树的集成：GBDT

集成方法之提升方法

小胖

目录

ONE 梯度提升
源自梯度下降法的启发

TWO GBDT算法细节
模型步骤

THREE 模型的联结
Facebook的经验

梯度提升

一个简单的回归问题

回归问题

假设有数据集{X, y}，以及模型F_0

$\{\mathbf{X}_i, y_i\} \quad F_0(\mathbf{X}_i)$

从错误中学习

计算得到模型F_0的预测误差

$r_i = y_i - F_0(\mathbf{X}_i)$

F和h可以认为是决策树模型

得到新的数据集{X, r}，训练模型h_1

$\{\mathbf{X}_i, r_i\} \quad h_1(\mathbf{X}_i)$

h_1 = 0时，F_1与F_0的效果一致

得到新的模型F_1，预测效果比F_0好

$F_1(\mathbf{X}_i) = F_0(\mathbf{X}_i) + h_1(\mathbf{X}_i)$

依次递进，得到效果更好的模型F_k

$F_k(\mathbf{X}_i) = F_{k-1}(\mathbf{X}_i) + h_k(\mathbf{X}_i)$

梯度提升

数学表达

回归问题

$$L = \frac{1}{n}\sum_i (\hat{y}_i - y_i)^2$$

损失函数

损失函数的梯度

$$\frac{\partial L}{\partial \hat{y}_i}(F_0(\mathbf{X}_i), y_i) = F_0(\mathbf{X}_i) - y_i = -r_i$$

计算得到模型F_0的预测误差

$$r_i = y_i - F_0(\mathbf{X}_i)$$

$$\{\mathbf{X}_i, -\frac{\partial L}{\partial \hat{y}_i}(F_0(\mathbf{X}_i), y_i)\} \quad \{\mathbf{X}_i, -\frac{\partial L}{\partial \hat{y}_i}(F_1(\mathbf{X}_i), y_i)\} \quad \cdots \quad \{\mathbf{X}_i, -\frac{\partial L}{\partial \hat{y}_i}(F_{k-1}(\mathbf{X}_i), y_i)\}$$

$$h_1(\mathbf{X}_i) \quad h_2(\mathbf{X}_i) \quad \cdots \quad h_k(\mathbf{X}_i)$$

$$F_1 = F_0 + h_1 \quad F_2 = F_1 + h_2 \quad \cdots \quad F_k = F_{k-1} + h_k$$

梯度提升

梯度下降的启示

假设损失函数：

$$L = \frac{1}{n}\sum_{i=1}^{n}(y_i - ax_i - b)^2$$

随机选取起点（a0, b0）：

$$\Delta L = L(a_1, b_1) - L(a_0, b_0)$$

$$\Delta L \approx \frac{\partial L(a_0, b_0)}{\partial a}\Delta a + \frac{\partial L(a_0, b_0)}{\partial b}\Delta b$$

如果令：

$$(\Delta a, \Delta b) = -\gamma\left(\frac{\partial L(a_0, b_0)}{\partial a}, \frac{\partial L(a_0, b_0)}{\partial b}\right)$$

$$\Delta L \approx -\gamma\left[\left(\frac{\partial L(a_0, b_0)}{\partial a}\right)^2 + \left(\frac{\partial L(a_0, b_0)}{\partial b}\right)^2\right] \leq 0$$

得到迭代公式：

$$a_{k+1} = a_k - \gamma\frac{\partial L(a_k, b_k)}{\partial a}, b_{k+1} = b_k - \gamma\frac{\partial L(a_k, b_k)}{\partial b}$$

假设损失函数：

$$L = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = L(\{\hat{y}_i\})$$

随机选取起点（{yi}0）：

$$\Delta L = L(\{\hat{y}_i\}_1) - L(\{\hat{y}_i\}_0)$$

$$\Delta L \approx \frac{\partial L(\{\hat{y}_i\}_0)}{\partial \hat{y}_1}\Delta\hat{y}_1 + \ldots + \frac{\partial L(\{\hat{y}_i\}_0)}{\partial \hat{y}_n}\Delta\hat{y}_n$$

同样的证明过程

得到迭代公式：

$$\{\hat{y}_i\}_{k+1} = \{\hat{y}_i\}_k - \gamma\frac{\partial L(\{\hat{y}_i\}_k)}{\partial \hat{y}_i}$$

梯度提升

$$F_{k+1} = F_k + h_{k+1}$$

梯度提升

梯度下降的启示

$$a_{k+1} = a_k - \gamma \frac{\partial L(a_k, b_k)}{\partial a}$$

$$\{\mathbf{X}_i, -\frac{\partial L}{\partial \hat{y}_i}(F_0(\mathbf{X}_i), y_i)\}$$

$$h_1(\mathbf{X}_i)$$

$$F_1 = F_0 + h_1$$

	梯度下降	梯度提升
更新对象	模型参数	模型预测值
更新方法	损失函数对参数的真实梯度值	模型对相应梯度的预测值
理论基础	以损失函数下降速度最快的方向更新模型参数	以模型效果上升速度最快的方向更新模型预测值

目录

ONE 梯度提升

源自梯度下降法的启发

TWO GBDT算法细节

模型步骤

THREE 模型的联结

Facebook的经验

GBDT算法细节

算法步骤

分类问题

$$L = \frac{1}{n}\sum_{i=1}^{n} y_i \ln(1+e^{-\hat{y}_i}) + (1-y_i)\ln(1+e^{\hat{y}_i})$$

1 定义模型的损失函数

$$L = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = L(\{\hat{y}_i\})$$

2 初始化预测值

$$F_0 = \frac{1}{n}\sum_{i} y_i$$

3 计算损失函数梯度
得到新的训练数据和更新模型

训练数据 $\{\mathbf{X}_i, -\frac{\partial L}{\partial \hat{y}_i}(F_0(\mathbf{X}_i), y_i)\}$

更新模型 h_1

4 得到最优的更新力度
得到新的预测模型

更新力度 $\gamma_1 = argmin_{\gamma} L(F_0 + \gamma h_1)$

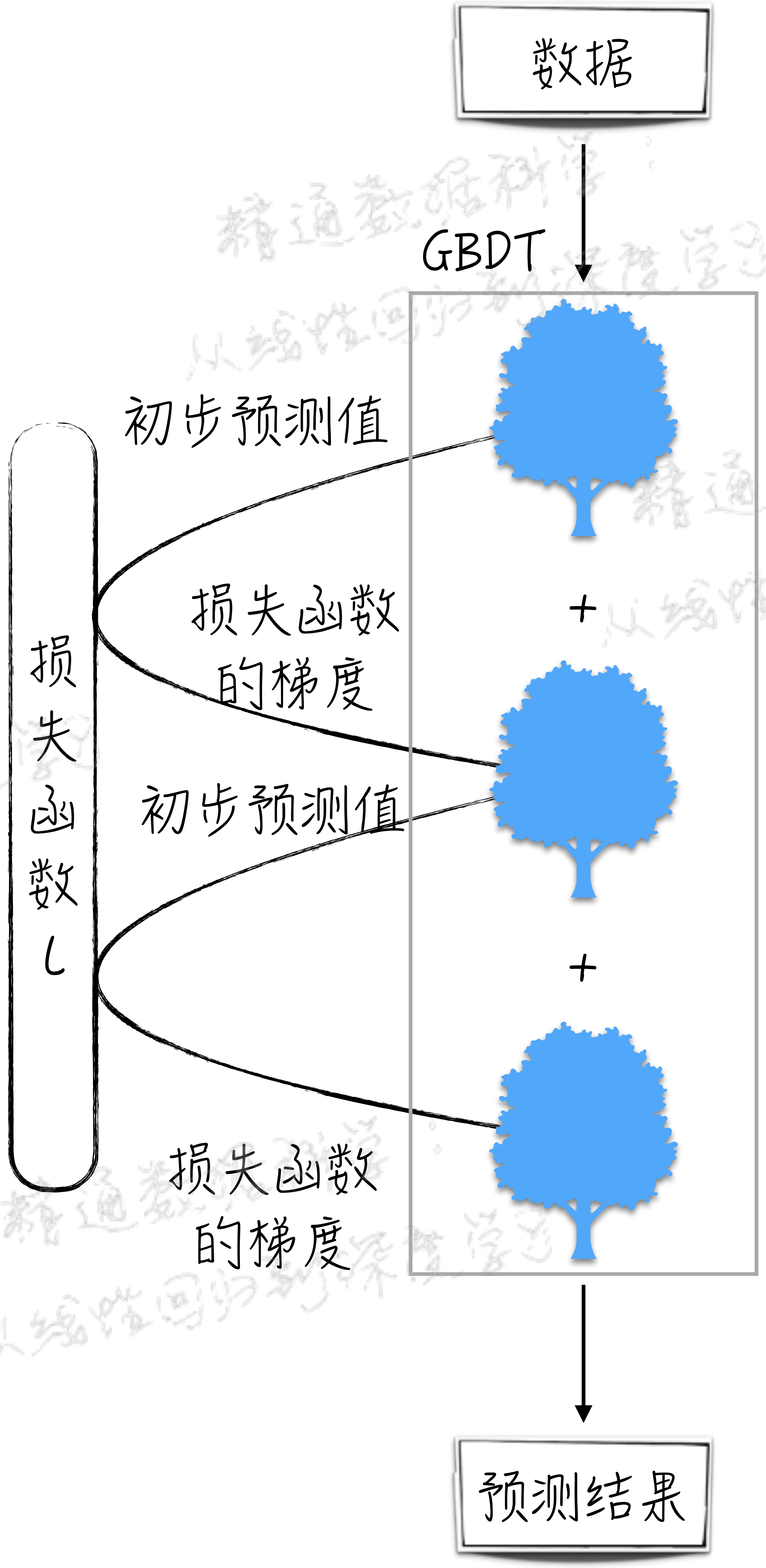
新的预测模型 $F_1 = F_0 + \gamma_1 h_1$

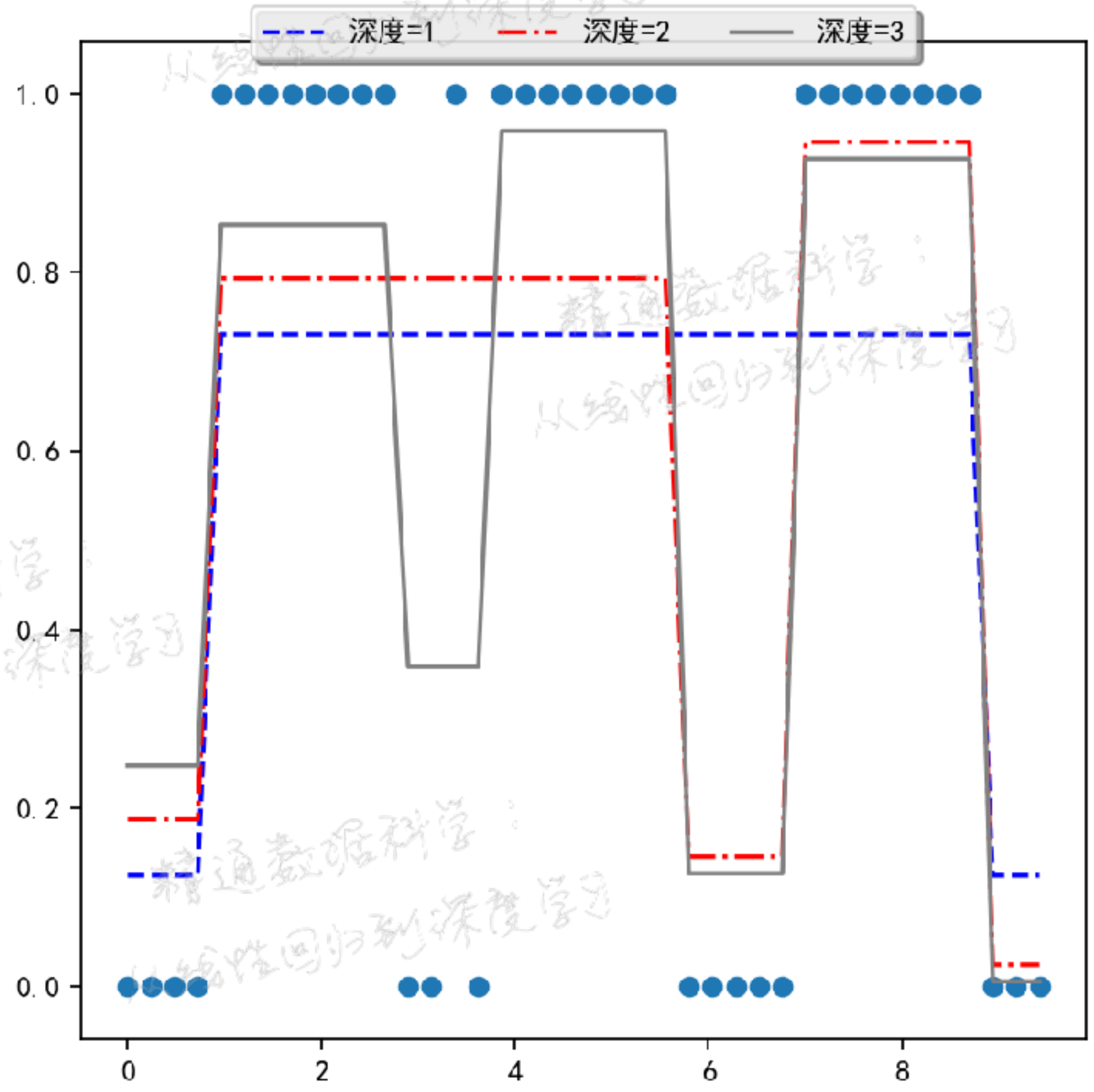
5 不断重复第3步和第4步

$$F_k = F_{k-1} + \gamma_k h_k$$

GBDT算法细节

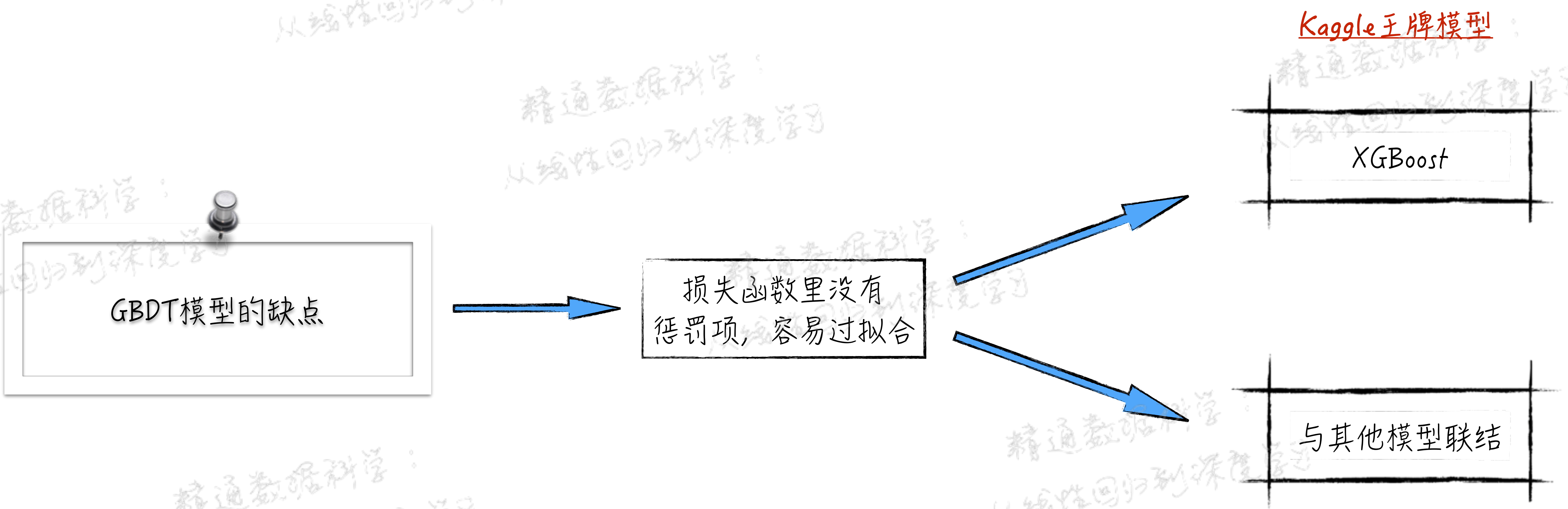
模型结果





GBDT算法细节

模型缺点



目录

ONE 梯度提升

源自梯度下降法的启发

TWO GBDT算法细节

模型步骤

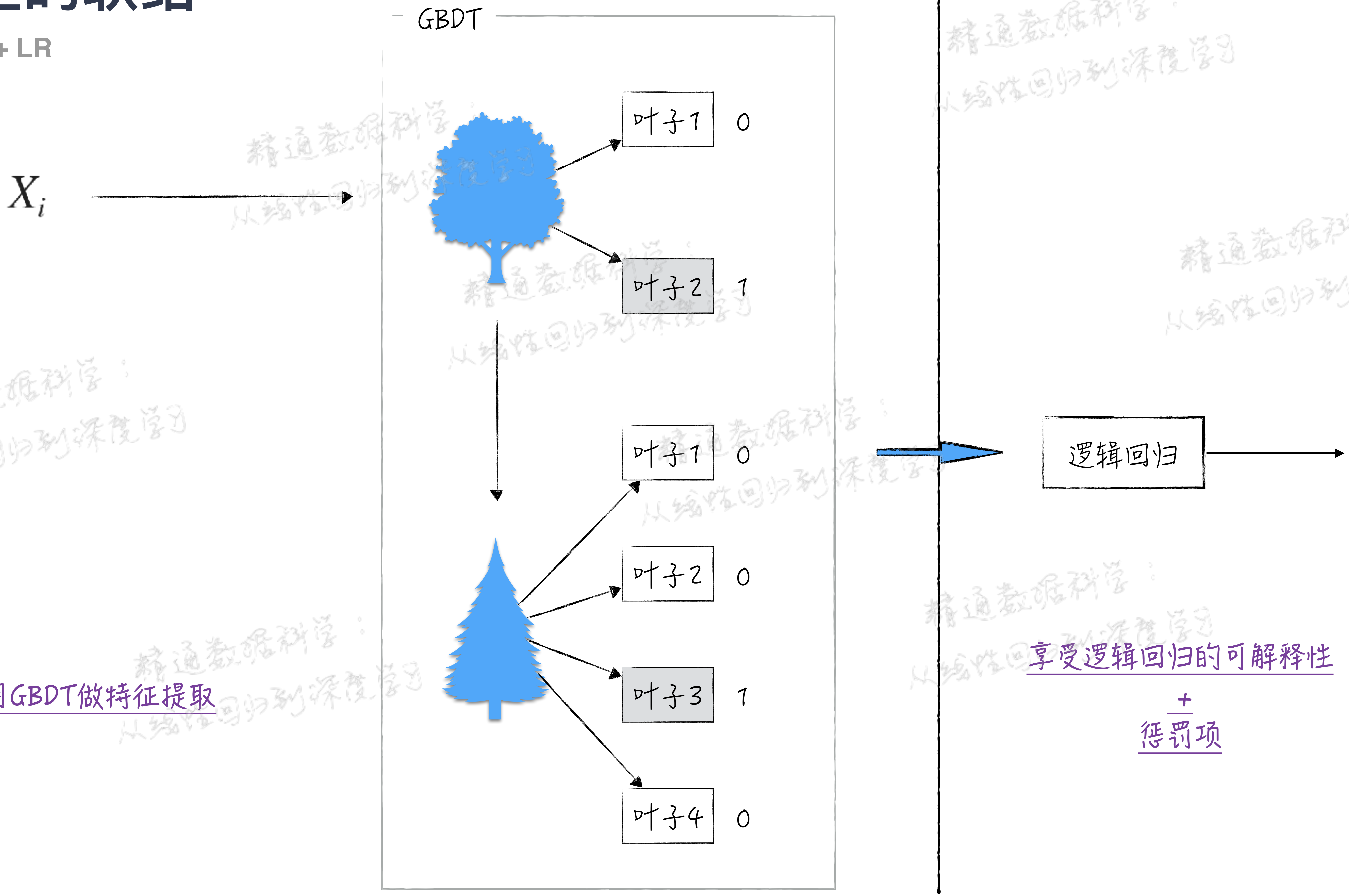
THREE 模型的联结

Facebook的经验

模型的联结

DBDT + LR

广泛应用于风控和广告预测（CTR）



利用GBDT做特征提取

享受逻辑回归的可解释性

+

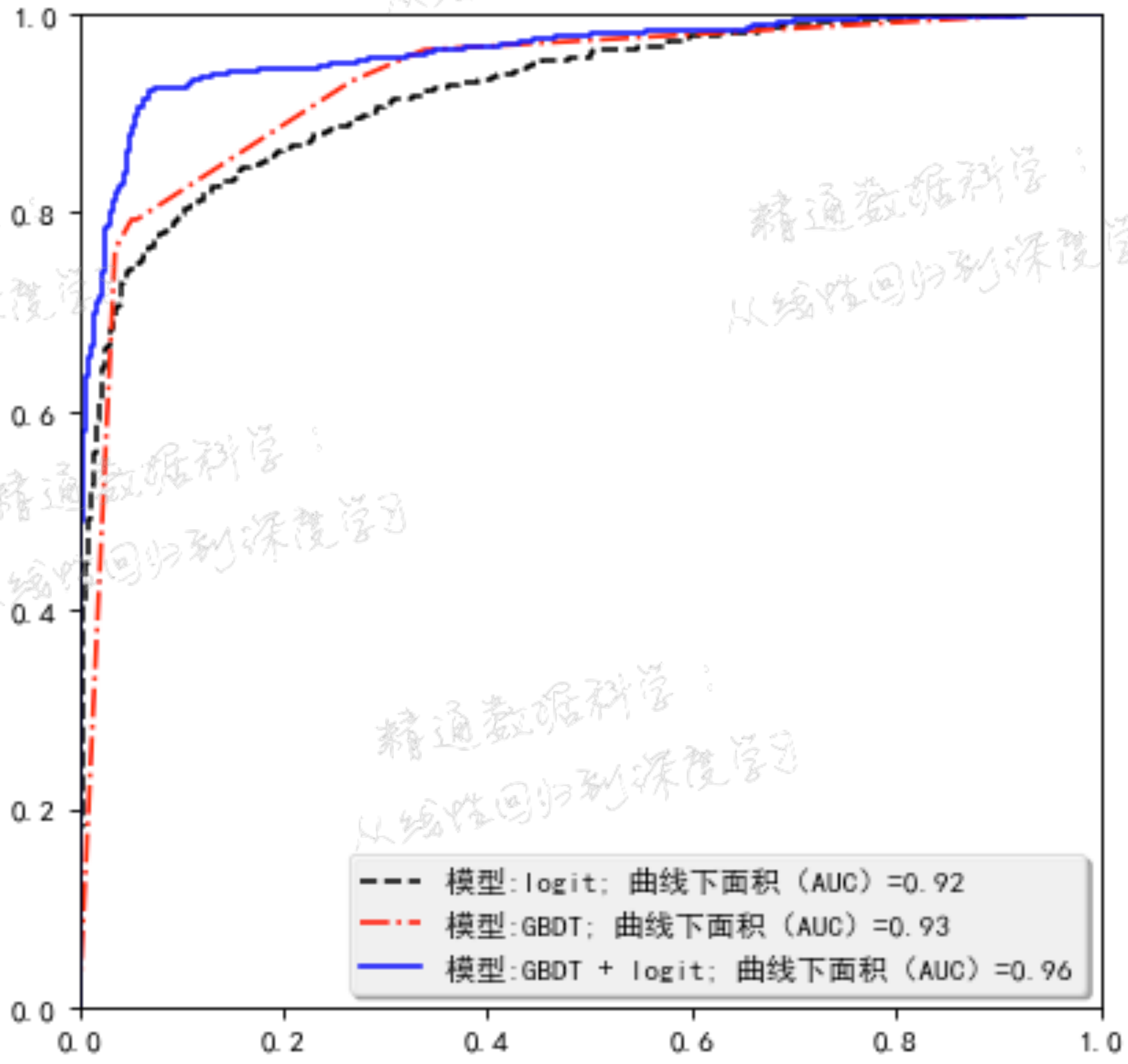
惩罚项

模型的联结

模型效果

同样的数据，分别使用LR、GBDT、GBDT + LR进行建模

GBDT + LR模型的效果更好



THANK YOU