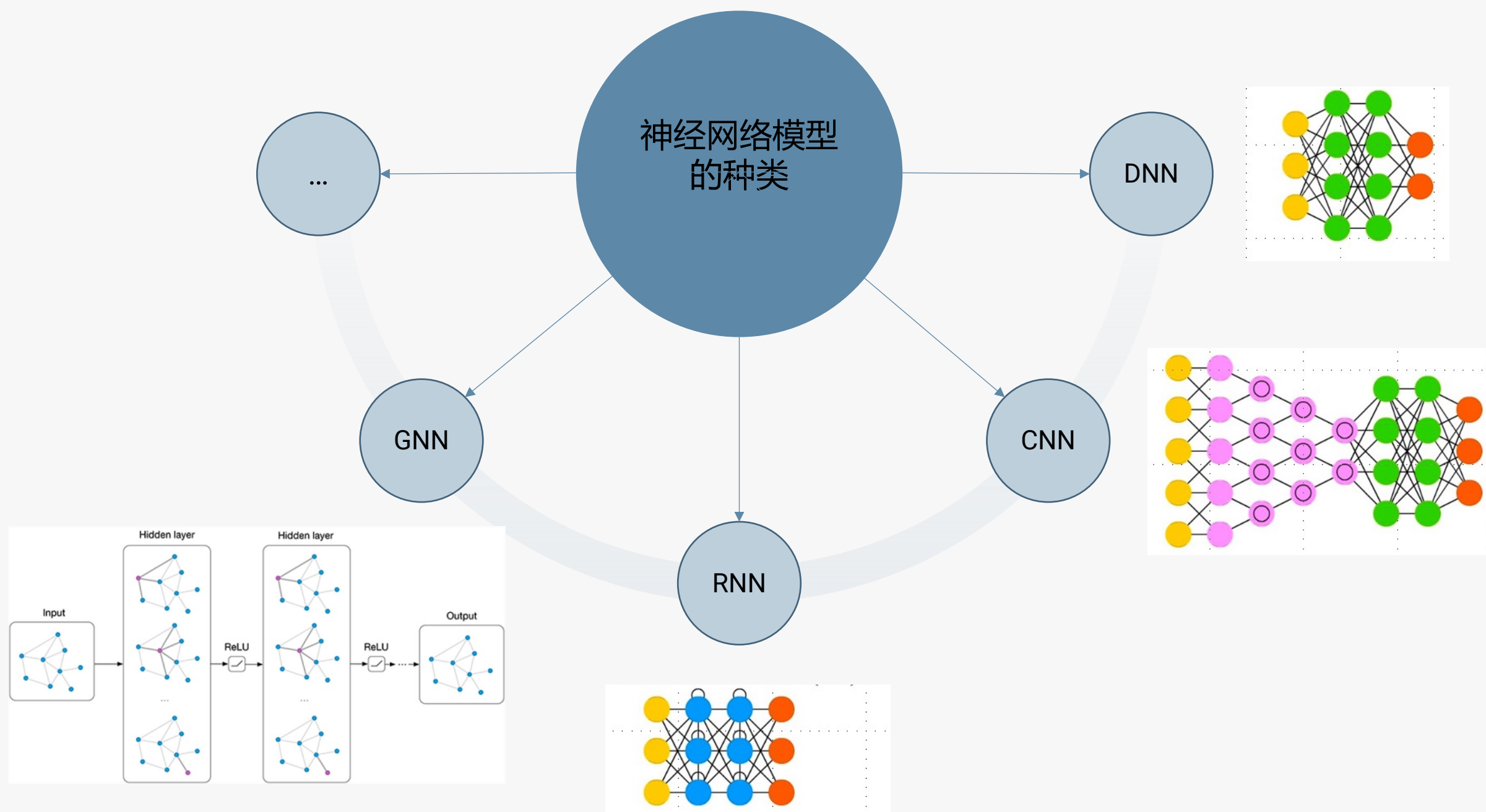


Main Types of Neural Networks



为什么有些模型在有些数据集上表现的好？偏好

我们只找合适的算法



Relational **inductive bias**, deep learning, and graph networks
arXiv:1806.01261 (值得研读的review)

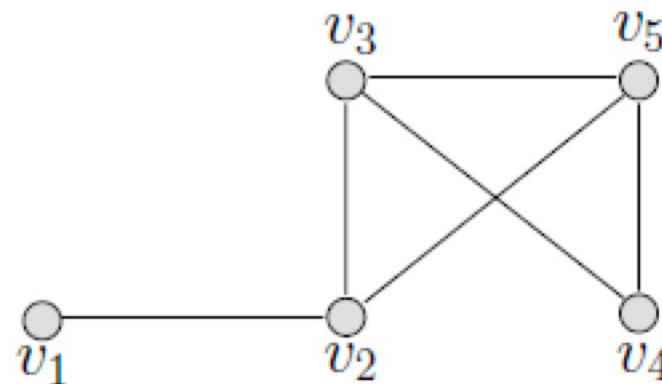
我们只找合适的算法

归纳偏好 Inductive bias

- NFL定理指出学习是不可能的，除非有先验知识。
- 通常情况下我们不知道上帝函数，但猜测它属于一个比较小的假设类别。
- 这种基于先验知识对目标模型的判断就是 inductive bias —— 归纳偏好。归纳偏好所做的事情就是将无限可能的目标函数约束在一个有限的假设类别当中。
- 这样模型学习才能成为可能。如果给出更加宽松的模型假设类别，也即用更弱的 inductive bias，那我们更有可能得到强力模型 —— 更接近目标函数 F ，但是训练变得非常困难，乃至不可能。
- 损失由近似损失和估计损失组成，弱bias虽然减少了近似损失，但会增大估计损失，模型更加难以学习，更容易过拟合。
- 学习者的归纳偏好是一组额外的假设，足以证明其归纳推理是演绎的推论。

图 (Graph)

图 (V, E)



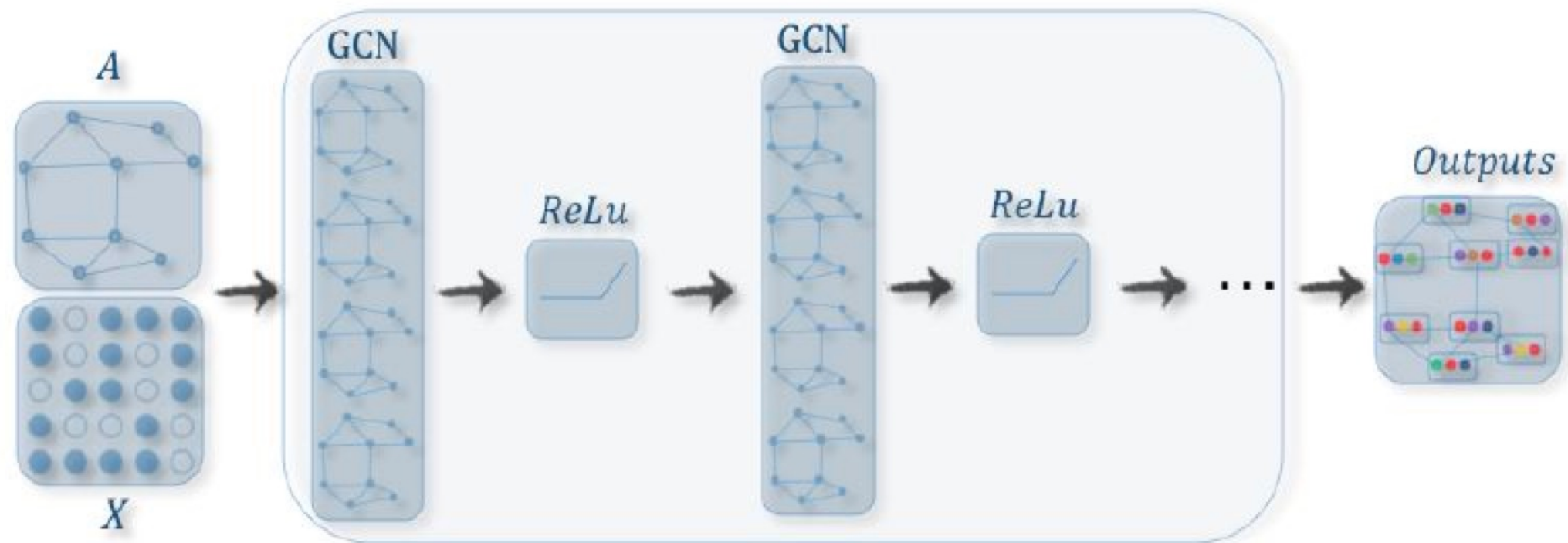
度矩阵和邻接矩阵(degree and adjacent degree)

$$\Delta(\mathcal{G}) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 3 \end{pmatrix} \quad \text{and} \quad A(\mathcal{G}) = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{pmatrix}.$$

<https://blog.csdn.net/luzaijiaoxia0618>

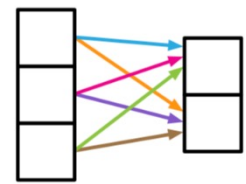
图表示的数据不再局限于欧氏空间

图形神经网络

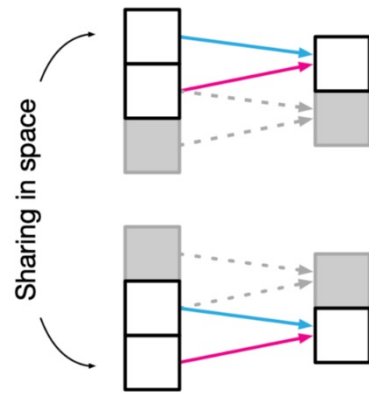


几何深度学习, ...
Geometric Deep
Learning, ...

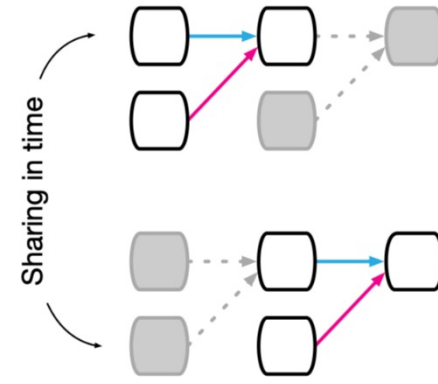
小结：不同算法的归纳偏好比较



(a) Fully connected

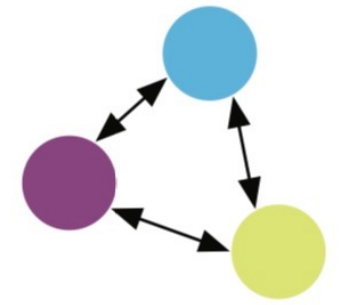


(b) Convolutional



(c) Recurrent

三体系统的质心



类型	输入形式	关系	偏好	变换	评论
FN	单位	All-to-all	弱	---	信息无重用，无孤立
CNN	均匀像素	局域	取决局域性	空间平移	局域特征和平移不变性
RNN	均匀时间步长	序列	取决于序列性	时间平移	信息重用+时间平移不变
GNN	节点	边	变化大	点、边的交换	顺序无关，IB来自某种东西的“无”（absence）

准备：环境+数据

两种环境

- Root& TMVA
 - ✓ 安装了Root即可（版本越新越好）
- Python
 - ✓ 选择框架：Keras(tensorflow)、PyTorch, ... 用

miniconda节约时间

三步执行例子

readme.txt

安装 miniconda

选择安装路径为\${INSTALL}
01.batch.sh

create 独立环境

02.batch.sh

```
export PATH=$PATH:${INSTALL}/bin/  
source ${INSTALL}/etc/profile.d/conda.sh  
conda info -e  
#conda remove energyflow  
conda create -n energyflow python=3.9  
conda activate energyflow  
pip install -r requirement.txt
```

运行程序

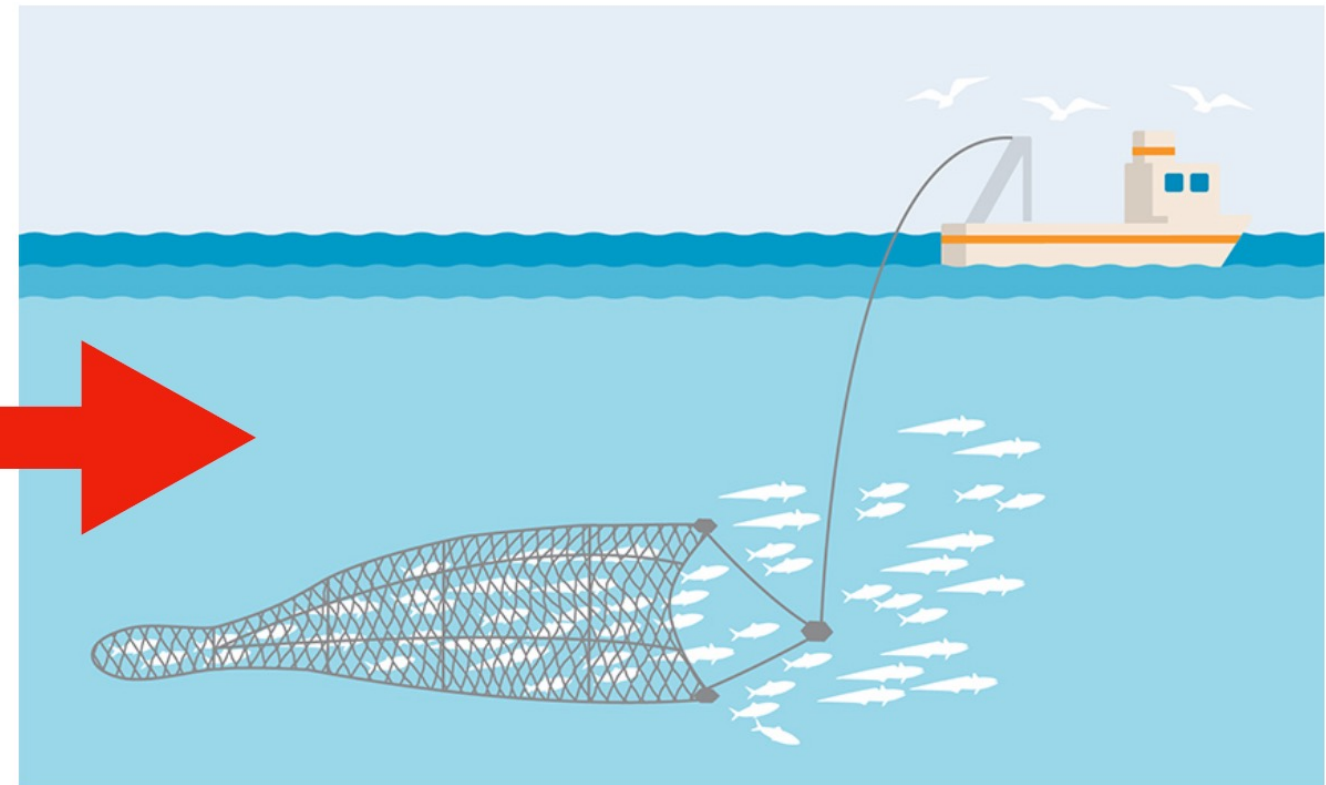
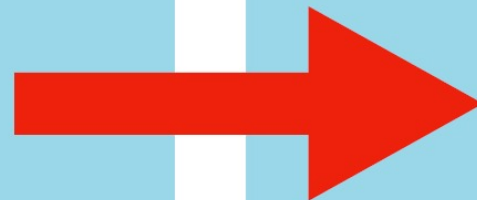
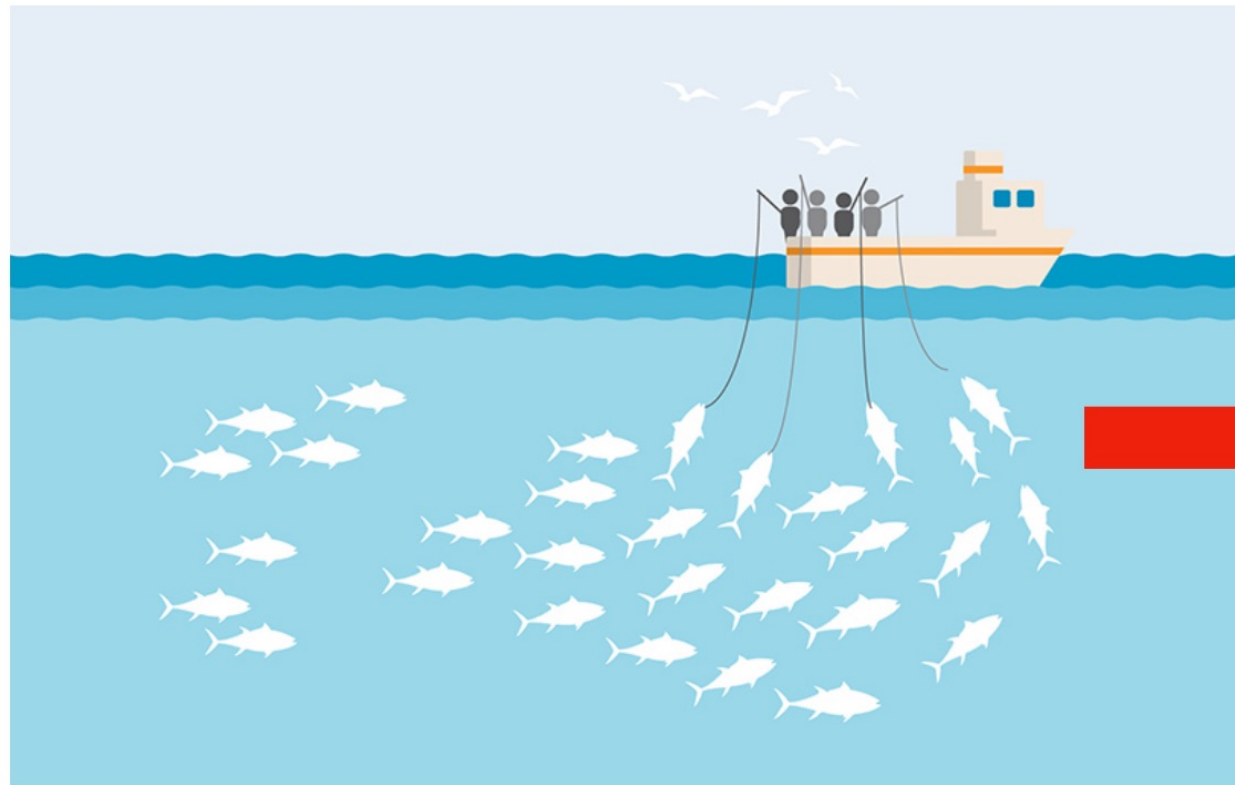
source 03.env.sh

cd example

conda activate energyflow

python pfn.py

实例



传统的高能数据分析

未来的高能数据分析

数据分析第一步：事例选择

- 假设对撞机上产生了n（9）类事例（ $N_1, N_2, N_3, \dots$ ）
- 我们需要选择其中一种作为信号来研究，比如测量 N_1 的分支比 B_1 ，其它过程都是本底
- 分类问题
- 此时只要分为信号和本底即可

$$B = \frac{N_1}{N_1 + N_2 + N_3 + \dots}$$

$$\begin{pmatrix} n_s \\ n_b \end{pmatrix} = [\epsilon_{11} \quad \underbrace{(\epsilon_{12} \quad \epsilon_{13} \quad \epsilon_{14} \quad \epsilon_{15} \quad \epsilon_{16} \quad \epsilon_{17} \quad \epsilon_{18} \quad \epsilon_{19})}_{\underbrace{\begin{pmatrix} N_1 \\ N_2 \\ N_3 \\ N_4 \\ N_5 \\ N_6 \\ N_7 \\ N_8 \\ N_9 \end{pmatrix}}_{\text{Signal}}}]$$

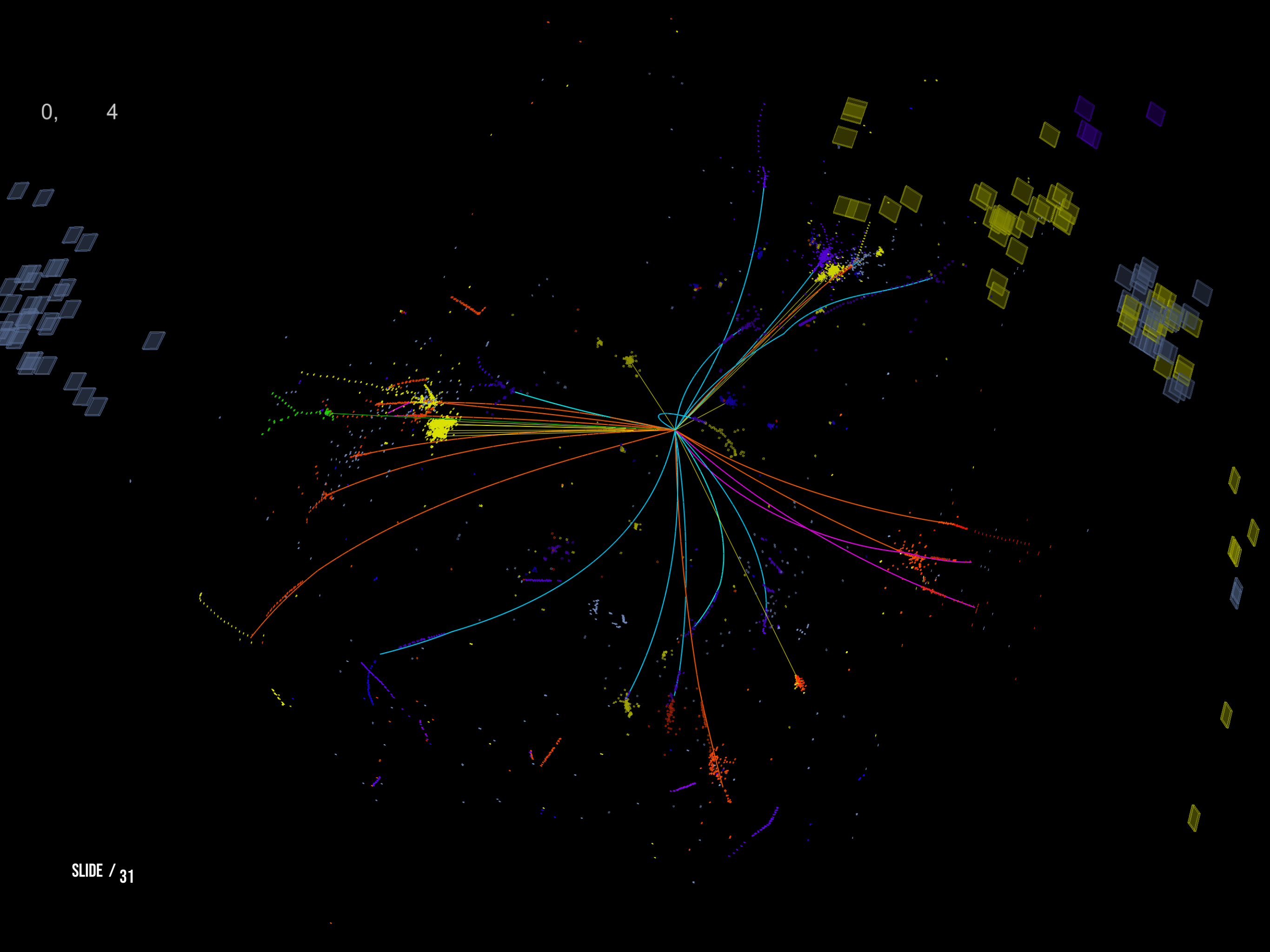
问题：如果分为9类会有什么效果？

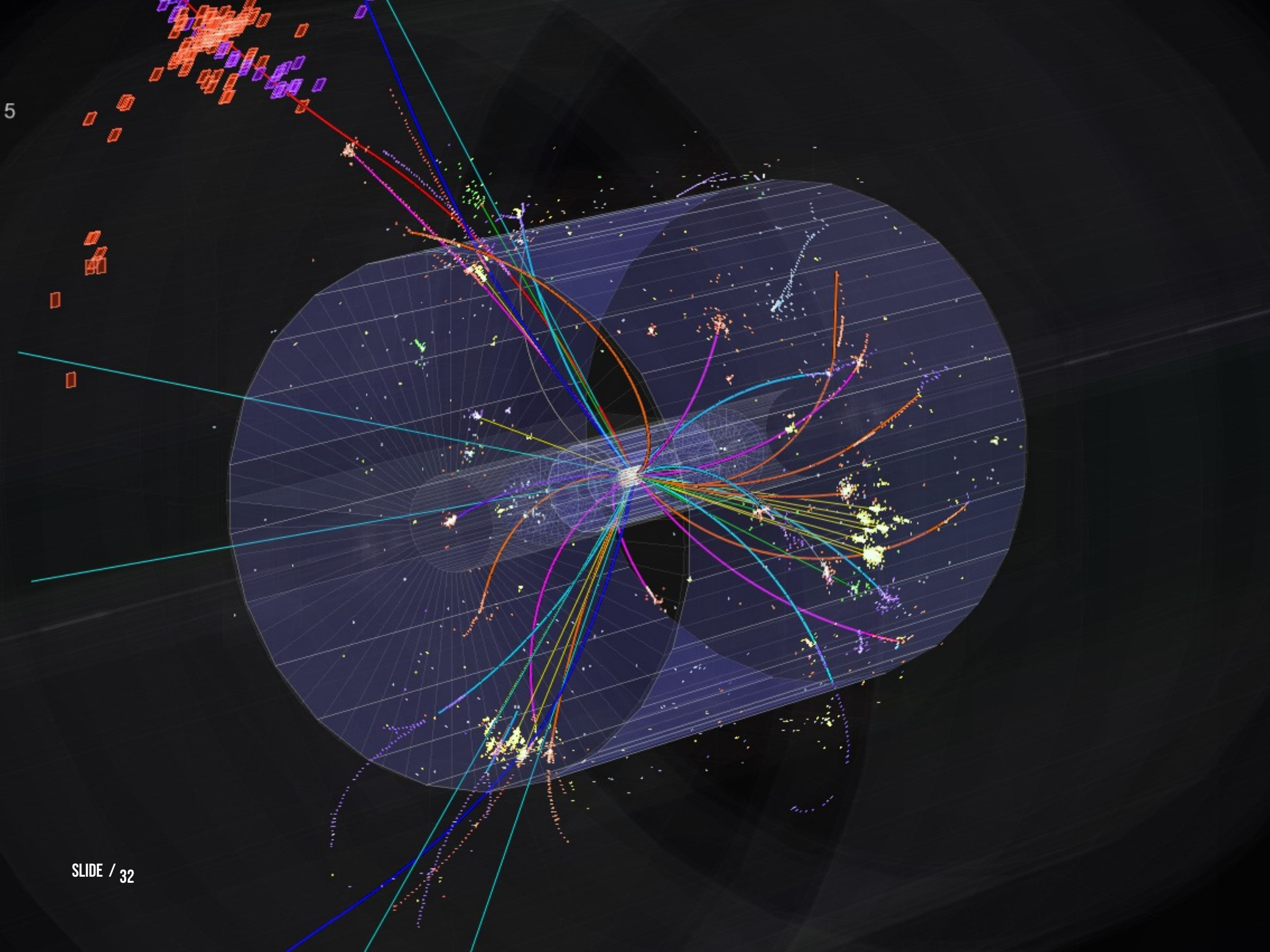
Efficiency matrix determined by DL multi-classification

$$\begin{pmatrix} n_1 \\ n_2 \\ \vdots \\ n_9 \end{pmatrix} = \begin{pmatrix} \epsilon_{11} & \epsilon_{12} & \cdots & \epsilon_{19} \\ \epsilon_{21} & \epsilon_{22} & \cdots & \epsilon_{29} \\ \vdots & \vdots & \ddots & \vdots \\ \epsilon_{91} & \epsilon_{92} & \cdots & \epsilon_{99} \end{pmatrix} \begin{pmatrix} N_1 \\ N_2 \\ \vdots \\ N_9 \end{pmatrix}$$

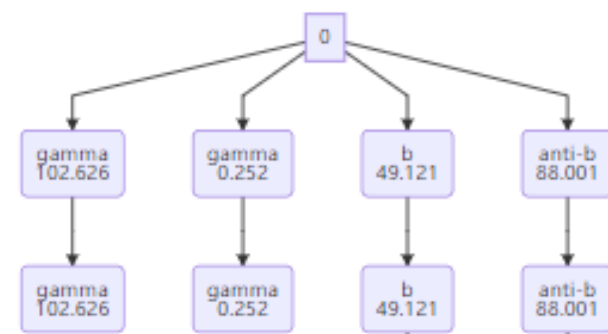
"on-stop" measurement of N quantities
more efficient

0, 4

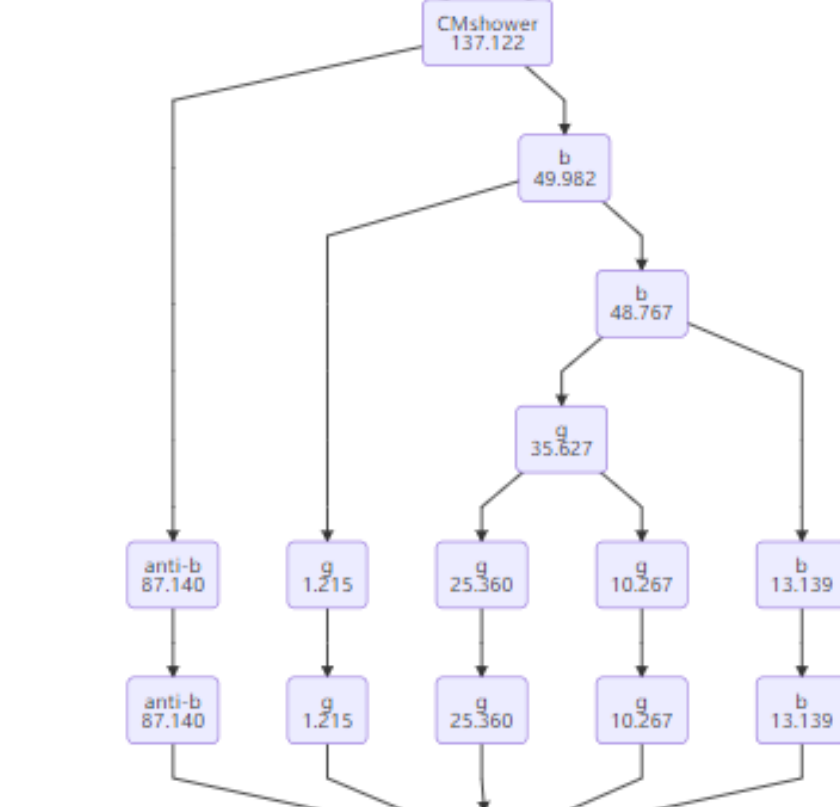
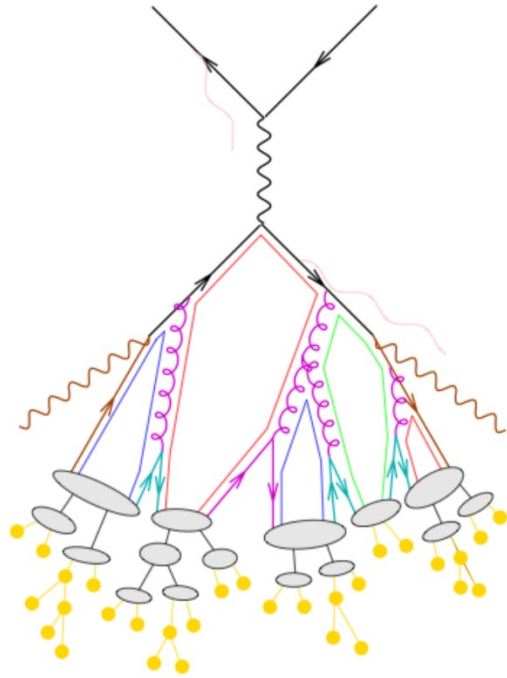




$$e^+e^- \rightarrow b\bar{b}$$

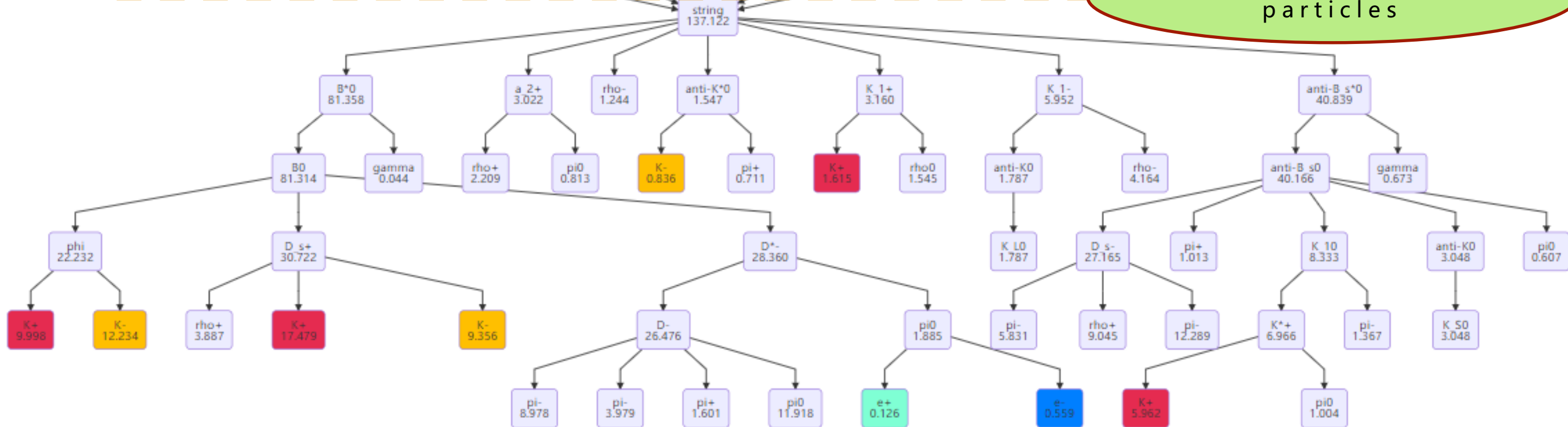


Hard process



Fragmentation

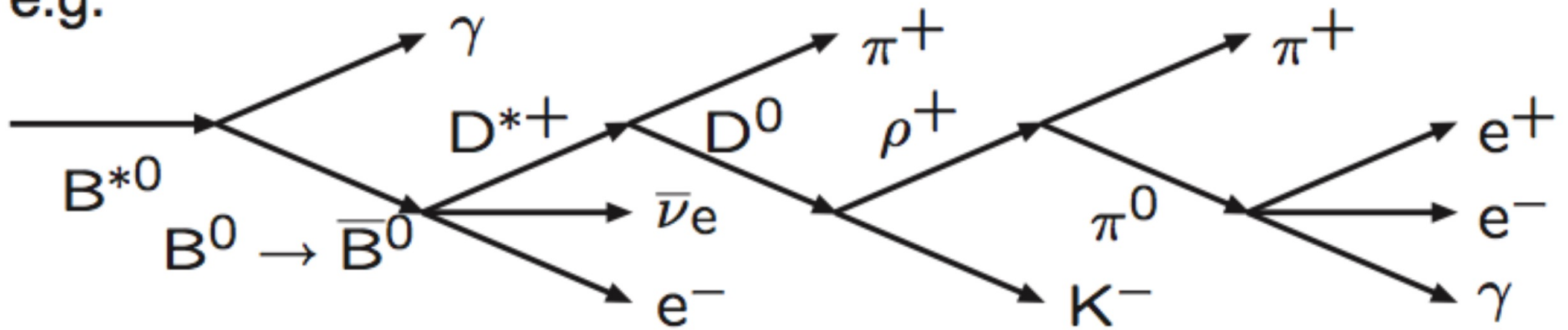
(Final state) particles



事例中间过程的一个例子

级联衰变直到稳定粒子

e.g.

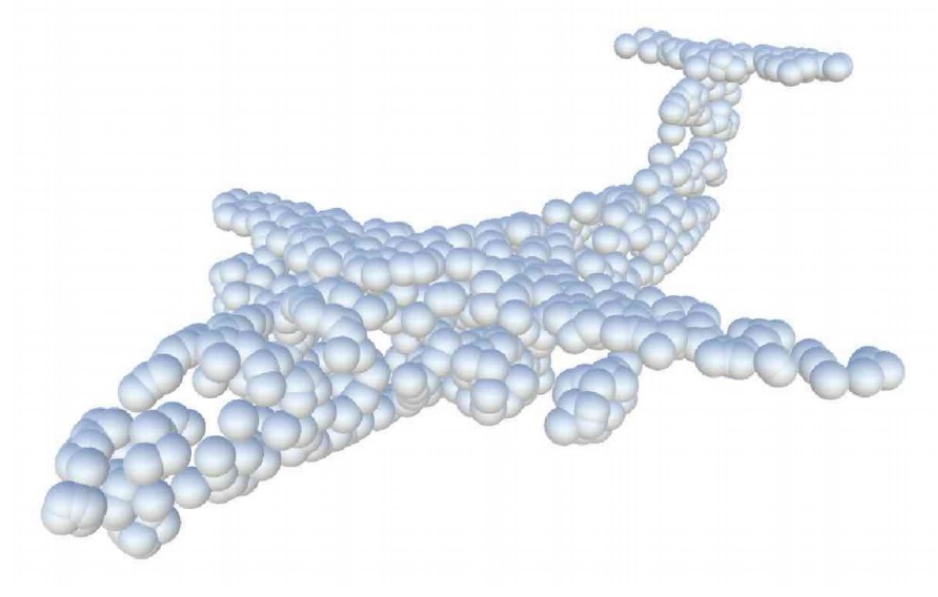


探测器最多只能知道末态粒子的

- 四动量（有事只有三动量）
- 粒子的起点（ r 、 z ）
- 粒子的终点（大小适中的话）

作为点云的 Jet ...

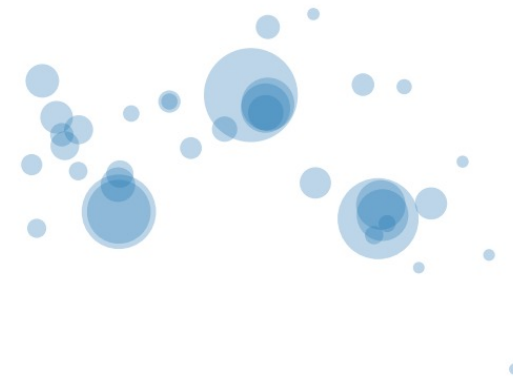
Point cloud $\rightarrow$ particle cloud (jet)



■ Point cloud

- points are intrinsically unordered
- primary information:
 - 3D coordinates in the xyz space

点和点之间的关系可以选择某种数学方式来表示，且不依赖点的顺序和总数；合适的、有物理背景的选择可以有比较强的表达能力，信息无损。



■ Particle cloud

- particles are intrinsically unordered
- primary information:
 - 2D coordinates in the η - φ space
- but also additional “features”:
 - energy/momenta
 - charge/particle type
 - track quality/impact parameters/etc.

比普通点云更加丰富的内容

一个事例可以表示为
无序、数目可变的粒子集合

交换 对称性

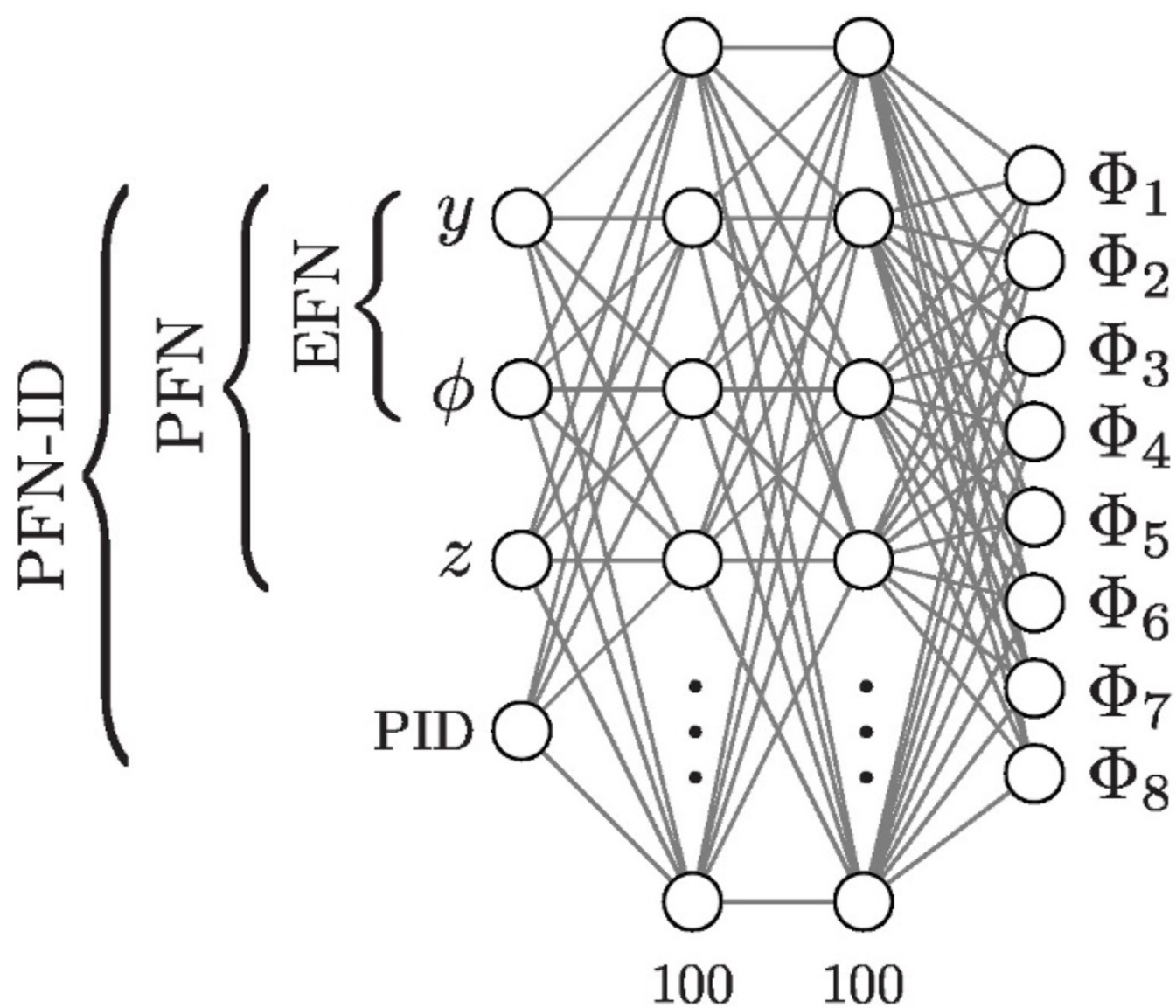
$$J(\{p_1^\mu, \dots, p_M^\mu\}) = J(\{p_{\pi(1)}^\mu, \dots, p_{\pi(M)}^\mu\})$$

p_i^μ represents *all* the particle properties:

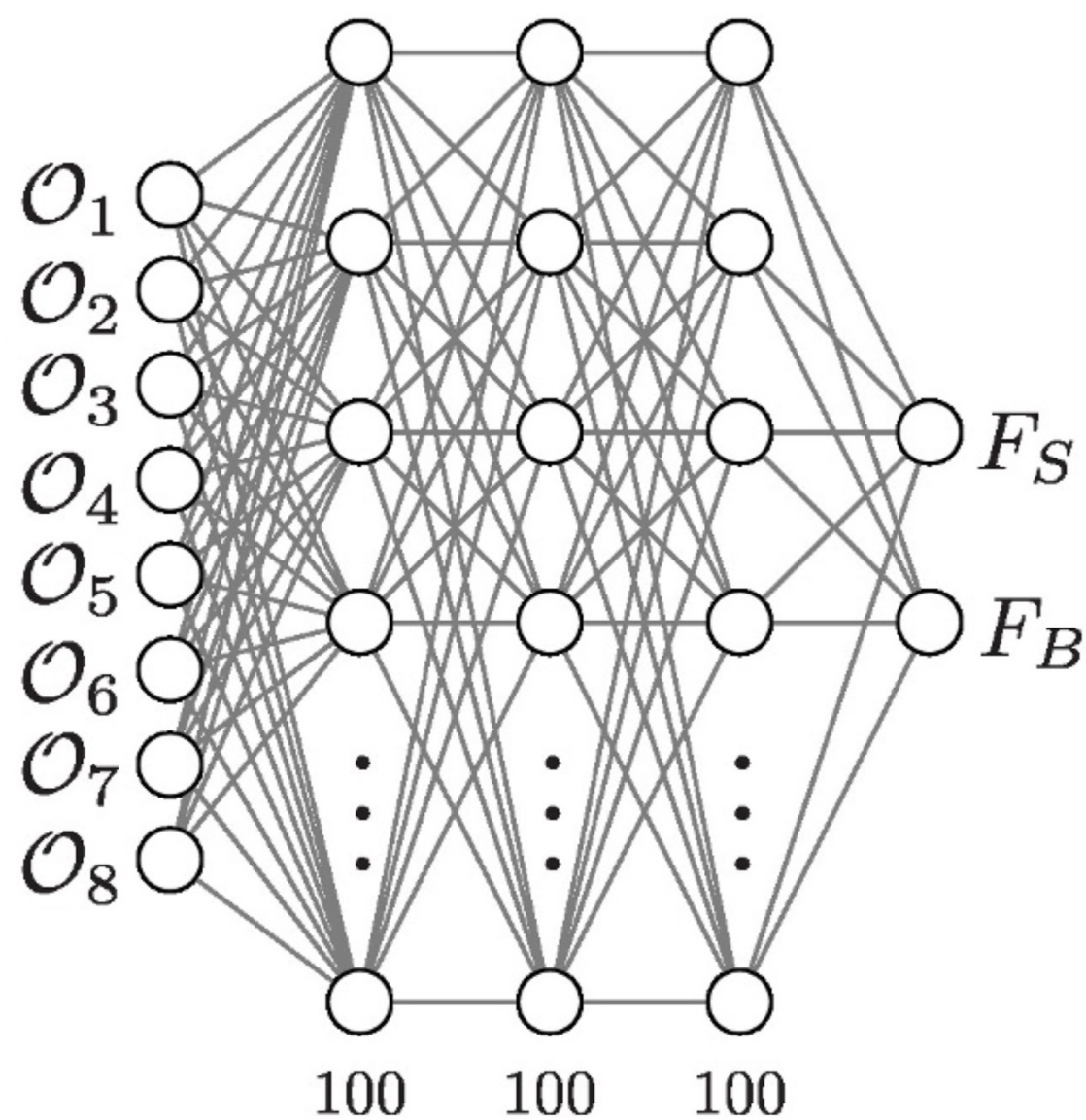
四动量，粒子质量，电荷，顶点信息 (~10 个 float)

Mapping

DNN

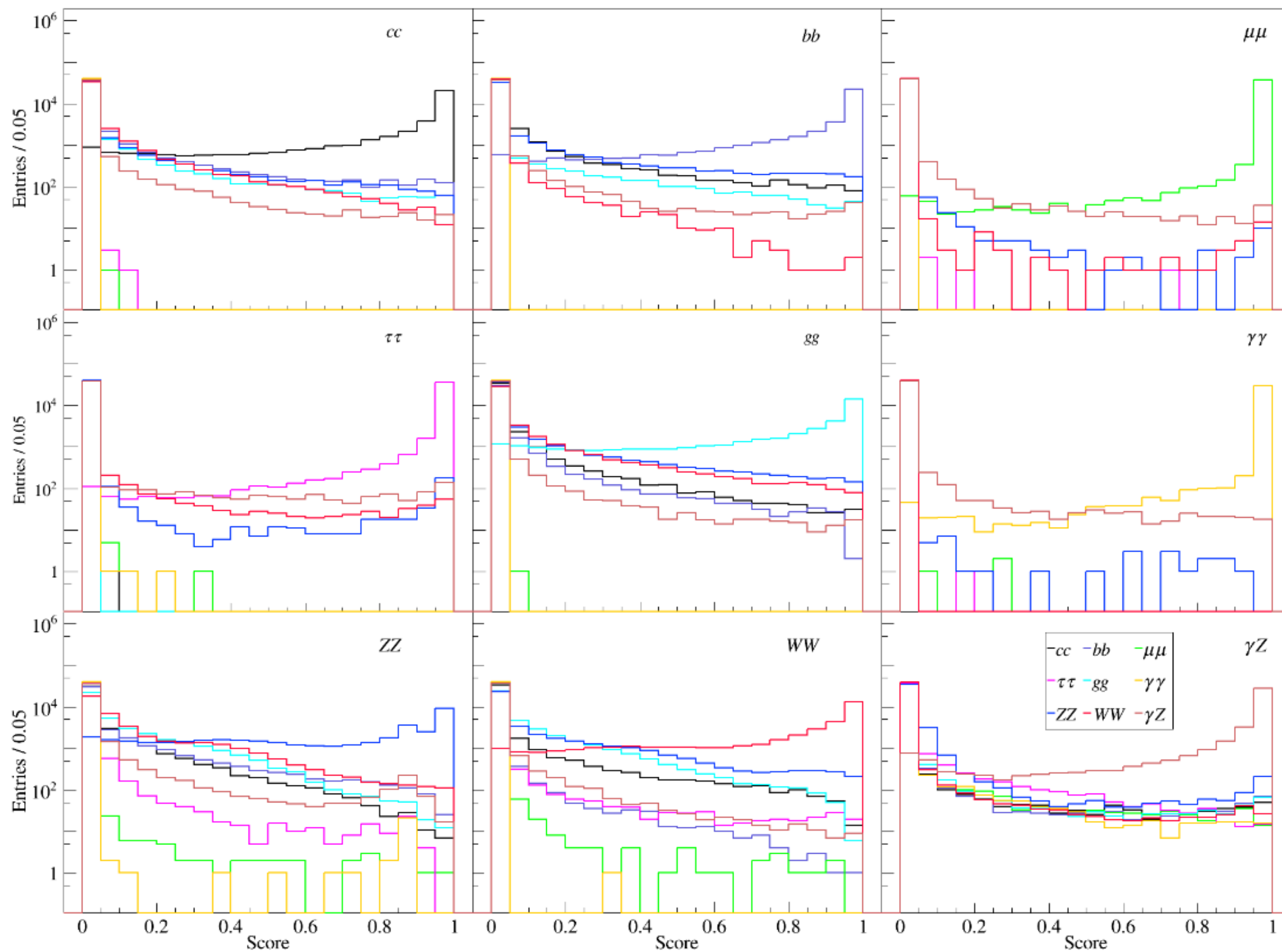


表示学习

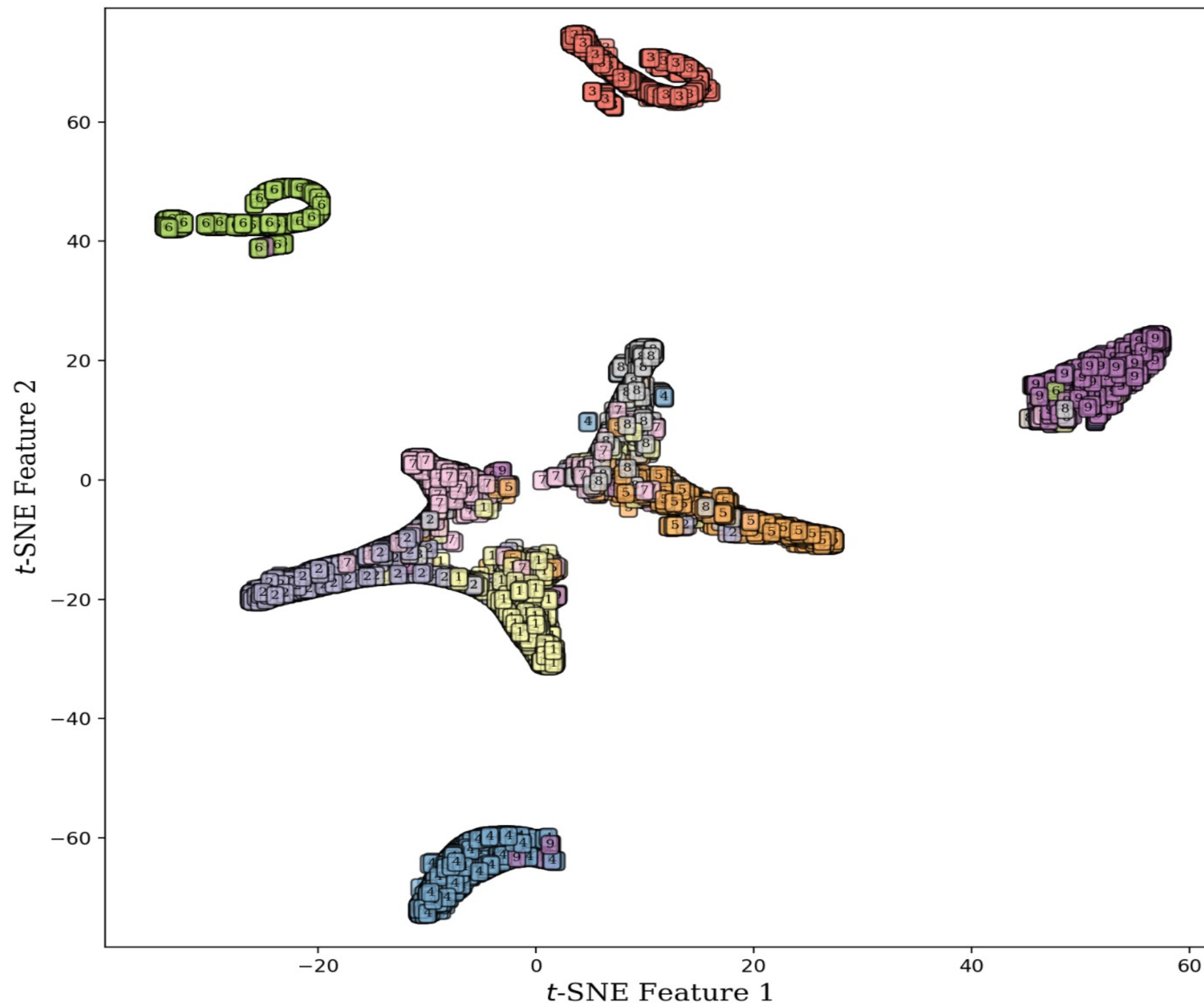


分类

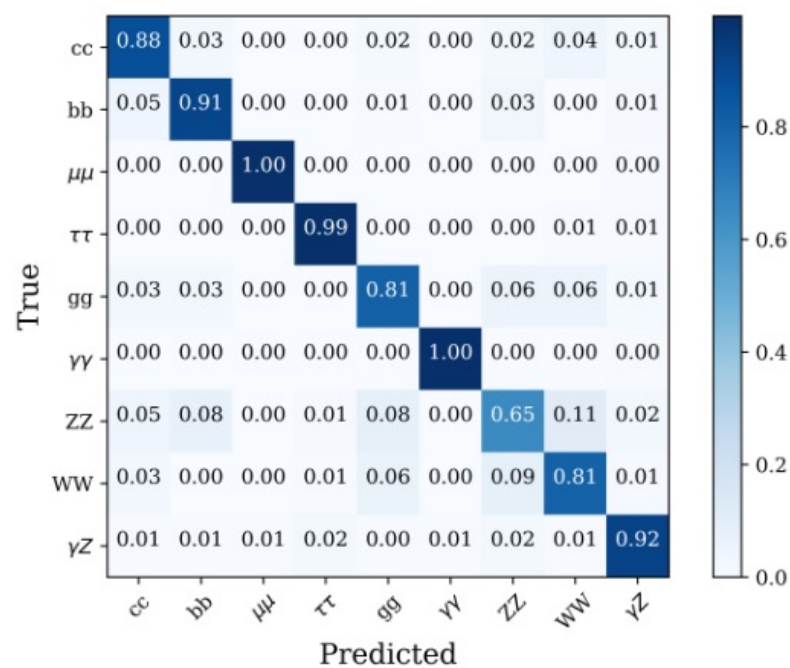
机器学习可以做到吗？答案是肯定的



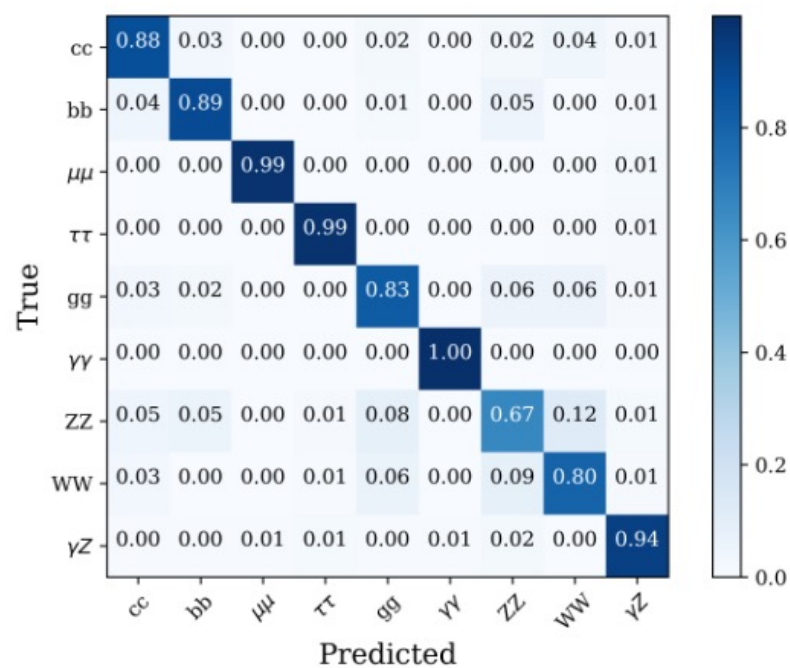
降维后的结果 (t-SNE作者是Hinton)



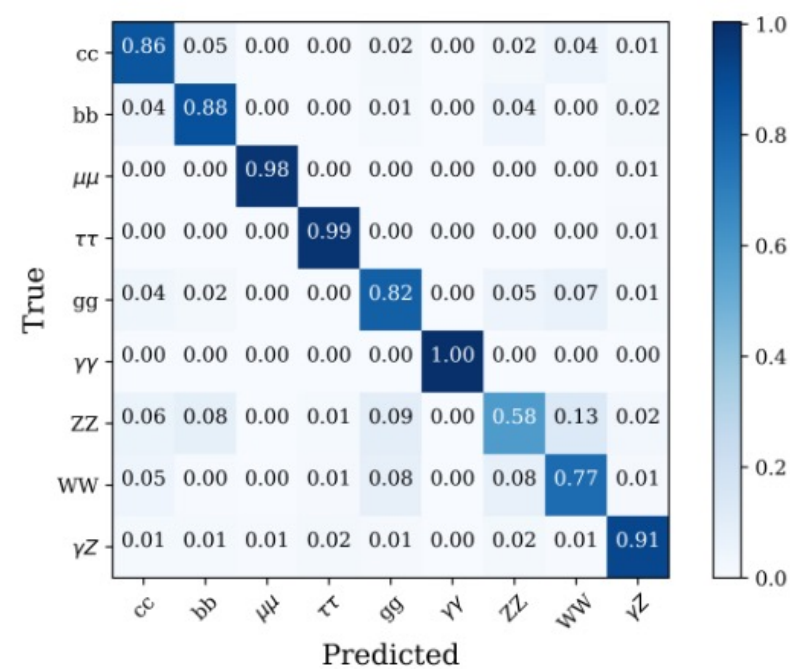
机器学习可以做到吗？答案是肯定的



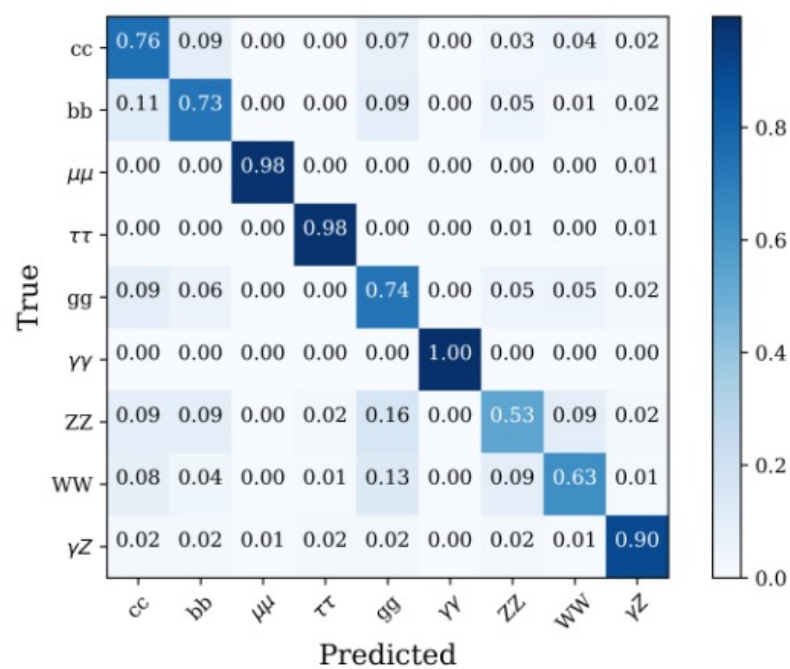
(a)



(b)



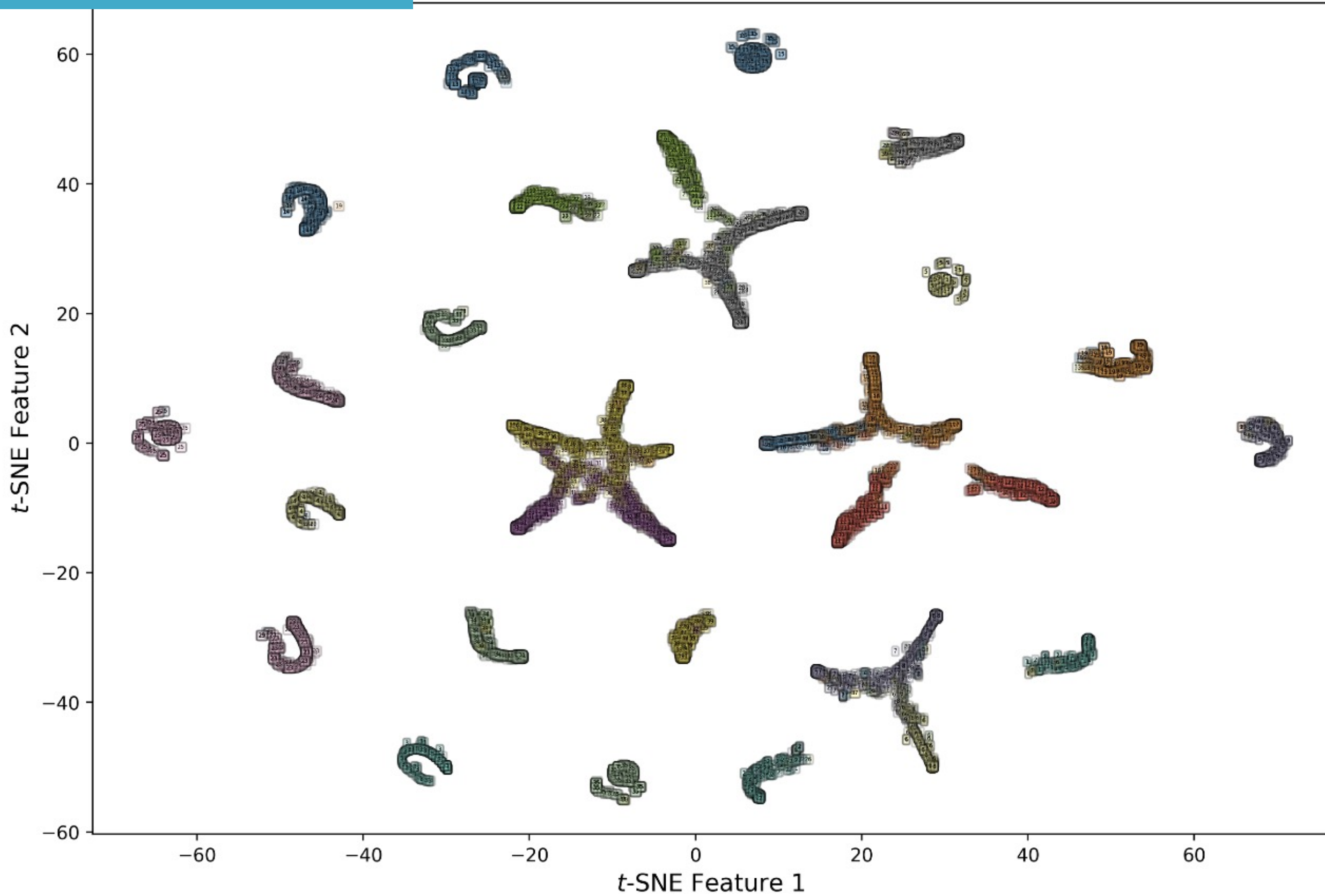
(c)



(d)

36 classes

ParticleNet features: t -SNE



小结

- 深度学习非常有用
- 深度学习不是黑盒子、也不是全能魔法
- 深度学习的可解释性正在发展中——突破的前夜 ...
- 多关注传统机器学习方法，对理解深度学习非常有帮助
- 与数学、物理的结合非常紧密：多关注问题的物理和统计本质
- 高能物理实验是非常好的应用场景
- 最终不再搬砖
- 学好线代、统计和物理

推荐阅读的综述文献

- A high-bias, low-variance introduction to Machine Learning for physicists , Physics Reports 810 (2019) 1–124
- Relational inductive bias, deep learning, and graph networks, arXiv:1806.01261
- Geometric Deep Learning, 5Gs, arXiv:2104.13478

谢谢大家

Extras

关于确定原理的知识很容易用来补偿我们对确定事实的无知
—— 爱尔维修, 1759

La connaissance de certains principes supplée facilement à
la connoissance de certains faits.
(Helvétius, 1759)

机器学习用于Jet 研究

Jet Representations $\longleftrightarrow$ Analysis Tools

Two key choices when analyzing jets

How to represent the jet

- Single expert observable
- A few expert observables
- Many expert observables
- Jet images
- List of particles
- Clustering tree
- N -subjettiness basis
- Energy flow polynomials
- Set of particles

PTK ML4Jets 2017

PTK ML4Jets 2018

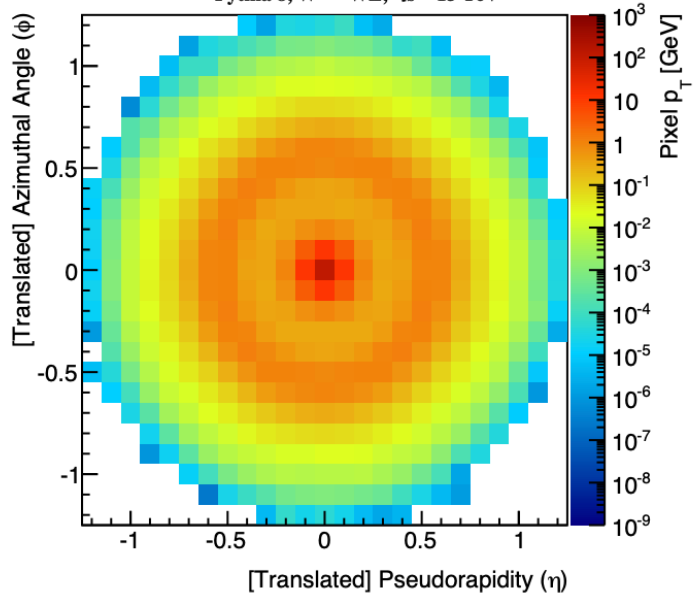
How to analyze that representation

- Threshold cut
- Multidimensional likelihood
- Boosted decision tree (BDT), shallow neural network (NN)
- Convolutional NN (CNN)
- Recurrent/Recursive NN (RNN)
- Fancy RNN
- Dense neural network (DNN)
- Linear classification
- Energy flow network

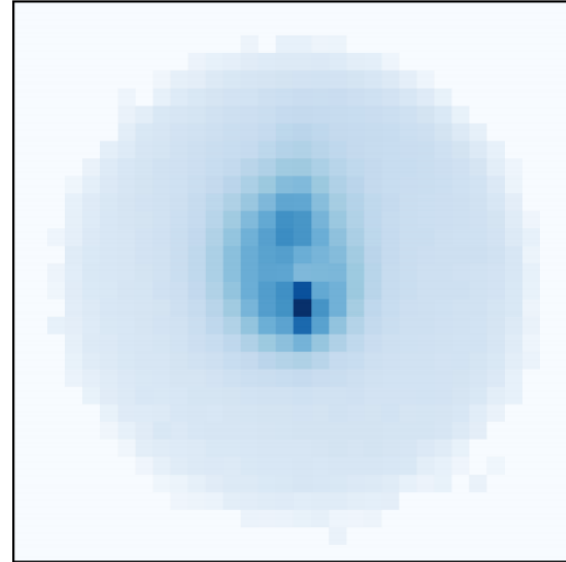
作为图片的 Jet ...

$250 < p_T/\text{GeV} < 260 \text{ GeV}$, $65 < \text{mass}/\text{GeV} < 95$

Pythia 8, $W' \rightarrow WZ$, $\sqrt{s} = 13 \text{ TeV}$

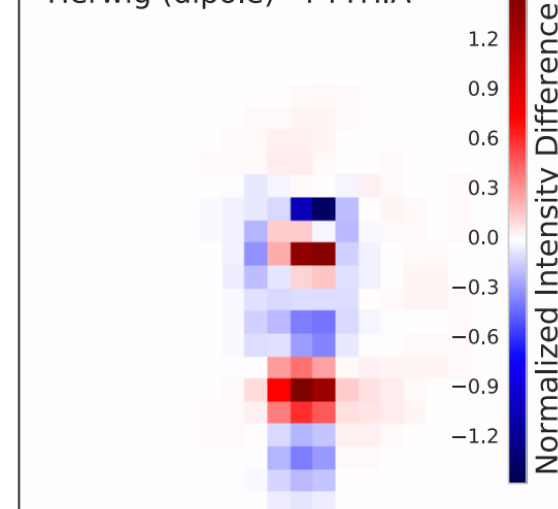


1511.05190



1603.09349

Herwig (dipole) - PYTHIA

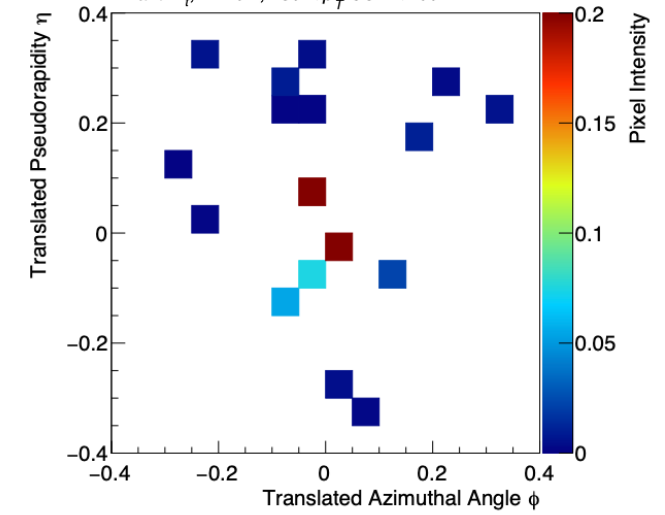


1609.00607

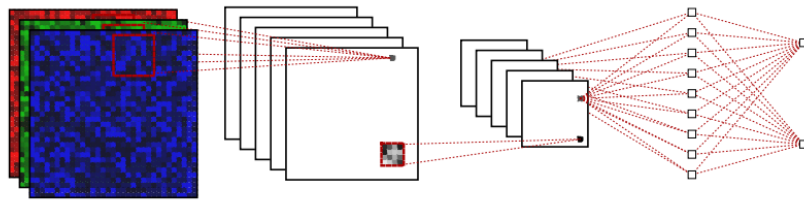
ATLAS Simulation Preliminary

Gluon Jets, Topocluster Constituents

anti- k_T , $R = 0.4$, $150 < p_T/\text{GeV} < 200$



ATL-PHYS-PUB-2017-017

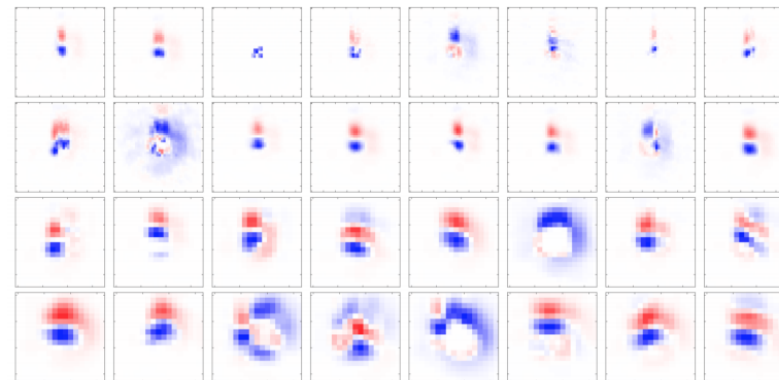


red = transverse momenta of charged particles

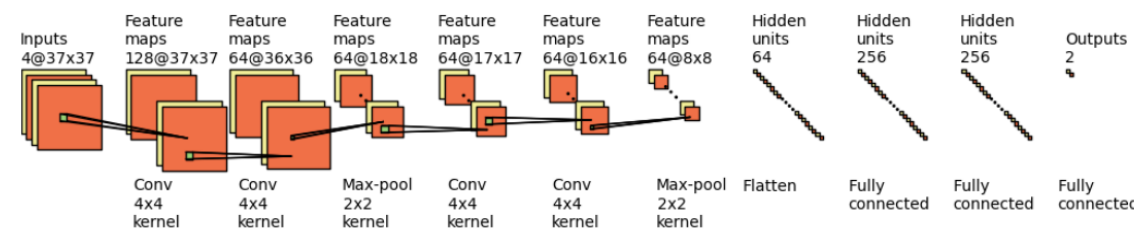
green = the transverse momenta of neutral particles

blue = charged particle multiplicity

1701.08784



1803.00107

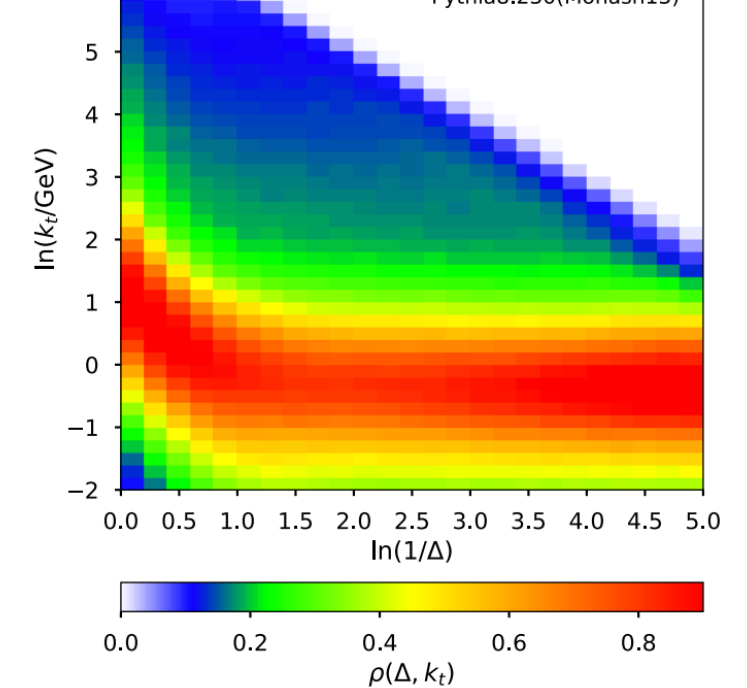


1612.01551

QCD jets, averaged primary Lund plane

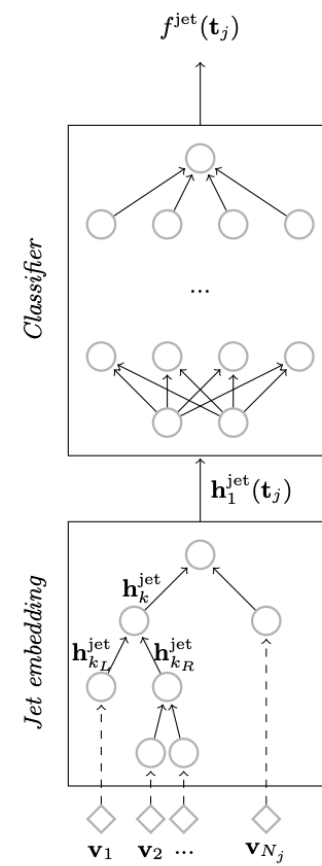
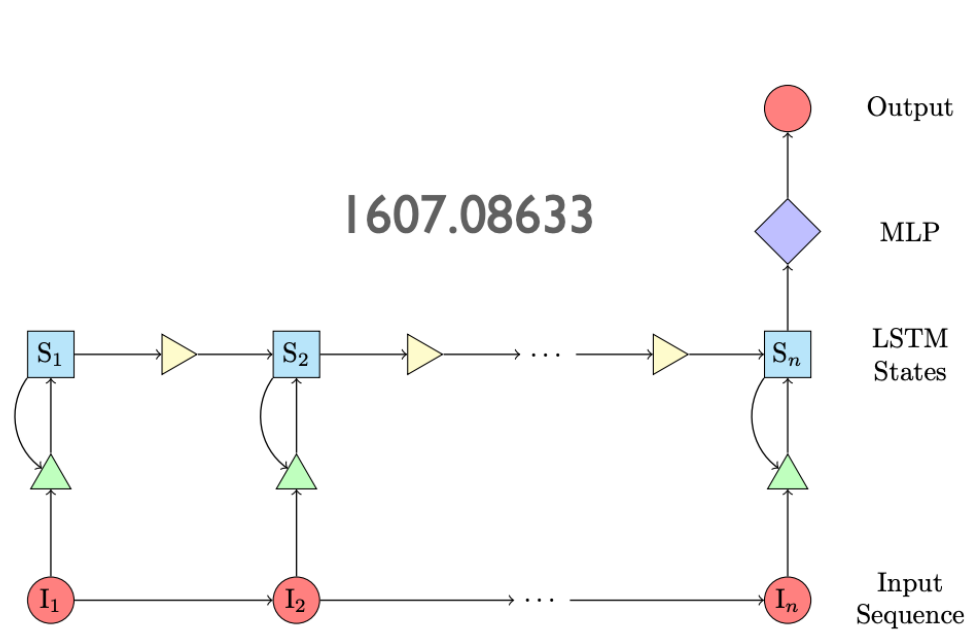
$\sqrt{s} = 14 \text{ TeV}$, $p_T > 2 \text{ TeV}$

Pythia8.230(Monash13)



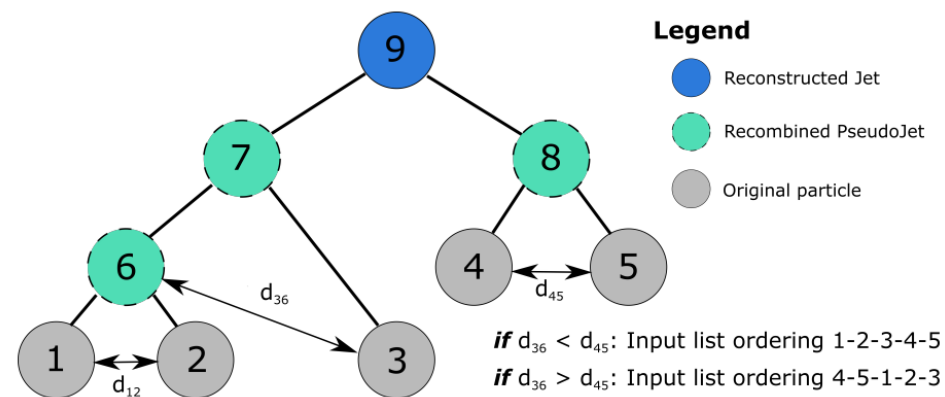
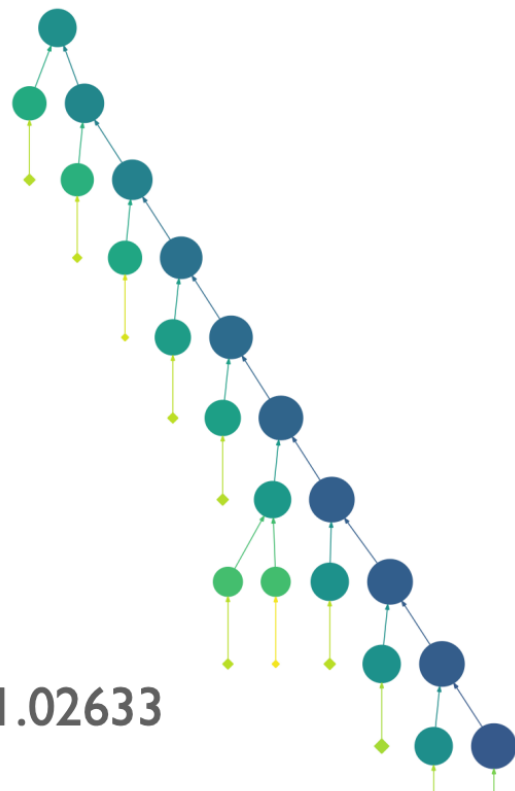
1807.04758

作为序列的 Jet ...



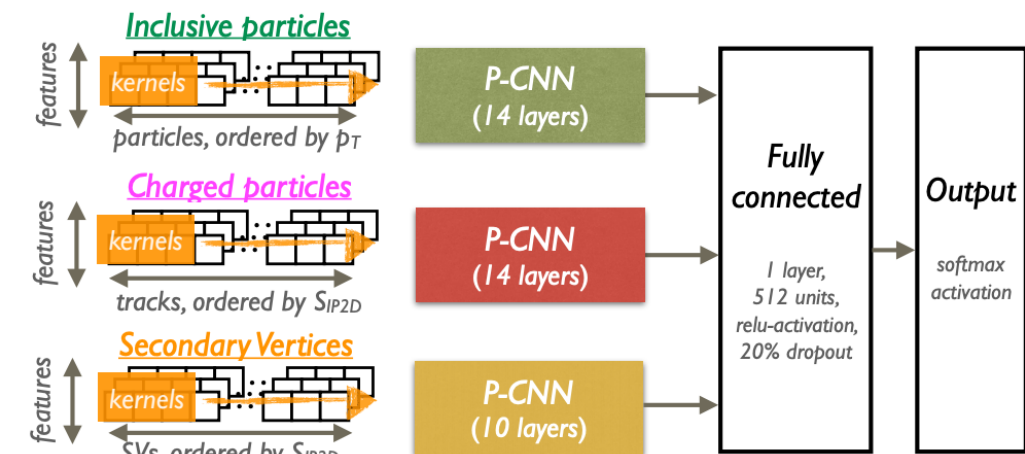
1702.00748

1711.02633



1711.09059

“DeepAK8”
CMS-DP-2017-049



P-CNN:

