

# 深度学习在CTR预估业务中的应用

新浪微博AI Lab 资深算法专家

张俊林



AI时代的移动技术革新

Era of AI: Innovations in Mobile Technologies

APICloud

## 张俊林

### 新浪微博AI Lab 资深算法专家

曾在阿里巴巴、百度、用友担任资深技术专家和技术总监等职位。中国中文信息学会理事，研发兴趣集中在：搜索技术、推荐系统、社交挖掘、自然语言处理等方面。本科毕业于天津大学，之后在中科院软件所直接攻读博士学位，研究方向是信息检索理论与自然语言处理，曾在 ACL/COLING等国际著名会议发表多篇学术论文。

技术书籍《这就是搜索引擎：核心技术详解》（该书荣获全国第十二届出版优秀图书奖）、《大数据日知录：架构与算法》的作者。



- 当深度学习遇到CTR任务
- Factorization Machine模型
- 深度学习CTR模型
  - 深度学习解决CTR任务的几个关键问题
  - 离散特征如何表达
  - 两种网络结构
- 模型选择与训练优化



# 深度学习:各个领域的成功



## 人脸识别



## 物体识别



## 语音识别



## 机器翻译



## 风格转换



## 图片生成





# CTR任务例子

Feature vector $x$															Target $y$							
$x^{(1)}$	1	0	0	...	1	0	0	0	...	0.3	0.3	0.3	0	...	13	0	0	0	0	...	5	$y^{(1)}$
$x^{(2)}$	1	0	0	...	0	1	0	0	...	0.3	0.3	0.3	0	...	14	1	0	0	0	...	3	$y^{(2)}$
$x^{(3)}$	1	0	0	...	0	0	1	0	...	0.3	0.3	0.3	0	...	16	0	1	0	0	...	1	$y^{(2)}$
$x^{(4)}$	0	1	0	...	0	0	1	0	...	0	0	0.5	0.5	...	5	0	0	0	0	...	4	$y^{(3)}$
$x^{(5)}$	0	1	0	...	0	0	0	1	...	0	0	0.5	0.5	...	8	0	0	1	0	...	5	$y^{(4)}$
$x^{(6)}$	0	0	1	...	1	0	0	0	...	0.5	0	0.5	0	...	9	0	0	0	0	...	1	$y^{(5)}$
$x^{(7)}$	0	0	1	...	0	0	1	0	...	0.5	0	0.5	0	...	12	1	0	0	0	...	5	$y^{(6)}$
	A	B	C	...	TI	NH	SW	ST	...	TI	NH	SW	ST	...	Time	TI	NH	SW	ST	...		
	User				Movie					Other Movies rated						Last Movie rated						

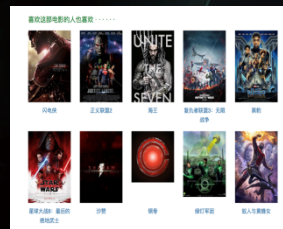
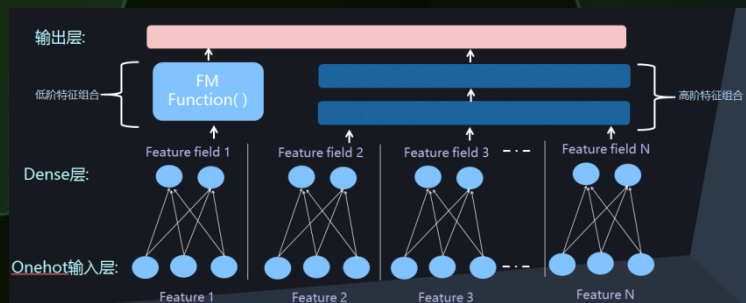


# CTR任务的特点

- 大量离散特征
- 大量高维度稀疏特征
- 特征工程：特征组合对于效果非常关键



# 当CTR预估遇到深度学习





- 当深度学习遇到CTR任务
- Factorization Machine模型
- 深度学习CTR模型
  - 深度学习解决CTR任务的几个关键问题
  - 离散特征如何表达
  - 两种网络结构
- 模型选择与训练优化



# 线性模型：思路及问题

$$\text{Linear: } \hat{y}(x) := w_0 + \sum_{i=1}^n w_i x_i$$

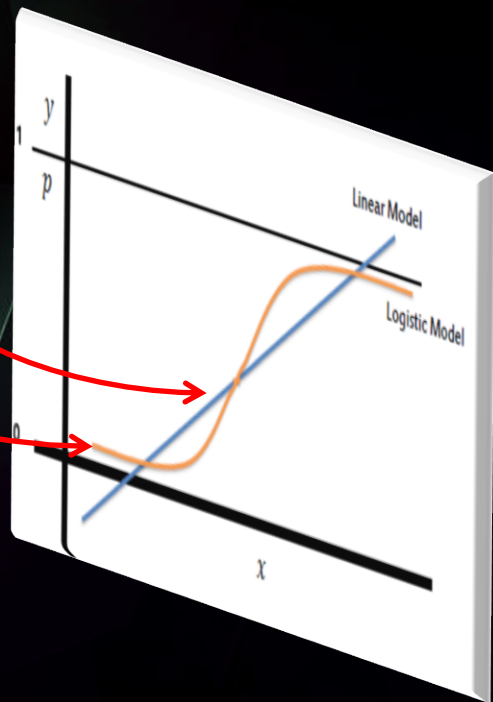
$$\text{LR: } \hat{y}(x) = \frac{1}{1 + w_0 \exp(-w^T x)}$$

优势：

简单；可解释；易扩展

缺点：

难以捕获特征组合



# 线性模型改进：加入特征组合

改进版本:  $\hat{y}(x) := w_0 + \sum_{i=1}^n w_i x_i + \underbrace{\sum_{i=1}^n \sum_{j=i+1}^n w_{i,j} x_i x_j}_{\text{两两特征组合}}$

两两特征组合

优势：

直接将两两组合特征引入模型

缺点：

组合特征泛化能力弱

$w_{i,j} = 0$  if 在训练数据中  $x_i x_j = 0$



# FM模型

$$\text{FM: } \hat{y}(x) := w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n \langle v_i, v_j \rangle x_i x_j$$



LR模型



Dense化两两特征组合

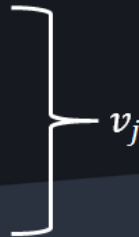
$$\langle v_i, v_j \rangle := \sum_{f=1}^k v_{i,f} \cdot v_{j,f} =$$

0.1 0.2 0.6 0.8 0.1 0.2



$v_i$

0.4  
0.2  
0.4  
0.2  
0.4  
0.2



$v_j$



# FM模型

$$\langle v_i, v_j \rangle := \sum_{f=1}^k v_{i,f} \cdot v_{j,f} =$$

$v_1$ :	0.3	0.2	0.6	0.8	0.1	0.2
$v_2$ :	0.1	0.8	0.6	0.8	0.4	0.6
$v_3$ :	0.4	0.2	0.7	0.2	0.1	0.2
$v_4$ :	0.1	0.2	0.6	0.8	0.5	0.2
$v_{n-1}$ :	0.3	0.2	0.6	0.8	0.1	0.2
$v_n$ :	0.5	0.8	0.9	0.8	0.4	0.6

$$w_{i,j} = \langle v_i, v_j \rangle \neq 0$$

even if 在训练数据中  $x_i x_j = 0$

only if 在训练数据中存在k使得  $x_i x_k \neq 0$

FM模型泛化能力强





- 当深度学习遇到CTR任务
- Factorization Machine模型
- **深度学习CTR模型**
  - 深度学习解决CTR任务的几个关键问题
  - 离散特征如何表达
  - 两种网络结构
- 模型选择与训练优化



# 深度学习CTR模型要解决的几个关键问题

- CTR任务特点：大量离散特征的代表问题
- CTR任务特点：如何快速处理大量高维度稀疏特征？（OneHot 2 Dense）
- 特征工程：如何从手工到自动？（深度学习的优势）
- 特征工程：如何捕获和表达两两组合特征？（FM机制神经网络化）
- 特征工程：如何捕获和表达多组组合特征？（利用Deep网络）



# CTR任务中的特征类型

- 连续特征
  - 收入，身高，体重.....
  - 适合DNN处理
- 离散特征
  - 职业，性别，毕业学校.....
  - 不适合DNN处理



# 离散特征如何让DNN可以处理？

- 直观思路：离散特征使用Onehot表达

feature\_time=friday

[0,0,0,0,1,0,0]

feature\_male=female

[0,1]

feature\_product=13584

[0,0.....1,0.....0]

x-feature=[0,0,0,0,1,0,0, 0,1, 0,0.....1,0.....0]



# 离散特征如何让DNN可以处理？

- 直观思路：离散特征使用Onehot表达

feature\_time = fr

[0,0,0,0,1,0,0]

feature\_product = 13584

[0,0.....1,0.....0]

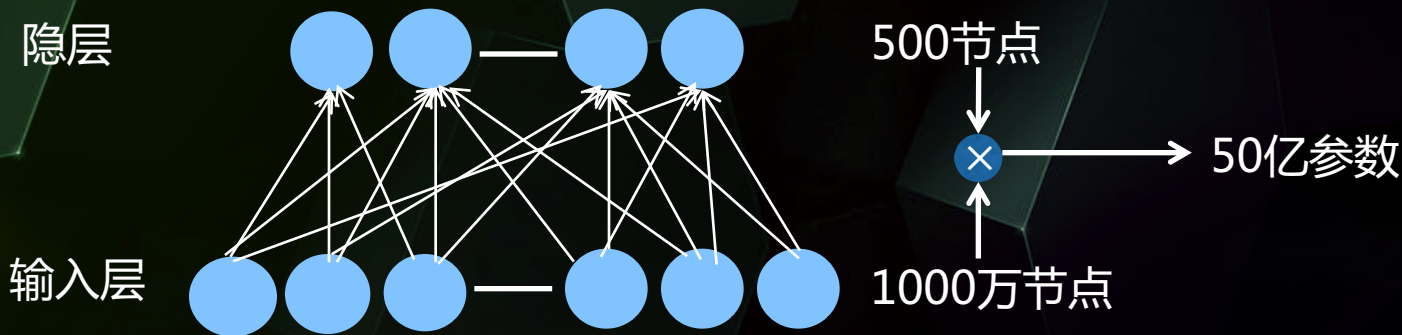
x-feature = [0,0,0,0,1,0,0, 0,1, 0,0.....1,0.....0]





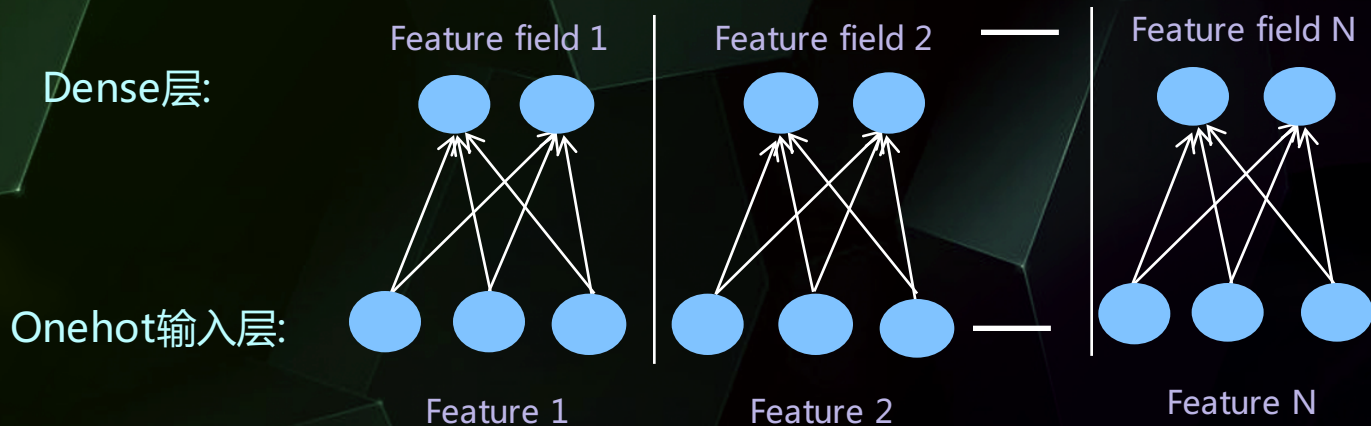
# 离散特征如何让DNN可以处理？

- Onehot作为DNN输入的问题：CTR预估任务里不可行



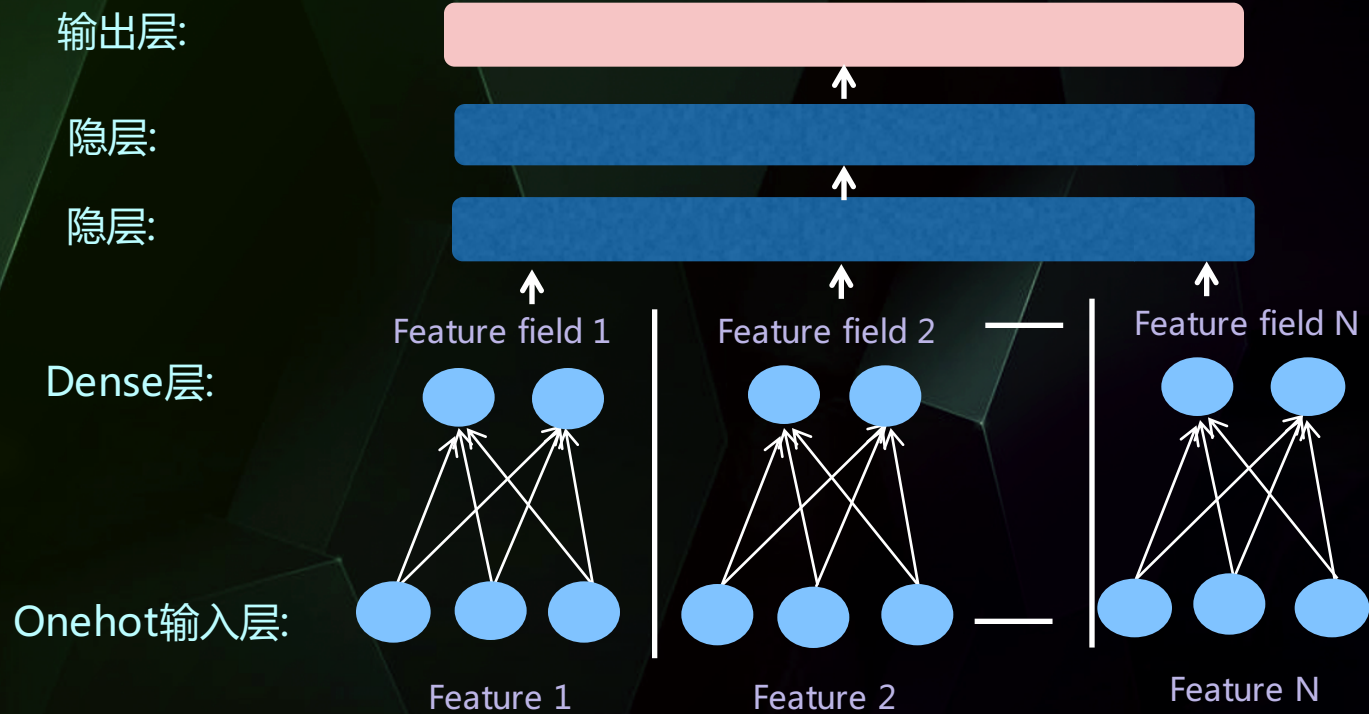
# 离散特征如何让DNN可以处理？

- 解决思路：从OneHot到Dense Vector

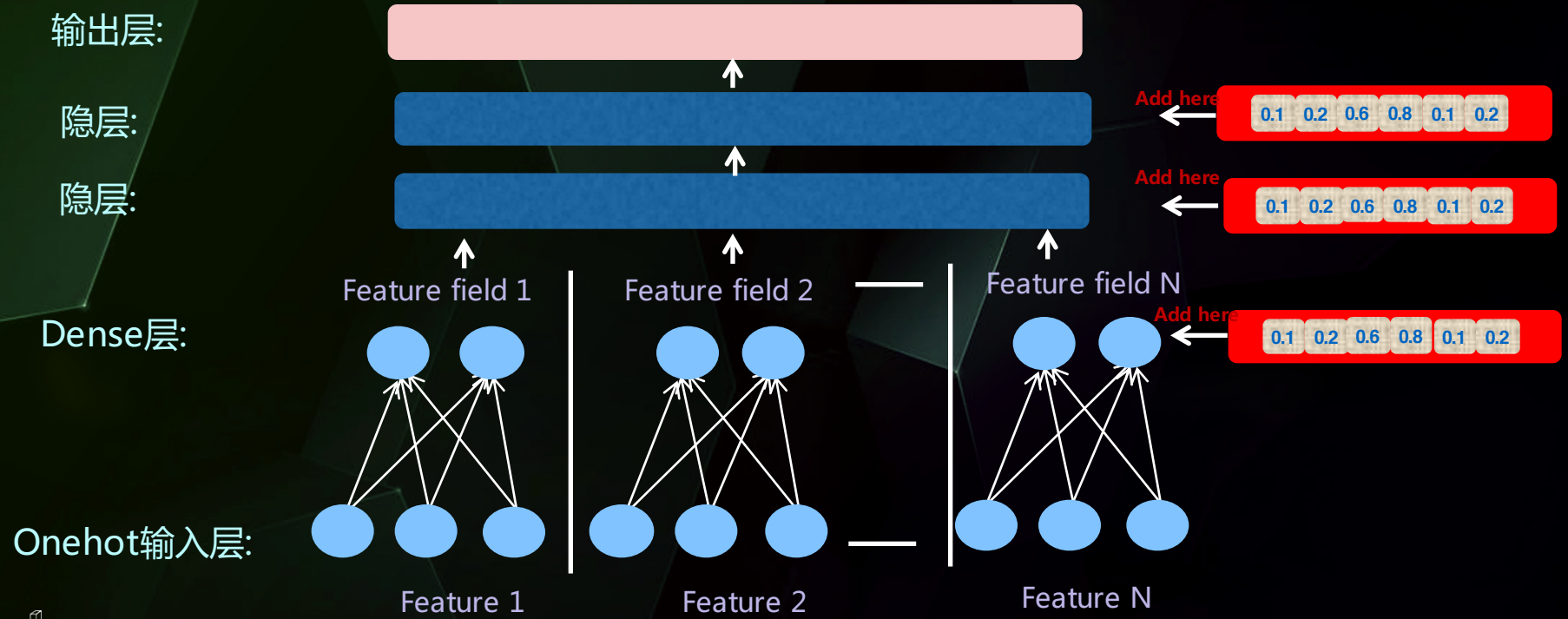


基本思想:避免全链接, 分而治之

# 形成DNN结构

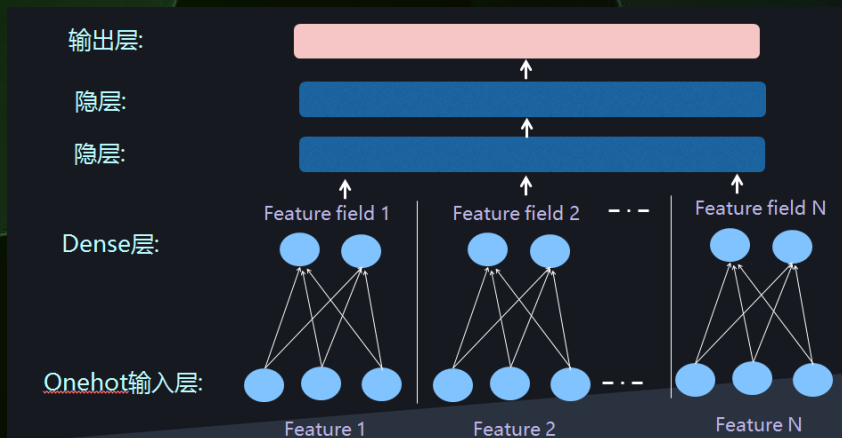


# 再加入连续特征



# 这是通用的深层模型结构

这就是FNN模型：Factorisation-Machine Supported Neural Networks



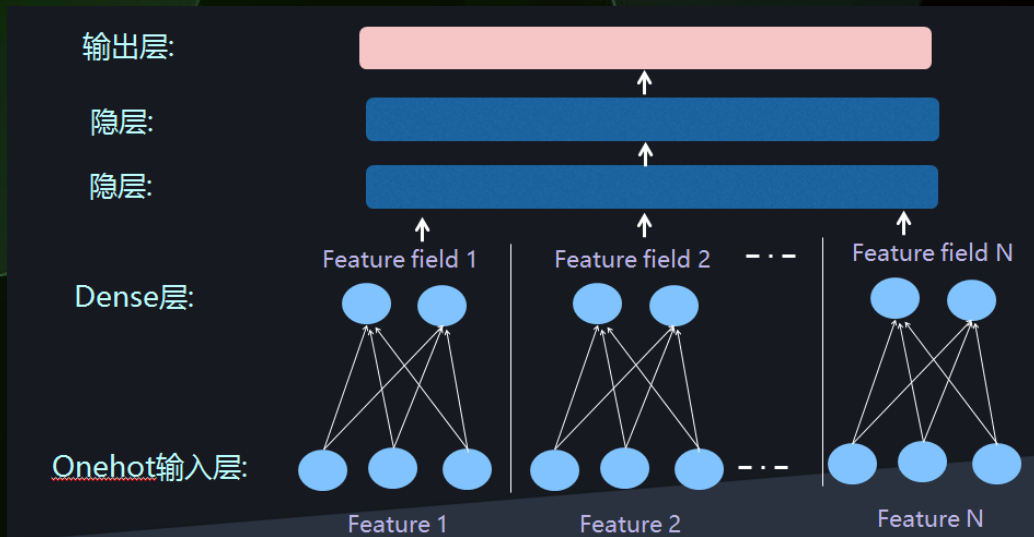
Wide&Deep模型的Deep部分：相同的结构

很多其它改进模型的Deep部分：相同的结构

几乎所有DL+CTR模型的输入部分:这种Onehot2Dense映射



# DNN输入问题解决了，但是.....



低阶和高阶特征组合隐含地体现在隐层

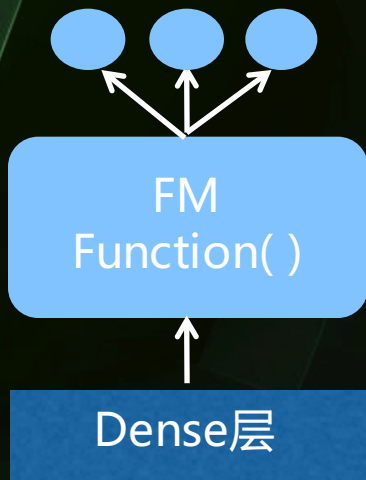
我们希望：把低阶特征组合单独建模

?



# 把低阶特征组合单独建模

- 首先需要：定义一个神经网络版的低阶特征组合模型



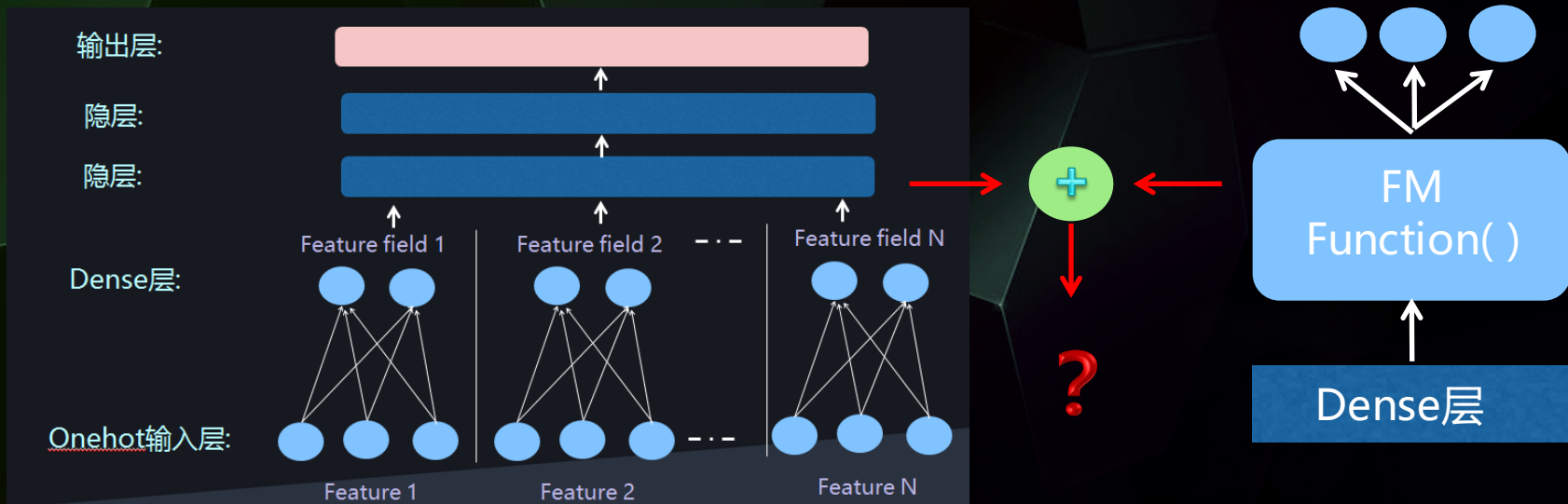
**1:** Function: feature \* feature  $\rightarrow R^d$

**2:** Function具有神经网络的表现形式：和现有网络融合

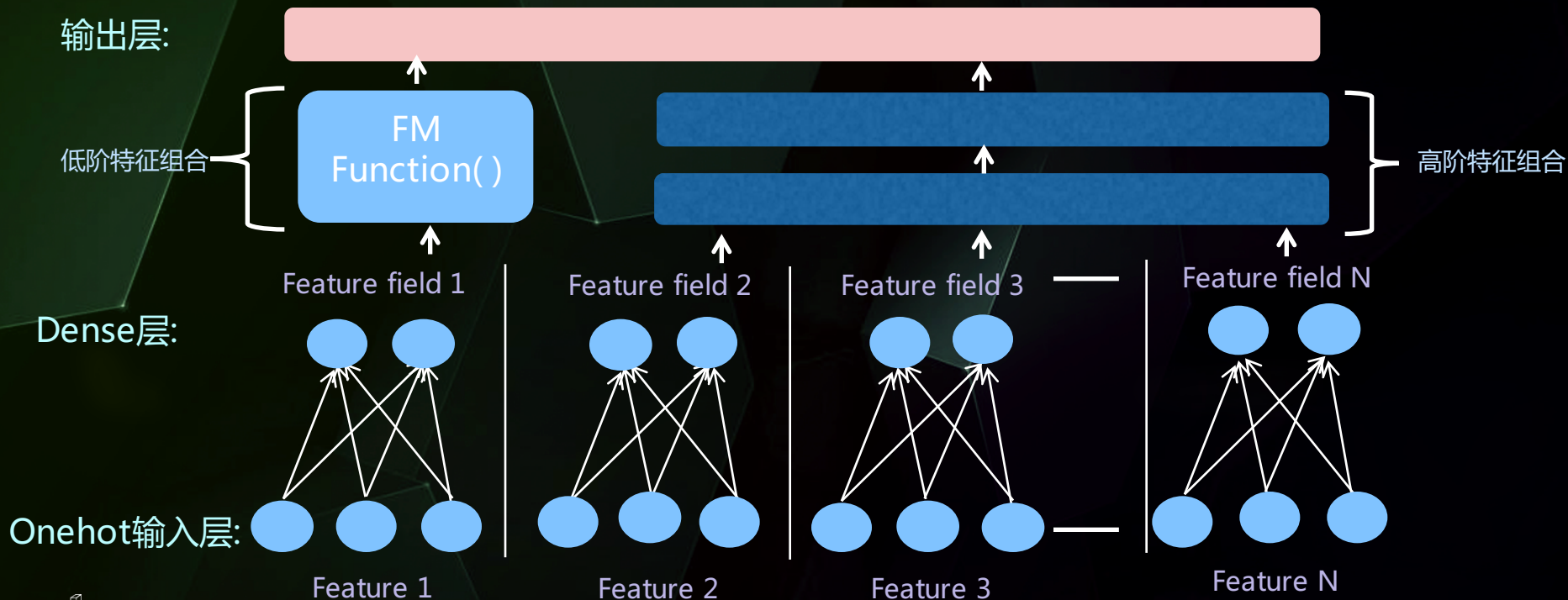
**3:** Function能够体现FM的思想：

$$\hat{y}(x) := w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n \langle v_i, v_j \rangle x_i x_j$$

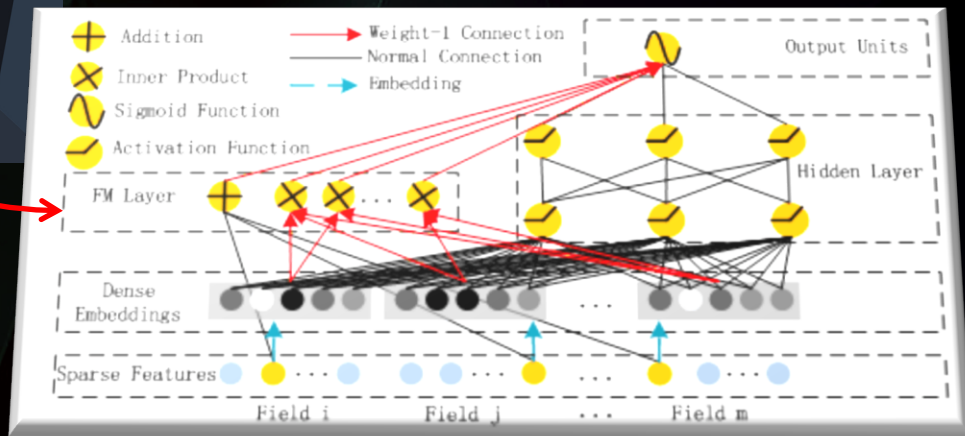
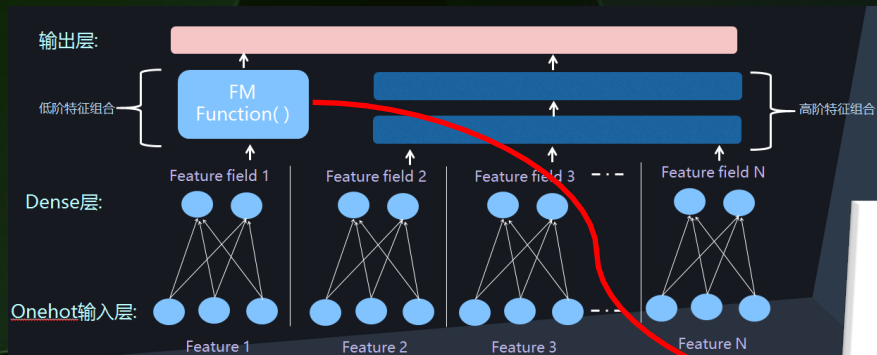
# 把低阶特征组合模型插入网络结构中



# 典型网络融合结构之一：并行结构



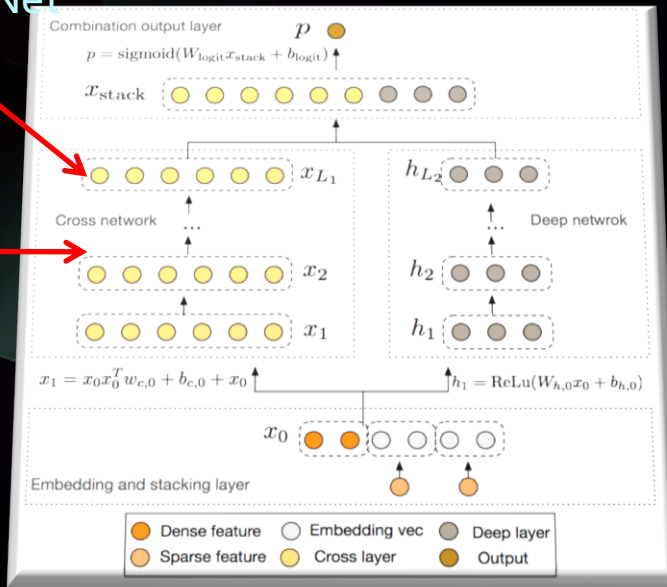
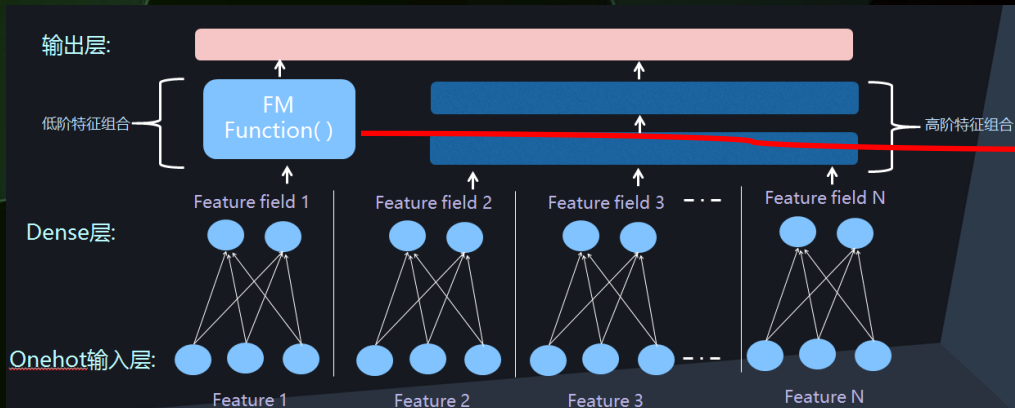
# 并行结构实例：DeepFM模型



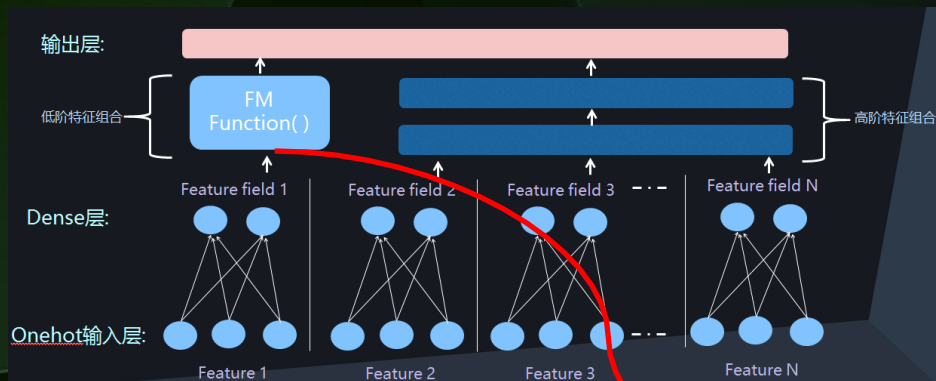


# 并行结构实例：Deep&Cross模型

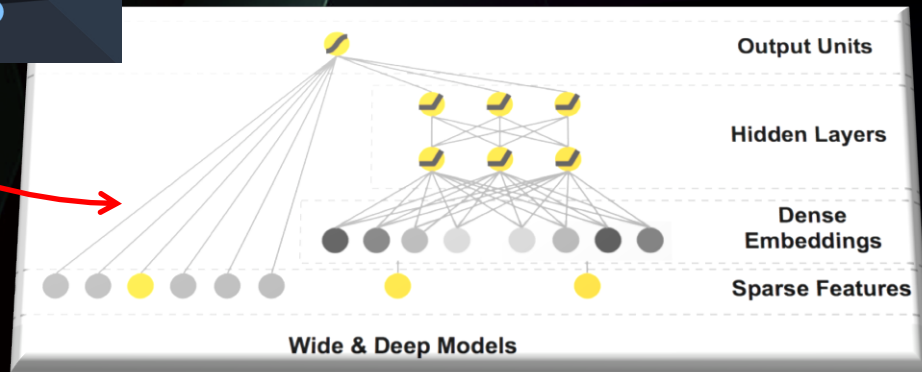
FM Function=Cross Network=ResNet



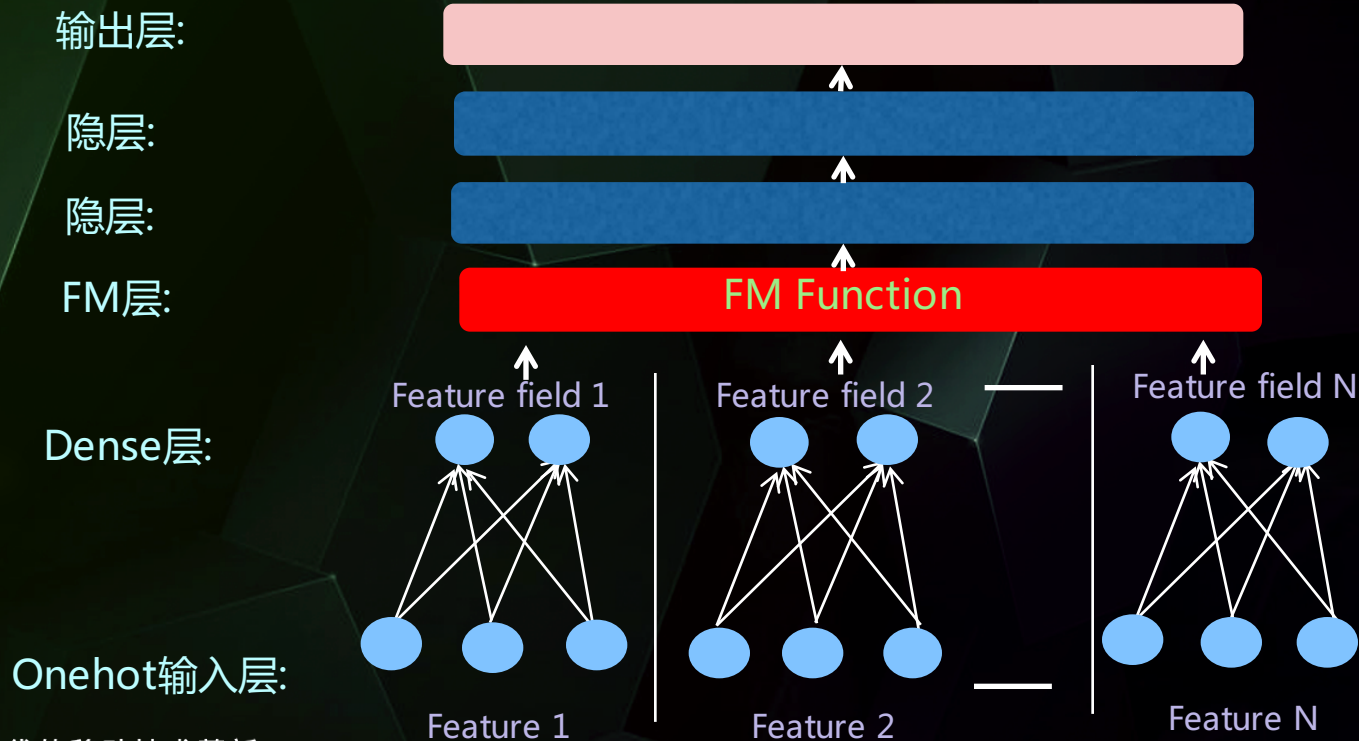
# 并行结构实例：Wide&Deep模型



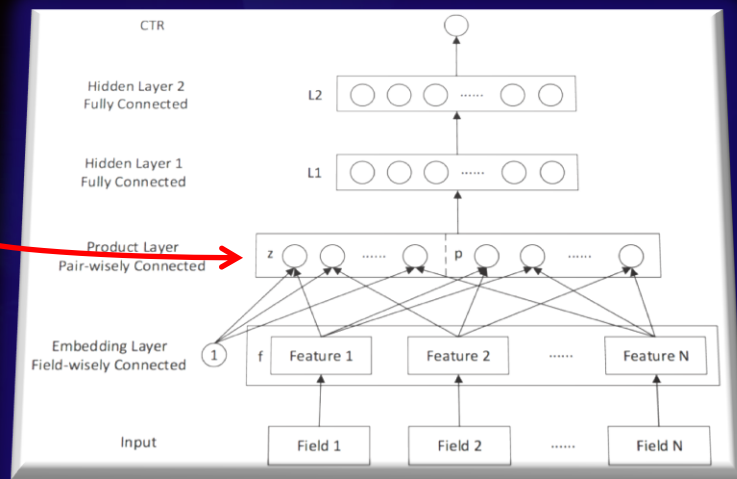
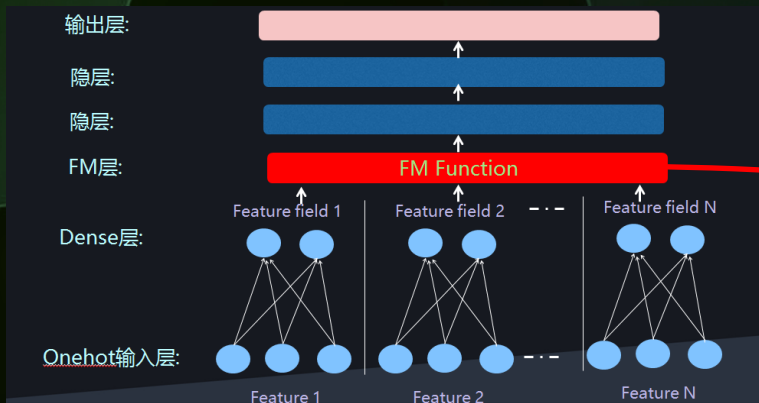
- 1:** Wide网络LR模型，无FM组合
- 2:** 两个不同的输入结构
- 3:** Wide部分手工特征工程+Cross 特征



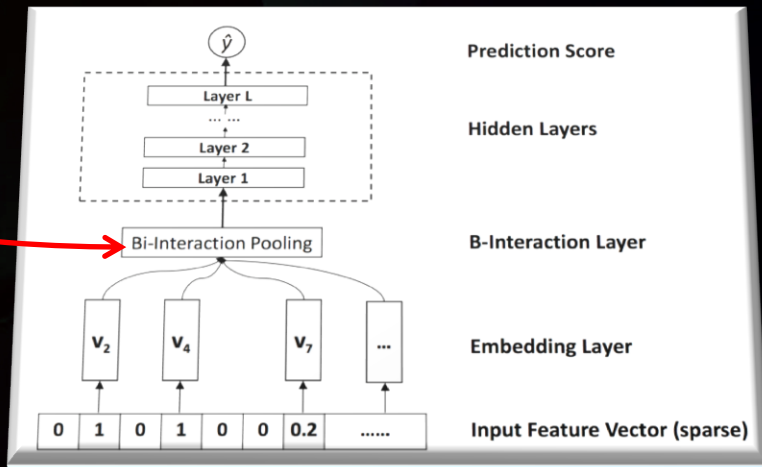
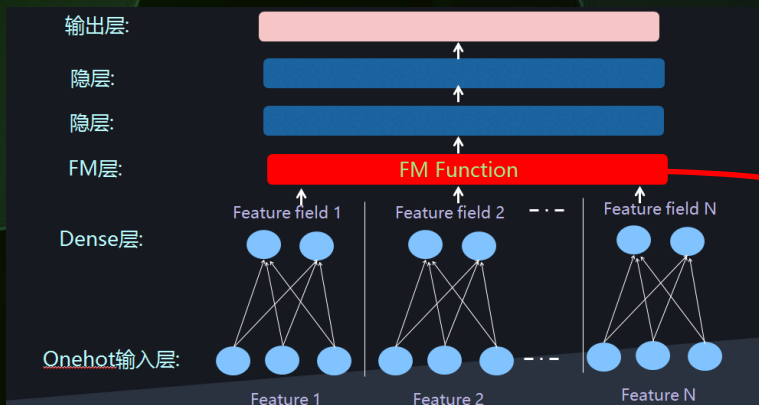
# 典型网络融合结构之二：串行结构



# 串行结构实例：PNN模型



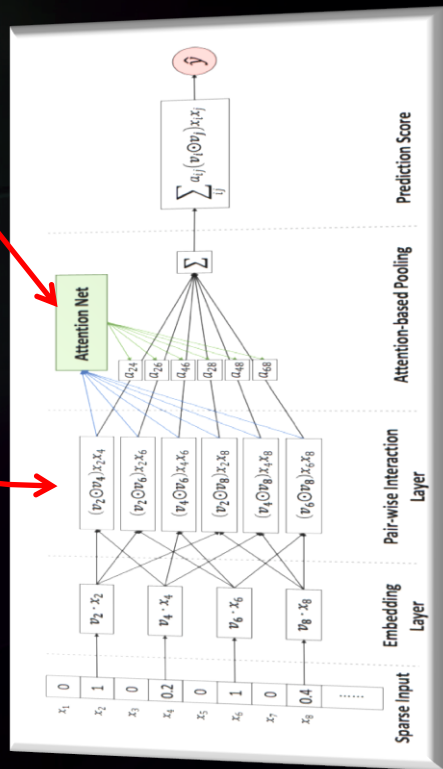
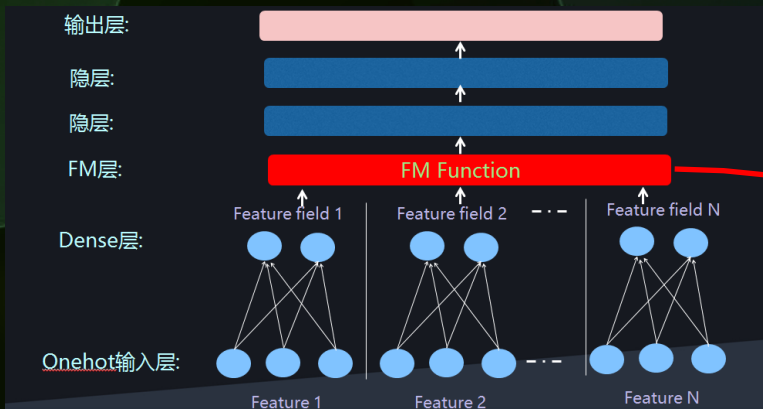
# 串行结构实例：NFM模型





# 串行结构实例：AFM模型

相比NFM模型只是多了针对组合特征的Attention



- 当深度学习遇到CTR任务
- Factorization Machine模型
- 深度学习CTR模型
  - 深度学习解决CTR任务的几个关键问题
  - 离散特征如何表达
  - 两种网络结构
- 模型选择与训练优化



# 关于Dense层的预训练

1: Wide & Deep模型Dense层需要预训练

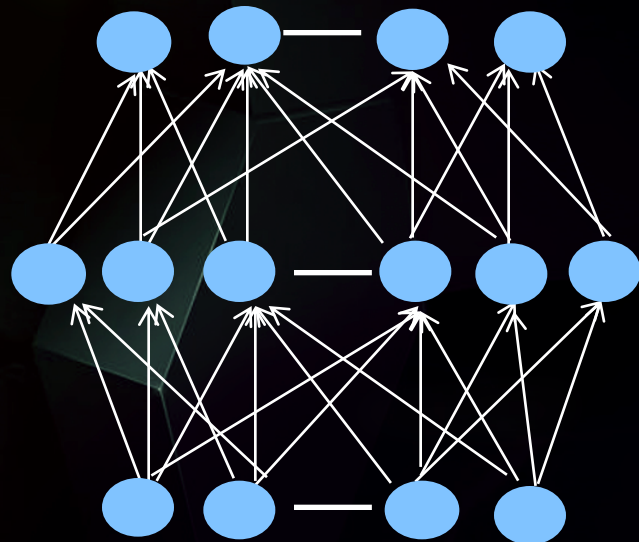
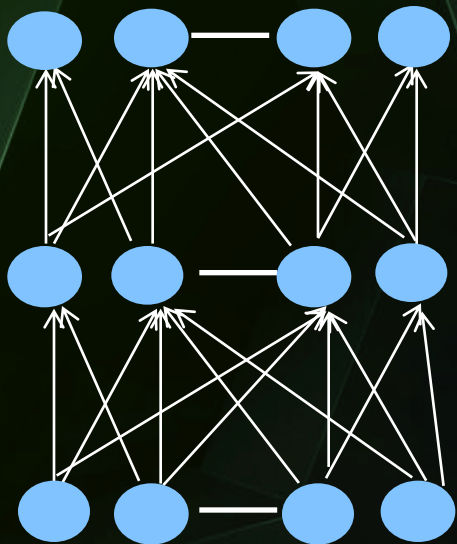
用FM初始化Onehot到Dense层的映射性能明显提升

2: 类似的FNN等无明显FM结构的模型Dense层需要预训练

3: 串行结构的模型Dense层需要预训练

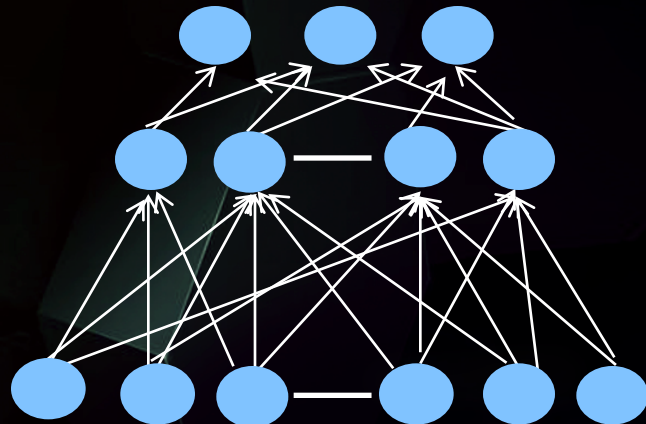
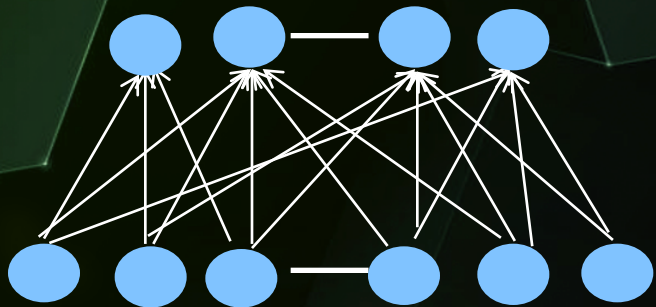


# 关于Deep网络隐层的网络结构



平行结构或者菱形结构效果较好

# 关于Deep网络隐层的层深



两层或三层







AI时代的移动技术革新

Era of AI: Innovations in Mobile Technologies

 APICloud

 IT大咖说  
知识共享平台

---

谢谢观看  
THANKS

APICloud