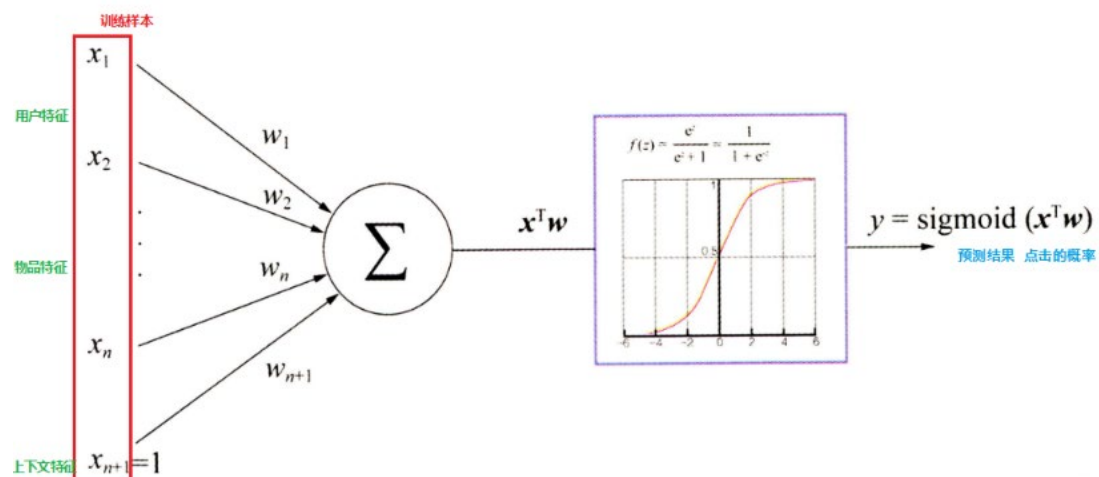


## 经典排序模型——GBDT+LR

原理：利用 GBDT 自动进行特征筛选和组合，生成新的离散特征向量，作为 LR 模型的输入，进行预测

**LR**：在线性回归的基础上加入 sigmoid 函数，将推荐问题转换为概率预测问题

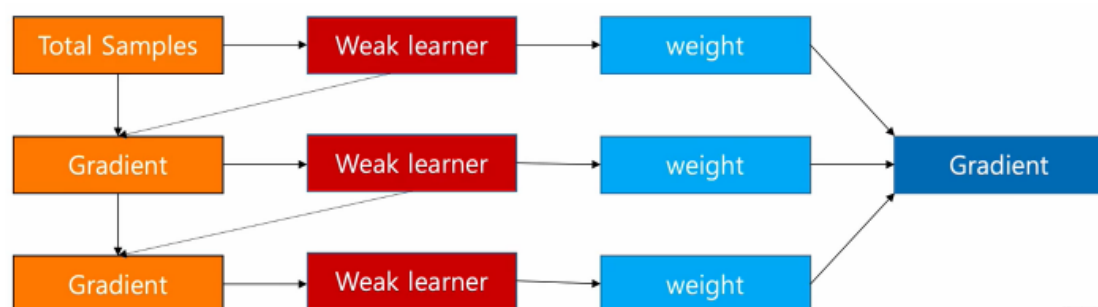
在推断过程中，需要将特征转换为数值型向量，将优化目标转换为二分类问题（是否点赞、收藏、转发等），然后训练模型并预测



优化的对象是每个特征的权重。首先进行初始化，然后计算梯度并更新，经过多次迭代后得到最终权重

缺点：无法进行特征交叉，模型简单准确率低，处理非线性数据麻烦，需要人工特征组合

**GBDT**：每一轮的训练都会生成一个弱分类器，在上一轮训练残差的基础上进行。通过降低训练残差提高分类器精度



GBDT 可以解决二分类和回归问题，但损失函数不同，前者是交叉熵函数，后者是 L2

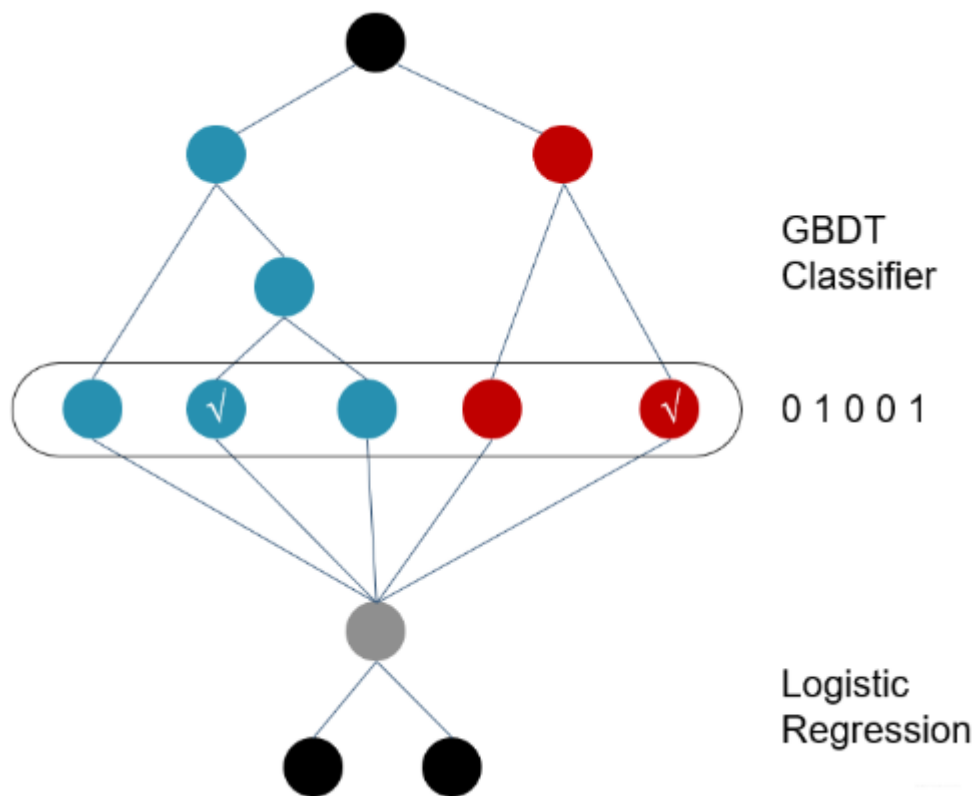
基本步骤：初始化 GBDT，得到一个叶子节点；循环生成决策树（计算负梯度值得到残差，用回归树拟合残差，计算叶子节点的输出，更新模型）

优缺点：

树的生成过程可以理解为自动进行多维度特征组合的过程

由于树的深度和棵树限制，存在性能瓶颈

### GBDT+LR:



GBDT 建树，进行特征组合和离散化，得到  $k$  个输出，每个样本输入后得到  $k$  维 one-hot 向量，如  $(0,1,0,0,1)$ ，作为 LR 的输入，进行线性加权预测