

[深度模型] Deep & Cross Network (DCN)

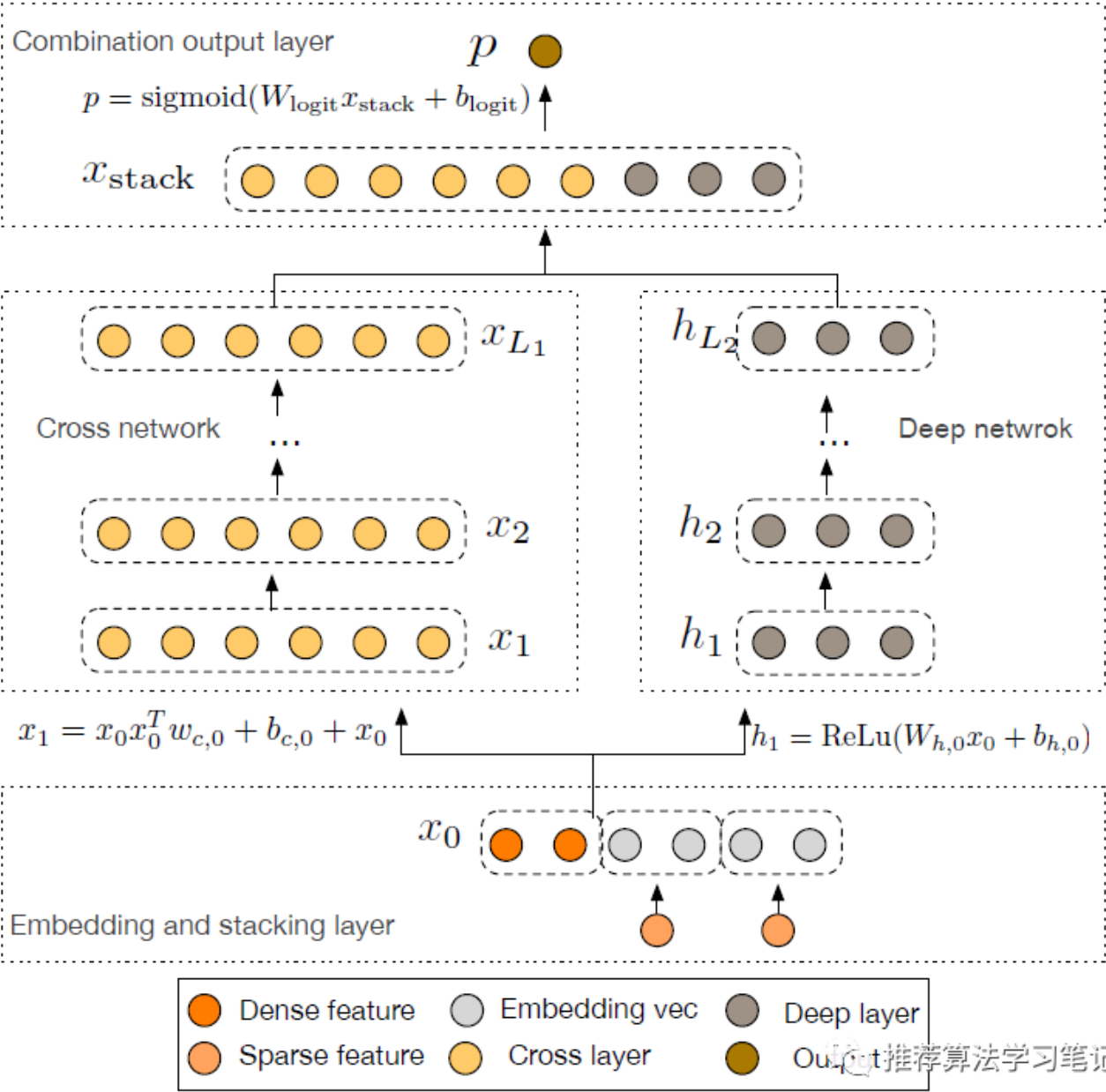
原创 xxxhuang 推荐算法学习笔记 5月6日

本文主要介绍的是经典的深度模型DCN。paper名称为《Deep & Cross network for Ad Click Predictions》

一. 概述

DCN是一个可以同时高效学习低维特征交叉和高维非线性特征的深度模型，不需要人工特征工程的同时需要的计算资源非常低。

DCN的模型结构如下图所示



推荐算法学习笔记

可以看到DCN分成4部分。最底下是“Embedding and stacking layer”，中间部分是“Cross network”和“Deep network”，最上面是“Combination output layer”。下面分别对每个部分进行讲解

二. Embedding and stacking layer

这一层是模型的输入层。我们知道输入的特征分为稠密和稀疏特征。

对于稀疏特征，我们一般采用one hot encoding的方式进行编码，然后再经过一层embedding将one hot embedding转换成稠密向量。

最后，将稠密特征和经过转换的稀疏特征对应的稠密向量concat起来组成模型的最终输入，也就是图中的 x_0

三. Cross Network

在cross network里面我们主要使用以下的公式进行一层一层的叠加：

$$\mathbf{x}_{l+1} = \mathbf{x}_0 \mathbf{x}_l^T \mathbf{w}_l + \mathbf{b}_l + \mathbf{x}_l = f(\mathbf{x}_l, \mathbf{w}_l, \mathbf{b}_l) + \mathbf{x}_l, \quad (3)$$

假如我们要叠加3层，则每一层计算如下

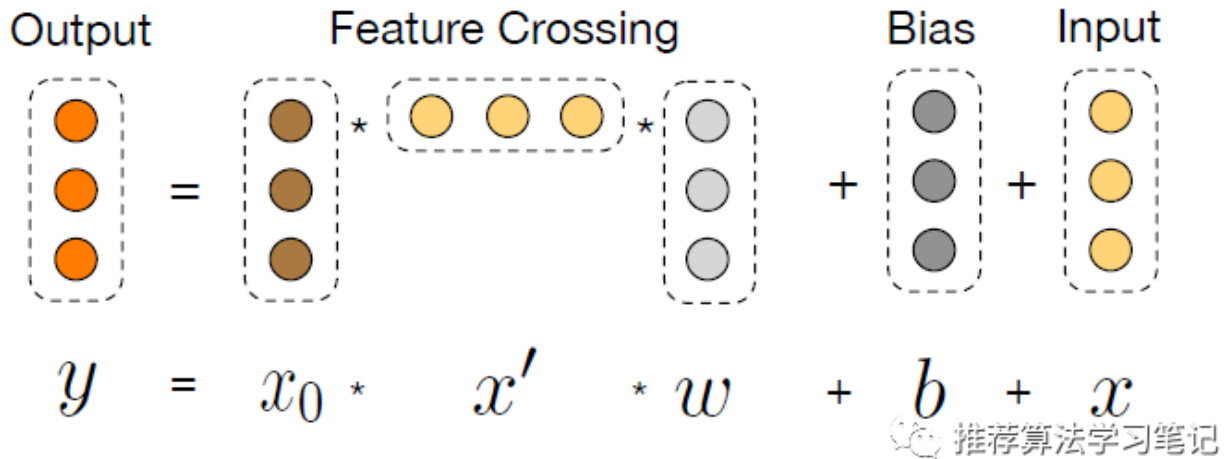
$$\mathbf{x}_1 = \mathbf{x}_0 \mathbf{x}_0^T \mathbf{w}_0 + \mathbf{b}_0 + \mathbf{x}_0$$

$$\mathbf{x}_2 = \mathbf{x}_0 \mathbf{x}_1^T \mathbf{w}_1 + \mathbf{b}_1 + \mathbf{x}_1$$

$$\mathbf{x}_3 = \mathbf{x}_0 \mathbf{x}_2^T \mathbf{w}_2 + \mathbf{b}_2 + \mathbf{x}_2$$

推荐算法学习笔记

可视化如下图所示



w和b是我们要学习的参数，可以看到在cross network里面，输入和输出的维度是一样的。

为什么这么设计？

cross network为1层的时候，我们可以得到的最高是2维的特征交叉；cross network为2层的时候，我们得到的是最高3维的特征交叉；cross network为3层的时候，我们得到的是最高4维的特征交叉；以此类推。。。

因此cross network以一种参数共享的方式，通过对叠加层数的控制，可以高效地学习出低维的特征交叉组合，避免了人工特征工程。

对cross network更深入的分析可参考paper

四. Deep Network

deep network部分就是传统的全连接前馈神经网络，用来学习高维非线性特征交叉组合。公式如下所示

$$h_{l+1} = f(W_l h_l + b_l),$$

推荐算法学习笔记

五. Combination Output Layer

最后，将cross network和deep network的输出concat起来，通过一个logit层，对于一个二分类问题，公式如下图所示

$$p = \sigma \left([x_{L_1}^T, h_{L_2}^T] w_{\text{logits}} \right),$$

推荐算法学习笔记

其中

$$\sigma(x) = 1/(1 + \exp(-x)).$$

损失函数为

$$\text{loss} = -\frac{1}{N} \sum_{i=1}^N y_i \log(p_i) + (1 - y_i) \log(1 - p_i) + \lambda \sum_l \|w_l\|^2, \quad (6)$$



公式最右边一项为l2正则

六. 总结

以上便是DCN的全部内容，如果有问题，欢迎随时联系。