



# 机器学习 (第二部分)：神经网络及其应用

第一届 CMS 中国冬令营

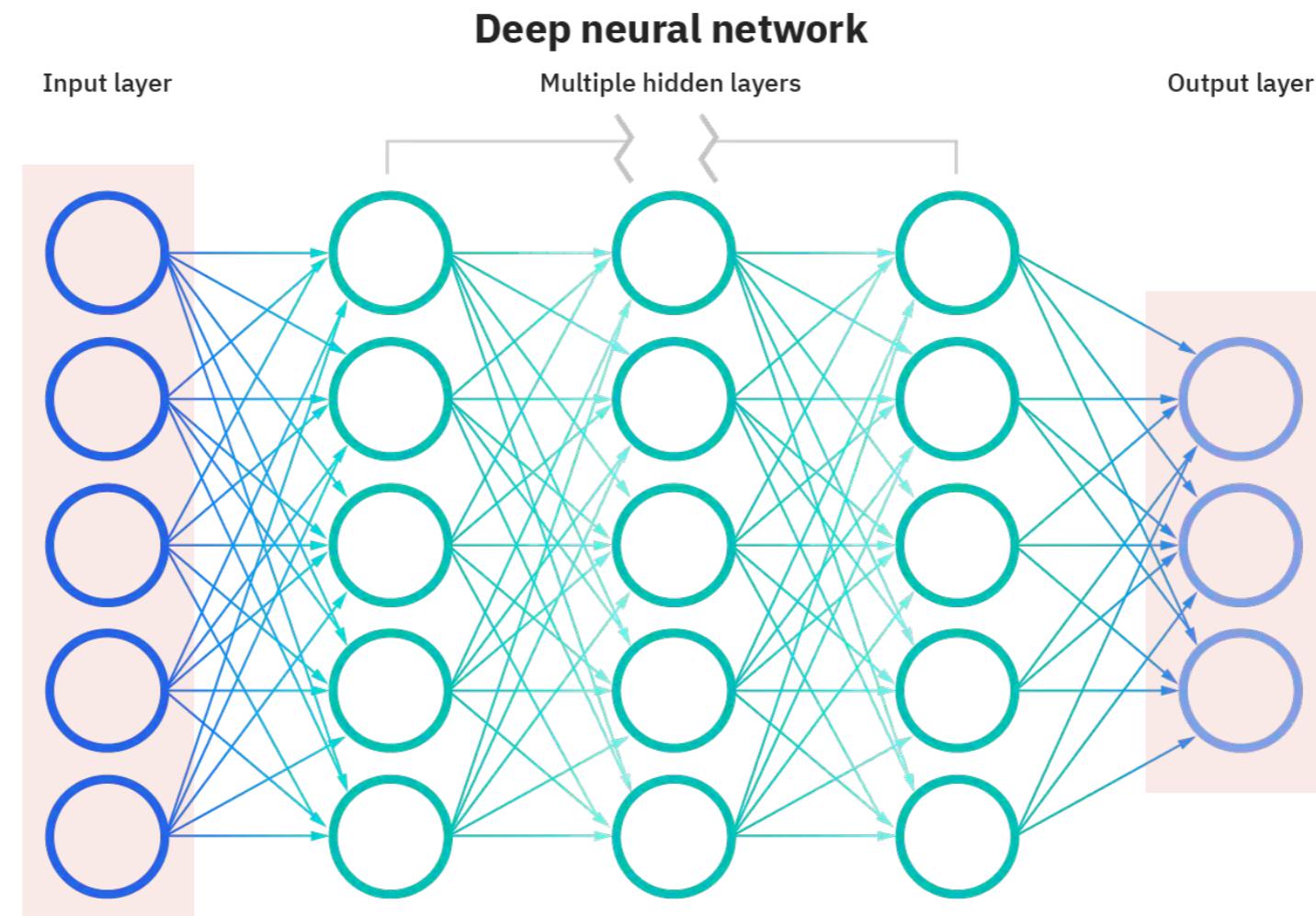
主讲人：李聪乔 钱思天

2021/12/20

# 内容简介

- 本 hands-on 环节将向同学们介绍基础的神经网络知识及它在 CMS 实验中的应用
  - ❖ 什么是神经网络？它的基本原理是什么？
  - ❖ 它在 CMS 分析中的典型应用是什么 (hands-on)
  - ❖ 学习一些实战技术、调参细节
  - ❖ 在 CMS 实验中的一些前沿应用 (hands-on)
- 本 hands-on 环节的全部代码请见：
  - ❖ <https://gitee.com/colizz/cmschina-dnn-tutorial>

# 神经网络介绍

 $X$ 

输入层  
这里为  $1 \times 5$  向量

$$Y = f(X)$$

神经网络本体  
包含许多隐藏层  
有许多自由参数

 $Y$ 

输出层  
这里为  $1 \times 3$  向量

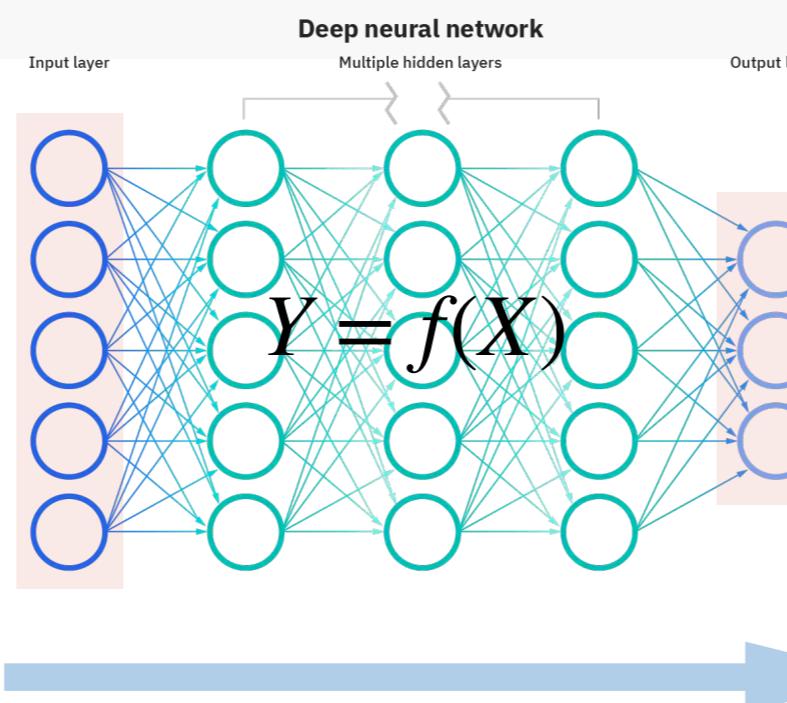
# 神经网络介绍

 $X$ 

输入手写  
数字图像

0	0	0	0	0	0
1	1	1	1	1	1
2	2	2	2	2	2
3	3	3	3	3	3

输入  
图像

 $Y$ 

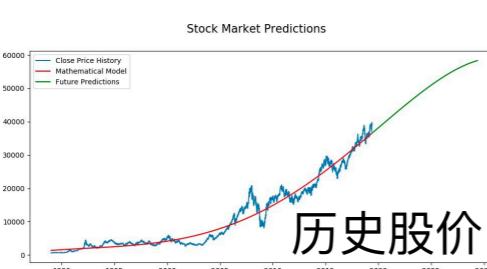
手写数字分类  
10 种类别

 $0 \ 1 \dots 0 \ 0$ 

ImageNet 分类：  
1000 种类别

 $0 \ 1 \dots 0 \ 0$ 

分类



嵌入到输入  
矢量

 $?$ 

下一时刻股价预测

中文文本

吃了吗您？

矢量嵌入

 $?$ 

How are you doing?

英文文本

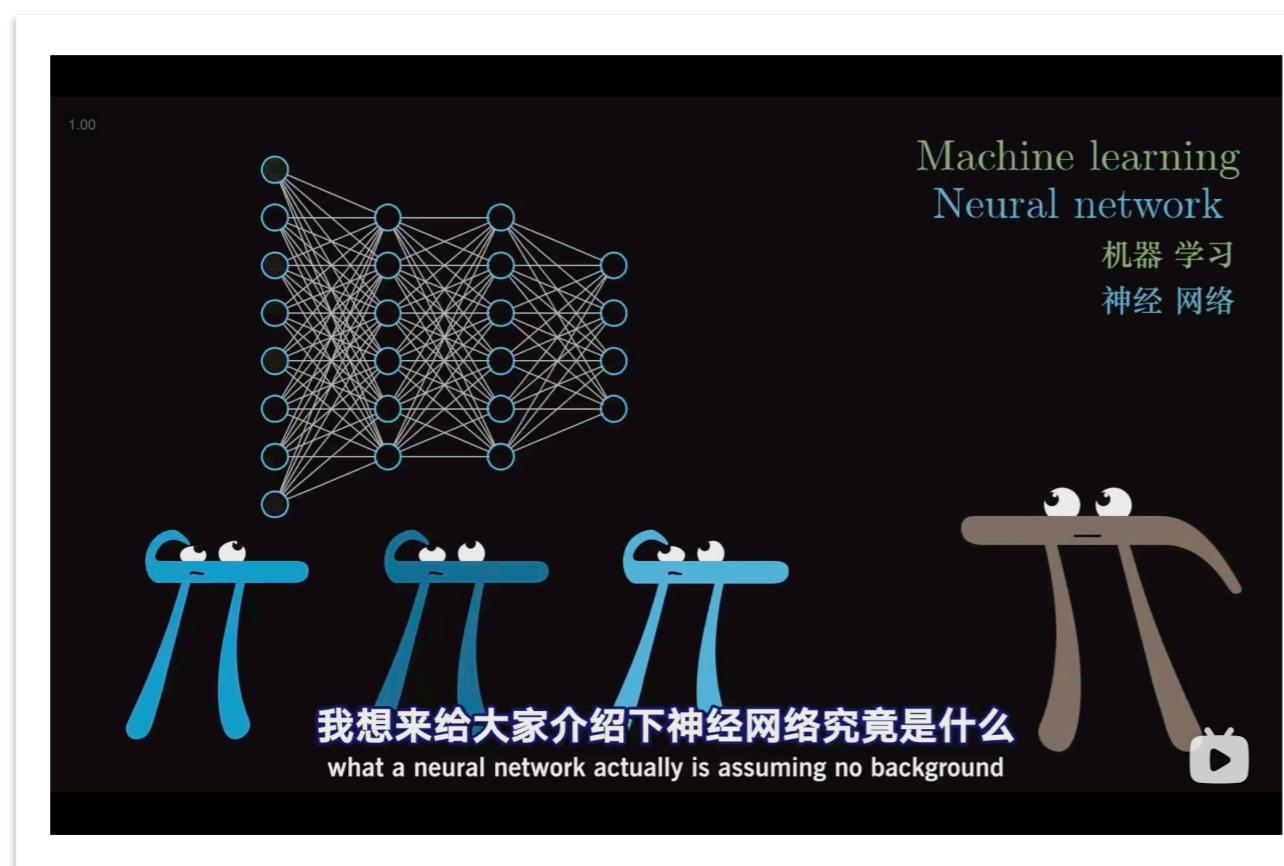
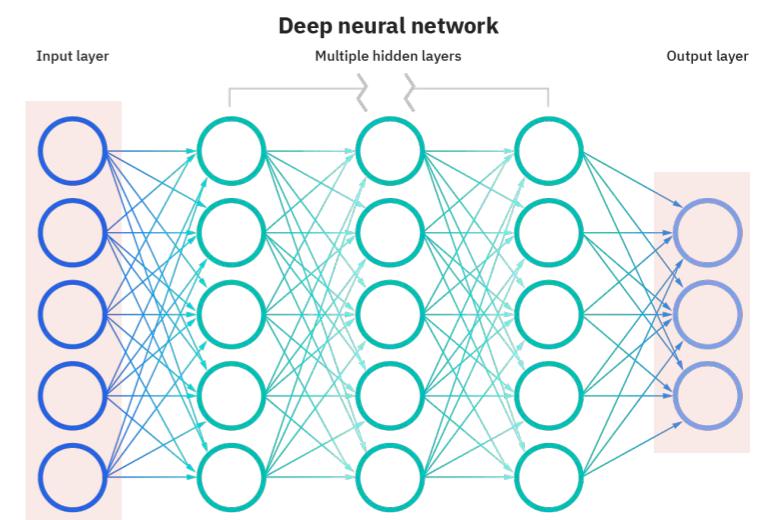
回归

# 神经网络工作原理

→ 观看视频短篇，了解神经网络的工作原理

→ 带着如下问题观看：

- ❖ 神经网络为什么要很多隐藏层？
- ❖ 神经网络为什么设置“激活”层？
- ❖ 神经网络函数的参数以什么为目标进行更新？



视频链接 [\[Part1\]](#) [\[Part2\]](#)

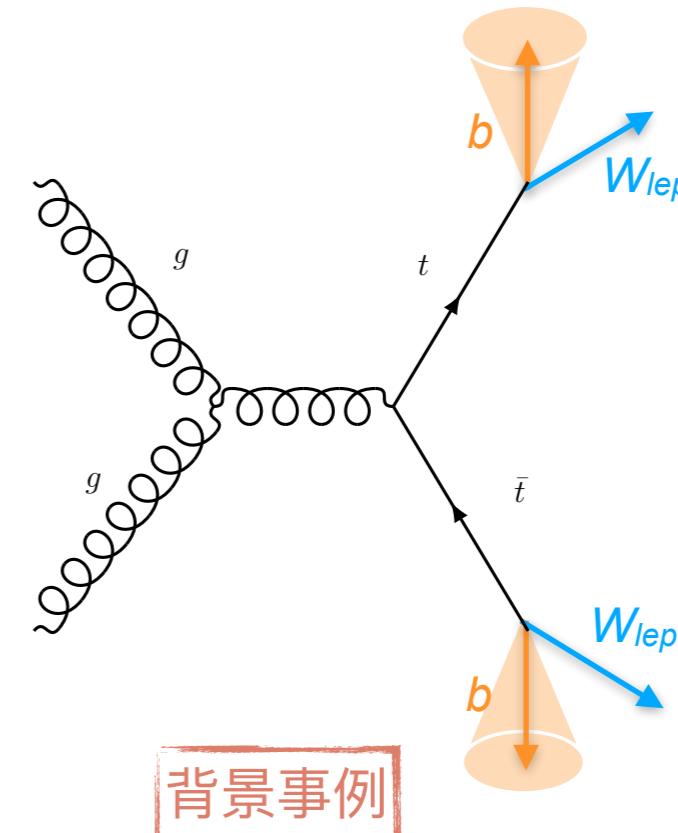
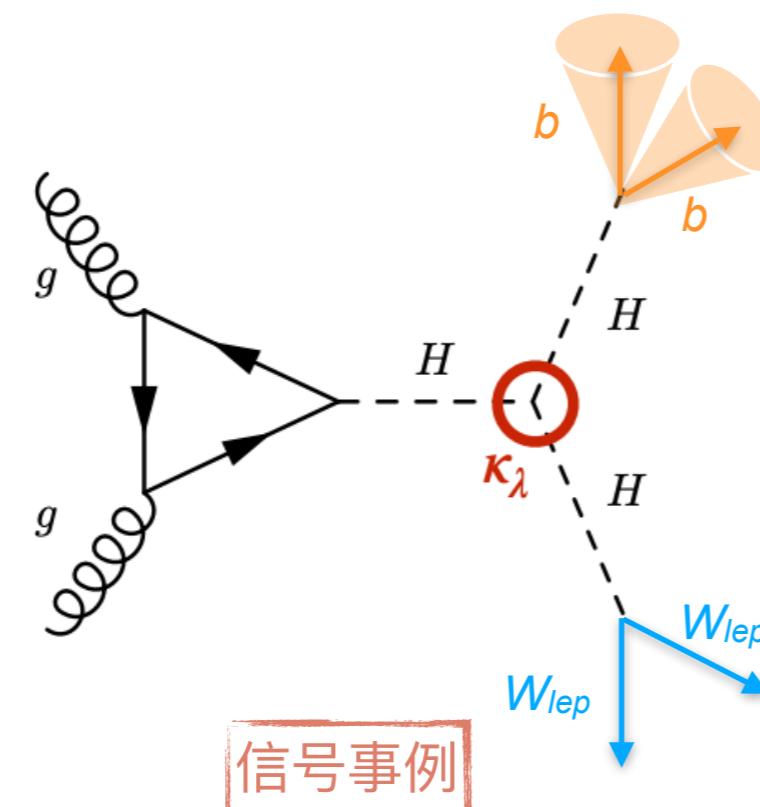
# 我们要处理的物理对象

- 首先让我们熟悉一下处理的物理对象
- 与 BDT 的分类任务相同，我们将分辨**双希格斯物理**中的物理信号  $\text{HH} \rightarrow \text{bbWW}$  与其最大的背景过程  $t\bar{t} \rightarrow \text{bWbW}$ 
  - ❖ 二者的末态完全相同：两个 b 夸克会形成**两个 b 喷注**；两个 W 玻色子研究其轻子化衰变，生成**两个轻子**（即电子或  $\mu$  子）和看不见的中微子
  - ❖ 但二者的事例结构稍有差别： $\text{HH}$  和  $t\bar{t}$  各自是头对头出射的（因极端相对论效应），所以前者 bb 和 WW 可能各自很近，后者两个 bW 对各自很近
  - ❖ 我们可以实验上测出信号和背景的末态信息，即两个 b 喷注 + 两个轻子的信息（四动量、喷注性质变量等等），那么，能否用神经网络技术区分两种过程？

物理动机：

双希格斯过程将出现  
**三希格斯耦合顶点**

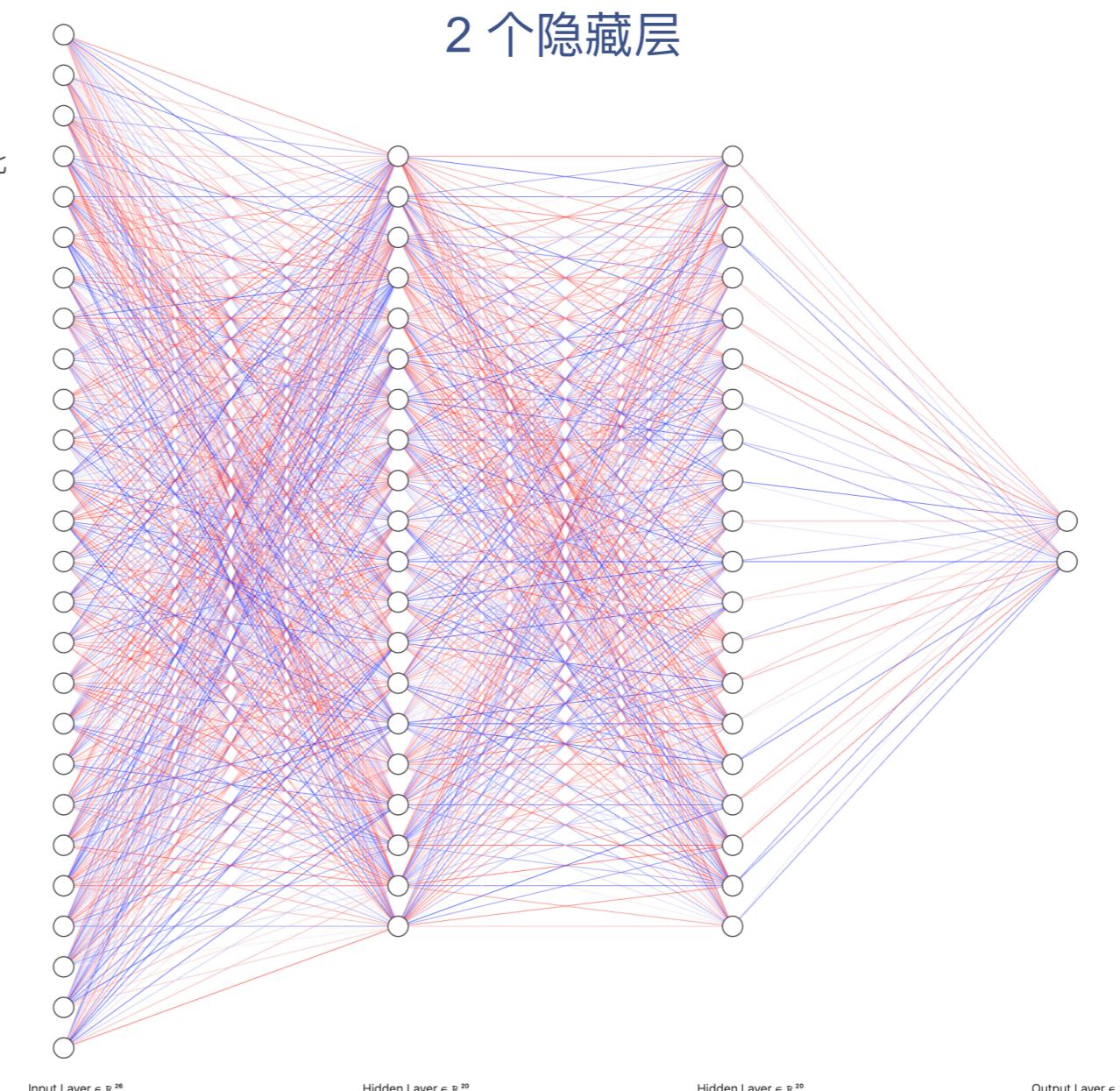
$\kappa_\lambda$ ，测出该过程出现  
概率即可**实验测定**  $\kappa_\lambda$



# 构建简单的前馈神经网络

26 个输入变量，分别为

bjet1_pt	1号b喷注的横向动量
bjet1_eta	赝快度 ( $\approx \operatorname{arctanh}(v_z/c)$ )
bjet1_phi	$xy$ 平面转角
bjet1_eratio	电磁量能器与强子量能器能量累计比
bjet1_mass	质量
bjet1_ncharged	所含带电强子个数
bjet1_nneutrals	所含中性强子个数
bjet2_pt	
bjet2_eta	
bjet2_phi	
bjet2_eratio	
bjet2_mass	
bjet2_ncharged	
bjet2_nneutrals	
lept1_pt	1号轻子的横向动量
lept1_phi	$xy$ 平面转角
lept1_eta	赝快度 ( $\approx \operatorname{arctanh}(v_z/c)$ )
lept1_charge	所带电荷
lept1_type	种类 (电子0, $\mu$ 子1)
lept2_pt	
lept2_phi	
lept2_eta	
lept2_charge	
lept2_type	
met	丢失横动量的大小
met_phi	丢失横动量的 $xy$ 平面转角大小



输出层: 二分类  
HH 信号为 (0, 1)  
 $t\bar{t}$  背景为 (1, 0)

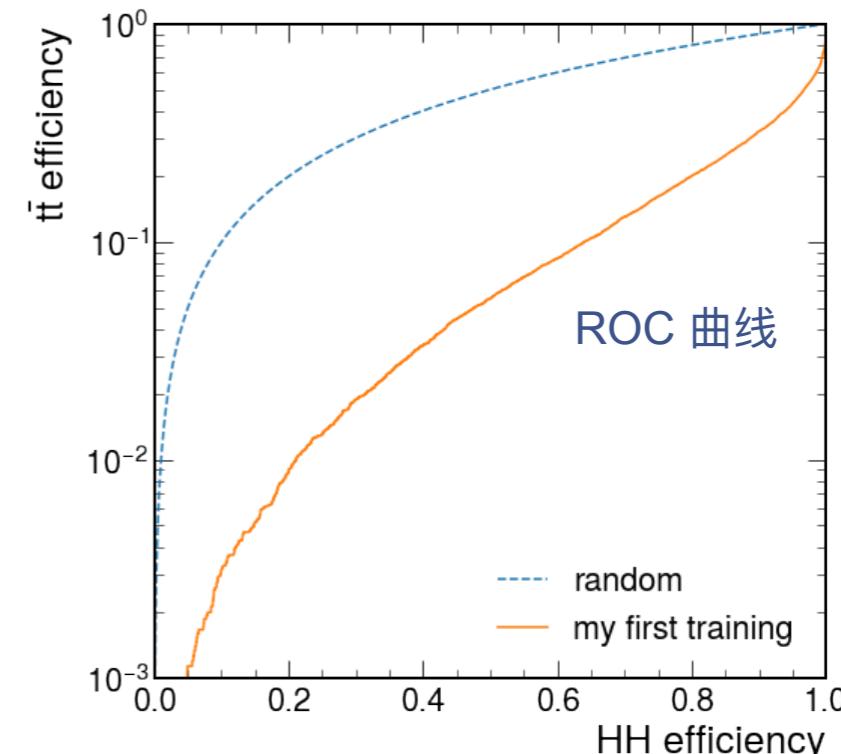
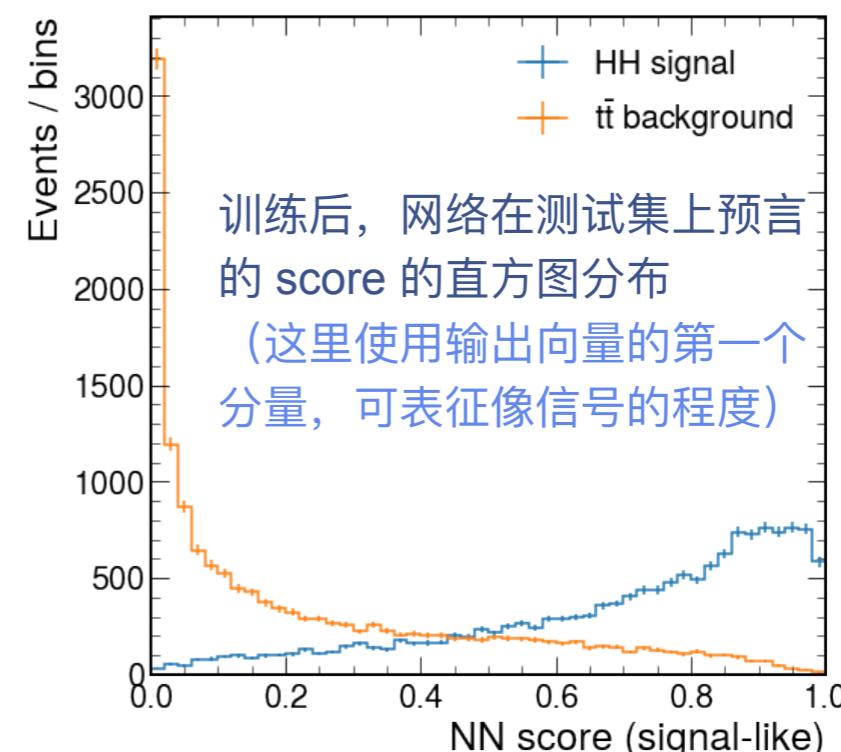
第 1 步: 首先运行 notebook: 01-visualize.ipynb

研究各个变量的分布直方图是什么样的, 信号和背景有什么差别?

# Hands-on: 开始训练!

第 2 步：运行 notebook: 02-dnn.ipynb 开始训练

- 使用 notebook 中的默认参数训练，可以看到不错的区分效果：分类的准确率达到 ~ 80% 🎉
- 用 notebook 绘制 ROC 曲线
- 动手尝试
  - ❖ 增加一个隐藏层，再次训练检查结果  
hands-on 任务，把新网络的预测分数绘制的 ROC 曲线与之前曲线画在一张图上
  - ❖ 我们先前的训练是否饱和？尝试增加训练的 epoch
  - ❖ 尝试给输入变量分别做平移、缩放的预处理，使各变量均值为 0，标准差为 1
  - ❖ 其它改进的方式？...



# 神经网络的调参

**人工智能训练师  
国家职业技能标准**

(2021 年版)

**1. 职业概况**

**1.1 职业名称**  
人工智能训练师

**1.2 职业编码**  
4-04-05-05

**1.3 职业定义**  
使用智能训练软件，在人工智能产品实际使用过程中进行数据库管理、算法参数设置、人机交互设计、性能测试跟踪及其他辅助作业的人员。

**1.4 职业技能等级**  
本职业共设 5 个等级，分别为：五级/初级工、四级/中级工、三级/高级工、二级/技师、一级/高级技师。

**1.5 职业环境条件**  
室内，常温。

**1.6 职业能力特征**  
具有一定的学习能力、表达能力、计算能力；空间感、色觉正常。

**1.7 普通受教育程度**  
初中毕业（或相当文化程度）。

**1.8 培训参考学时**  
五级/初级工 60 标准学时；四级/中级工 50 标准学时；三级/高级工 40 标准学时；二级/技师 40 标准学时；一级/高级技师 30 标准学时。

**1.9 职业技能鉴定要求**

**1.9.1 申报条件**  
具备以下条件之一者，可申报五级/初级工：

(1) 累计从事本职业或相关职业<sup>①</sup>工作 1 年（含）以上。

---

① 相关职业：人工智能工程技术人员、呼叫中心服务员、电子商务师等职业。下同。

来源：<http://www.mohrss.gov.cn/wap/zc/zqyj/202106/W020210617509883457681.pdf>

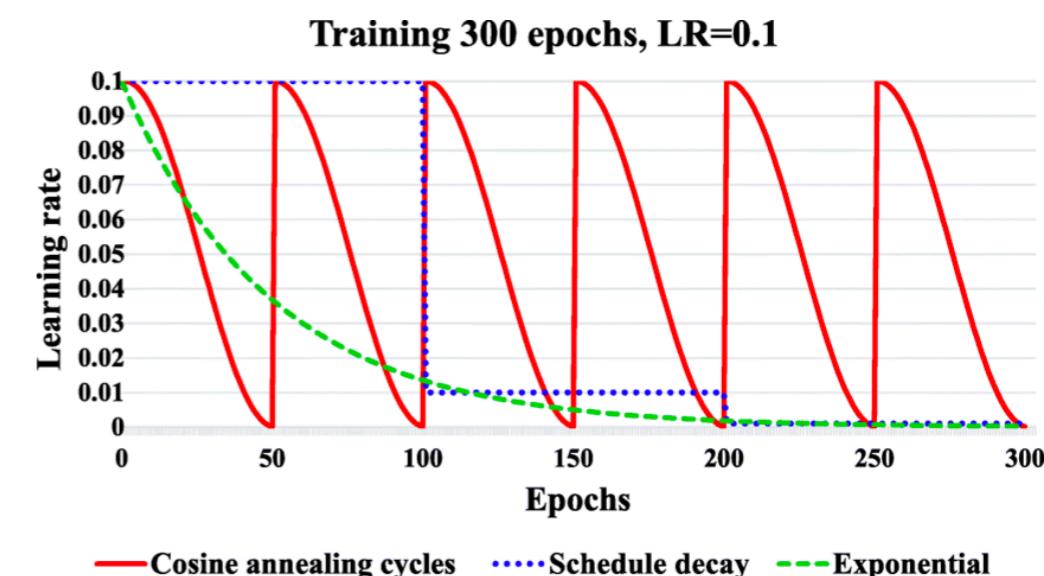
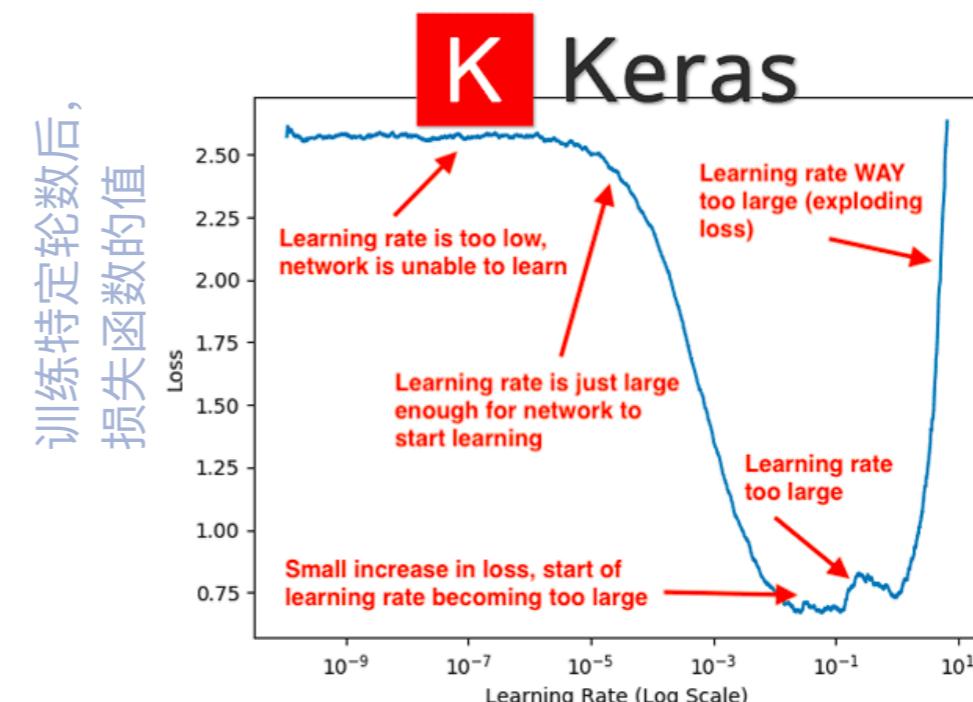
# 神经网络的调参

## → 学习率 (learning rate)

- ❖ learning rate 是最重要的超参数
- ❖ 反映训练的每一步按照梯度更新参数的步长，可以以指数为跨度进行调节，最后挑选始终的值

## → 学习率的衰减 (decay rate)

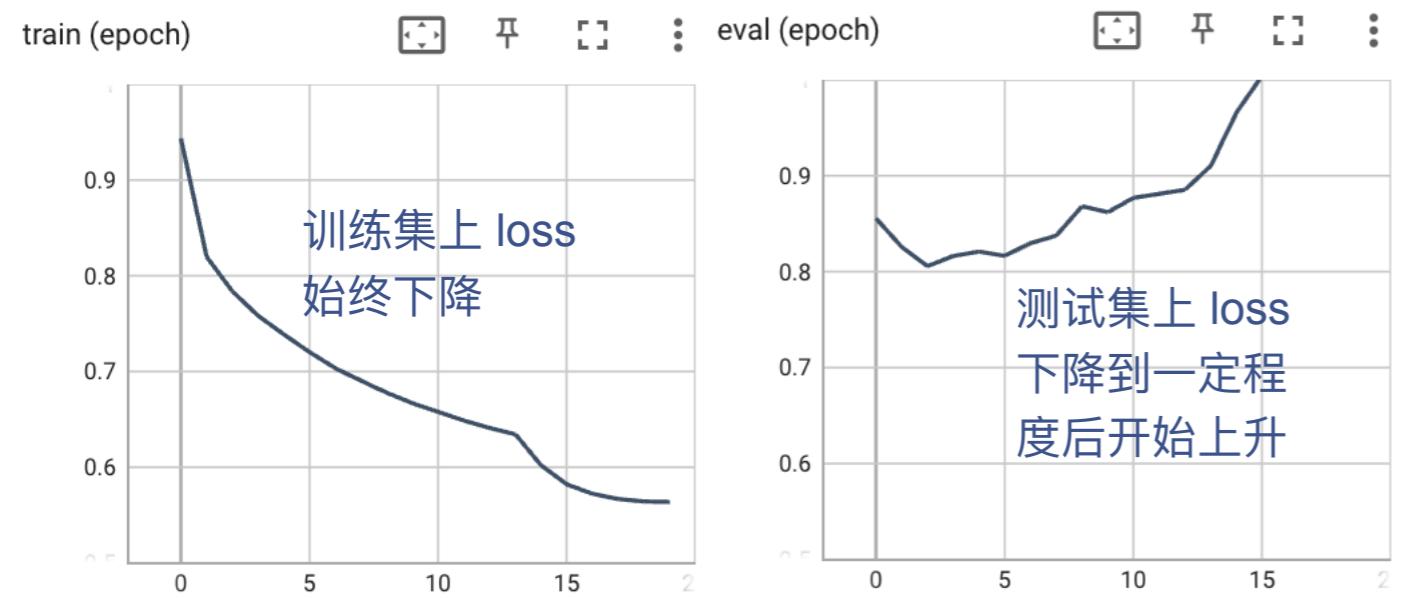
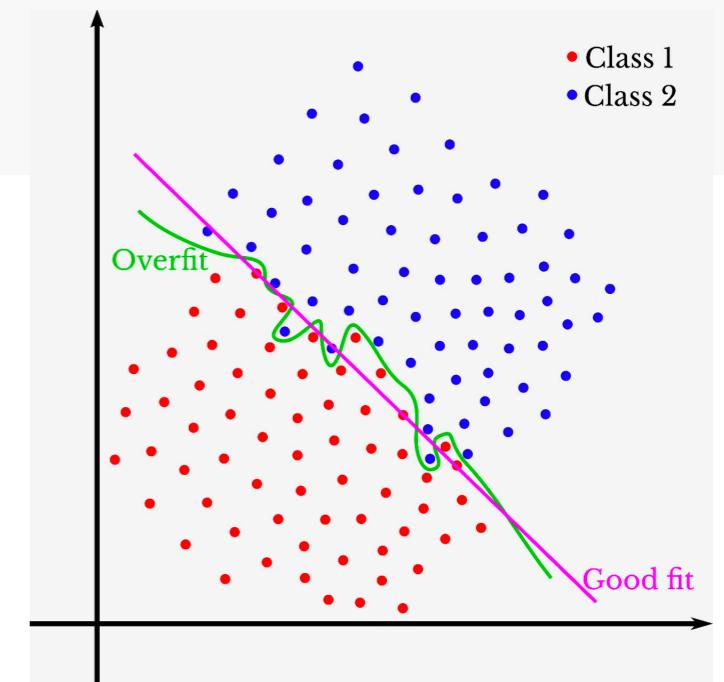
- ❖ 学习率选用何种模式衰减？阶梯式衰减、指数衰减、余弦退火衰减，在不同的场景下各有应用
- ❖ 对于基础的神经网络，选择简单的阶梯式或指数衰减即可



# 神经网络的调参-有关过拟合

## → 过拟合

- ❖ 表现：loss 在训练样本上始终下降，但每轮 epoch 结束后在测试样本上运行发现不再下降，甚至有所上升



### ❖ 原因：

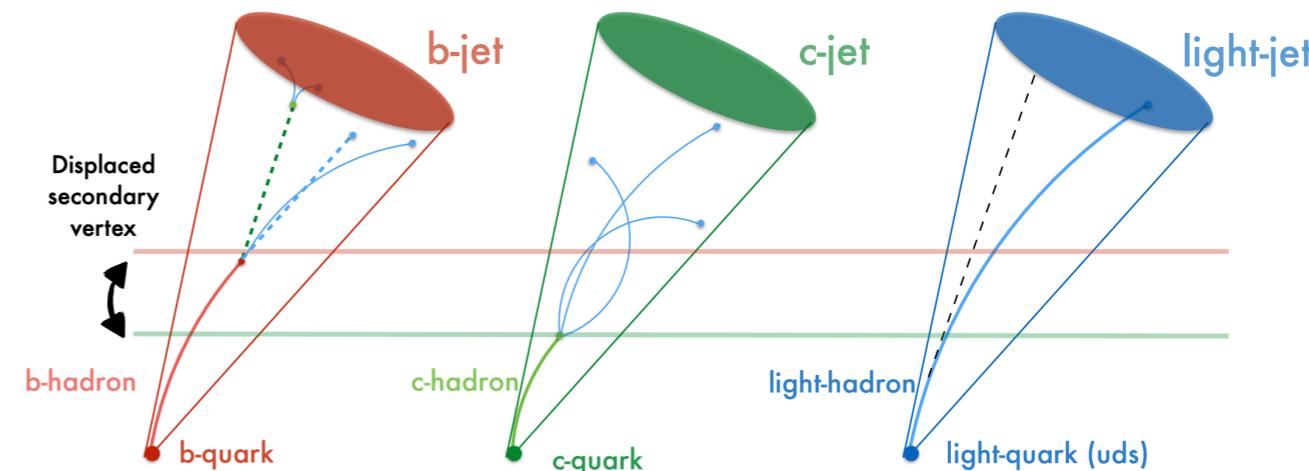
- ▶ 模型设计过大，有很多冗余参数
- ▶ 训练的样本量太小，无法支持深层神经网络进行训练（经验上，NN 比 BDT 先天需更多样本）

### ❖ 改进方案：

- ▶ 减小模型尺寸（如减少层数、减小每层的大小）
- ▶ 加 dropout：一种训练技巧，每批训练时随机删除一些神经元 → keras 里有相应设置

# 深入神经网络在 CMS 中的前沿应用

- CMS 在喷注鉴别的任务中首先部署了前沿的神经网络架构
- ❖ 喷注鉴别属于传统的多分类问题，因此最早展开前沿神经网络的尝试
- ❖ 采用底层变量进行训练（构成每个喷注的 **~100 个粒子信息直接作为信息输入**，而不是描述喷注的高级变量（如我们之前采用的变量）



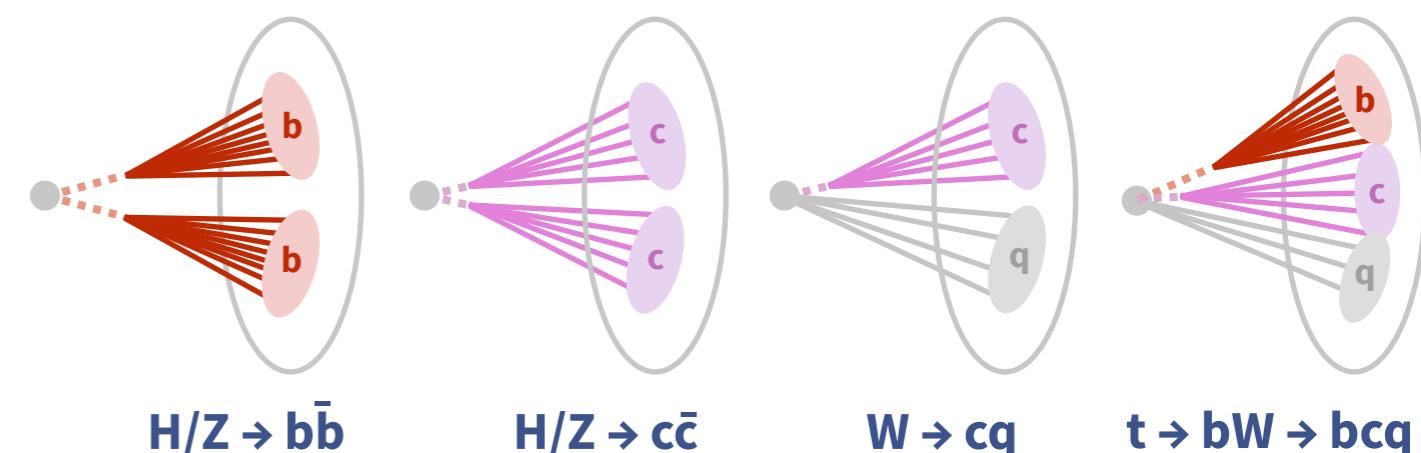
**喷注鉴别：**即判断一个喷注来自一个 b 夸克、c 夸克、或是轻夸克或胶子 u/d/s/g

- 它们主要差别在于产生强子的数目，以及次级顶点偏移的距离（次级顶点是由 b/c/轻味道强子飞出一定距离后才衰变导致的，因寿命不同，所以有差异）

相关材料 [链接](#)

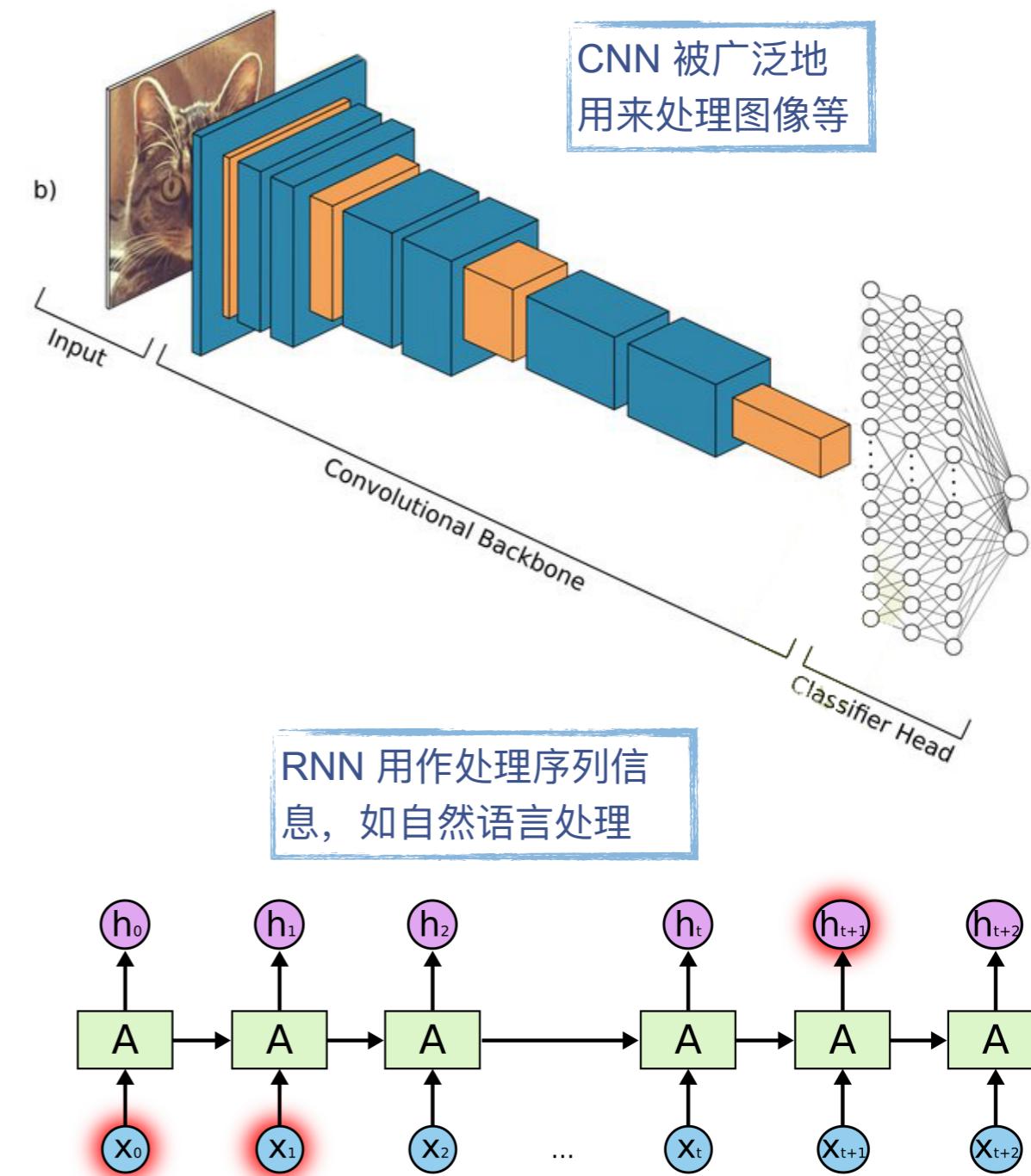
相关材料 [链接](#)

**粗喷注鉴别：**选择用更大的喷注半径重建末态，则一个粗喷注可能是由 2 到 3 个原始夸克导致的，而它们来源于一个共振态粒子（Higgs、Z、W 玻色子等）



# 深入神经网络在 CMS 中的前沿应用

- CMS 在喷注鉴别的任务中首先部署了前沿的神经网络架构
  - ❖ 喷注鉴别属于传统的多分类问题，因此最早展开前沿神经网络的尝试
  - ❖ 采用底层变量进行训练（构成每个喷注的 ~100 个粒子信息直接作为信息输入，而不是描述喷注的高级变量（如我们之前采用的变量）
  - ❖ 用更大的、更复杂的、有特定特征的神经网络模式（如卷积神经网络 CNN、循环神经网络 RNN，及近期发展的图神经网络 GNN）
    - ▶ 这些网络相比传统的前馈神经网络 (feed-forward NN) 性能优越，也被很多独立分析所采用

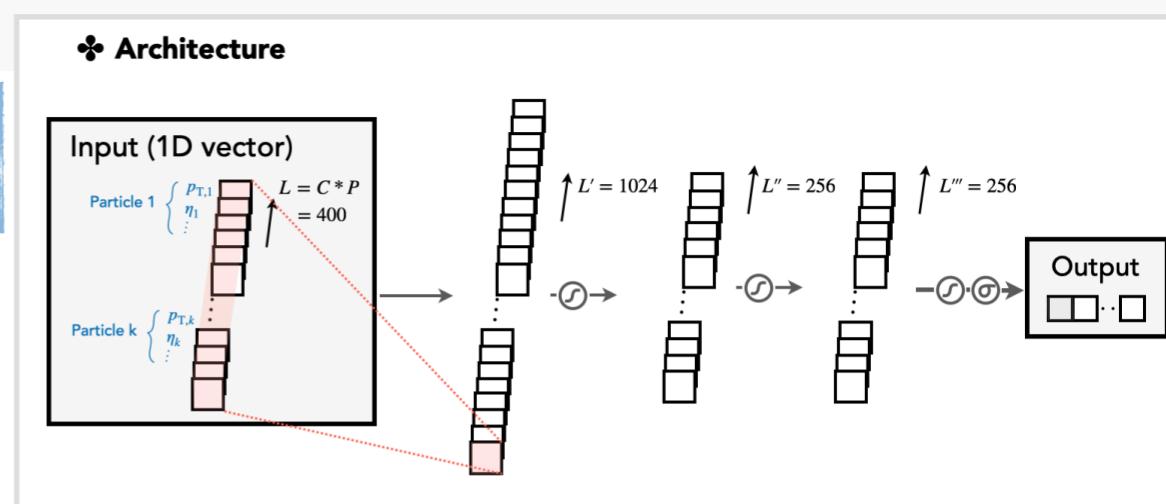


# 训练一个复杂的网络

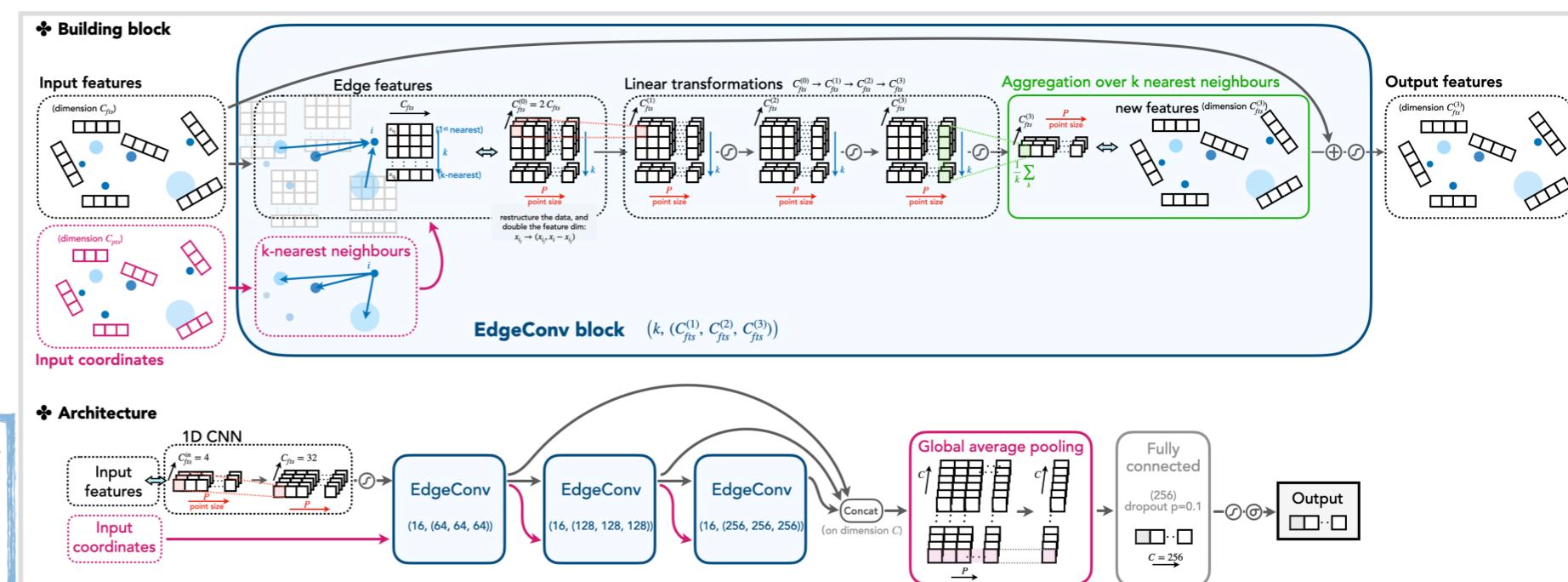
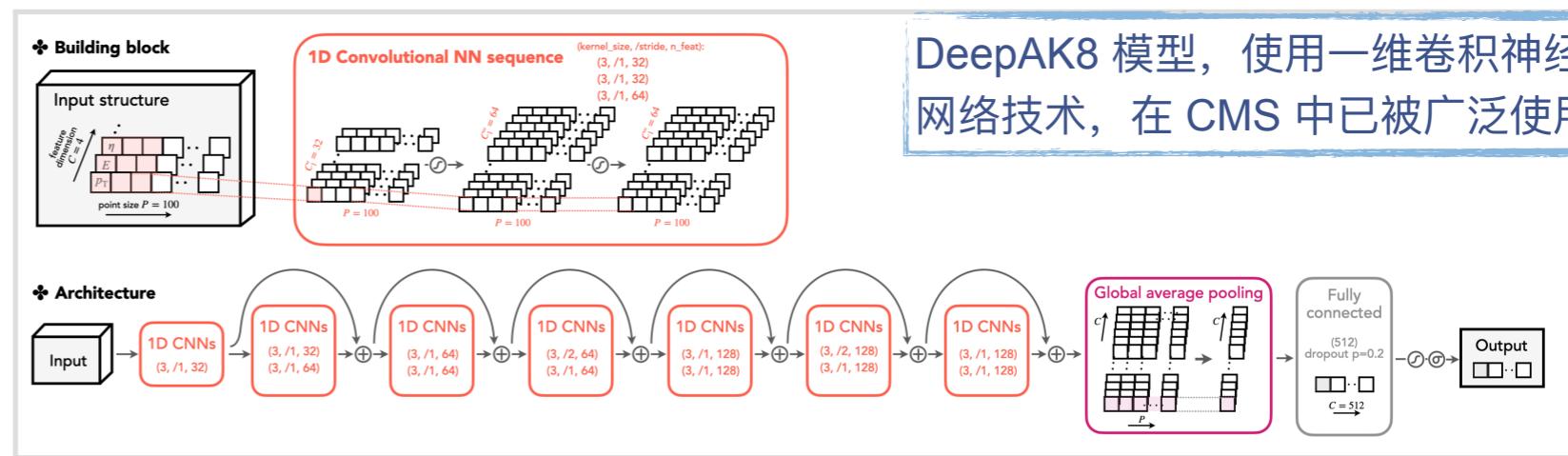
详见 CMS ML 文档

- 我们将简单地重复 CMS 粗喷注鉴别的训练工作
- [CMS ML 文档](#) 介绍了如何使用复杂网络
  - ❖ 用更先进的训练框架：PyTorch
  - ❖ 提供了三种神经网络模型的训练样例
  - ❖ 今天我们尝试简单重复第一个例子（使用 CPU 训练）

一个简单的前馈神经网络的例子



DeepAK8 模型，使用一维卷积神经网络技术，在 CMS 中已被广泛使用



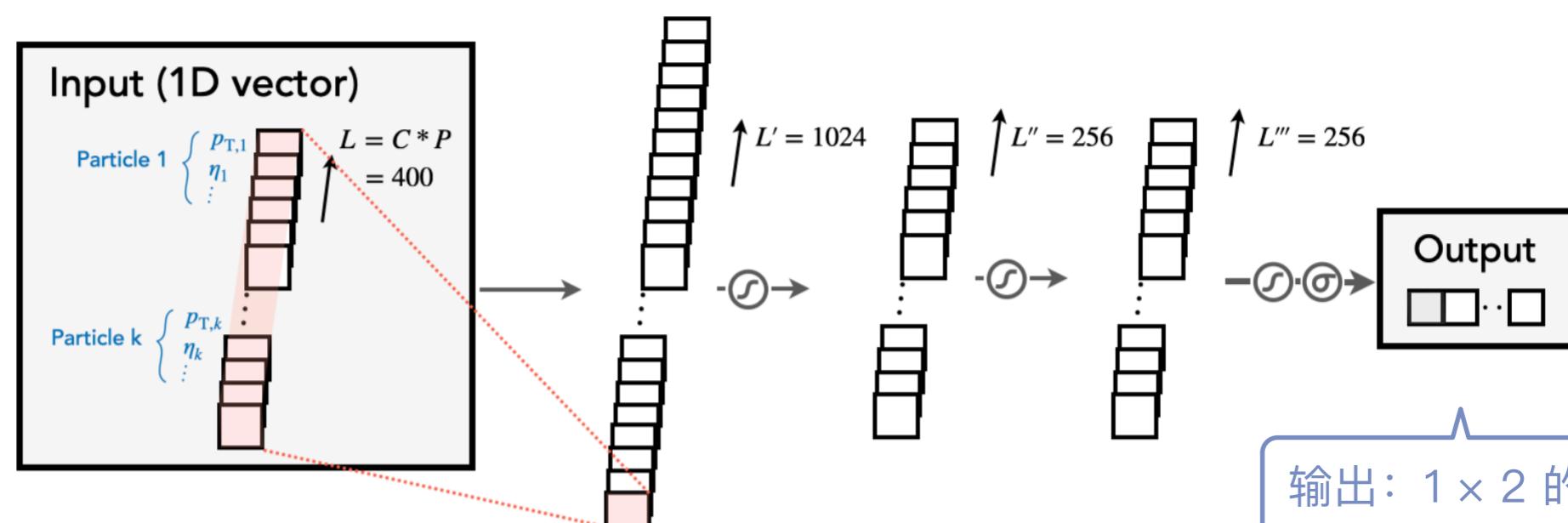
ParticleNet 模型，使用图卷积技术，是目前效果最佳的喷注鉴别神经网络，开始在 CMS 被广泛了解并使用

# Hands-on: 训练用于粗喷注鉴别的复杂网络

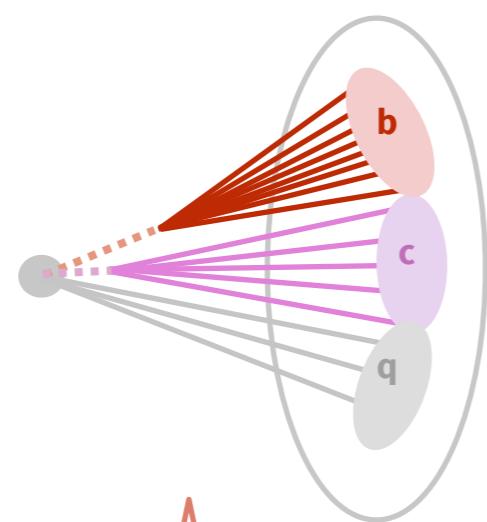
hands-on : 运行  
03-jet-tagging.ipynb

输入: 每个喷注  
取前 100 个粒  
子, 各 4 个粒子  
信息, 排列成  $1 \times 400$  的向量

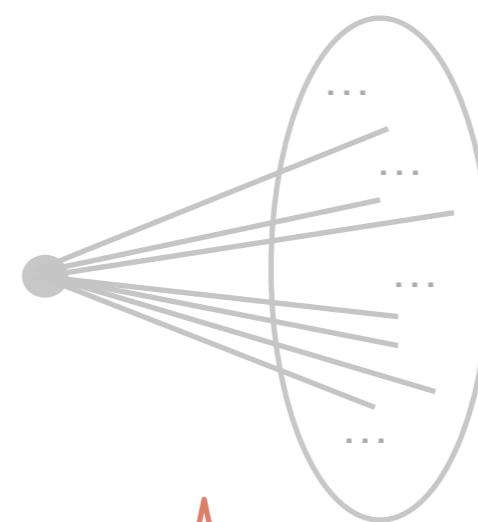
## Architecture



输出:  $1 \times 2$  的向量,  
反映 isTop / isQCD



信号喷注: 来自 t 夸克的 top 喷  
注, 因  $t \rightarrow bW \rightarrow bcq/bqq$  的衰变  
特征而形成典型的 3 叉结构



背景喷注: 由强相互作用随机  
辐射出的夸克/胶子, 进入粗  
喷注中, 称为 QCD 喷注

# 总结

→ 机器学习在高能物理中应用广泛

- ❖ 传统的 BDT 方法仍在广泛用于分类任务中，已有几十年的历史

→ 深度学习的飞速发展为高能物理带来更多使用场景

- ❖ 多隐藏层的深度神经网络结构开始取代 BDT 被用于分析的事例鉴别任务
- ❖ 新型神经网络架构（如 DeepAK8、ParticleNet 模型）紧跟深度学习领域前沿，直接采用底层信息作为输入，用大样本量训练一些通用任务（如喷注鉴别）
- ❖ 同时也在开辟除了传统的分类、回归任务以外的新方法，如使用生成模型直接生成末态粒子，代替现有 particle-flow 重建算法等

# 附录：了解深度学习前沿

## → 计算机视觉课程：

- ❖ [UMich EECS 498: Deep Learning for Computer Vision](#) [ [Youtube](#)]  
[ [bilibili](#)]

## → 经典文章导读系列：

- ❖ Bilibili: 跟李沐学AI [ [link](#)]