



CTR预估

CSDN学院



► 大纲

- CTR出现的背景
- Logistic回归（LR）
- GBDT
- 因子分解机（FM/FFM）
- 案例分析

► 展示广告

例：百度搜索广告

百度为您找到相关结果约46,600,000个

为您推荐: 儿童自行车 自行车品牌 自行车简笔画 自行车图片 山地自行车

自行车_百度图片

自行车: 简笔画 儿童 结构 搞笑图片 唯美 卡通图片 停车场 制作 轮胎 更多

image.baidu.com - 查看全部1,272,387张图片

自行车_百度百科

自行车,又称脚踏车或单车,通常是二轮的小型陆上车辆。人骑上车后,以脚踩踏板为动力,是绿色环保的交通工具。英...
[名字由来](#) [发展历史](#) [构成](#) [分类](#) [选购技巧](#) [自行车运动](#) [更多>>](#)
[查看“自行车”全部4个含义>>](#)

baike.baidu.com/

相关交通工具

- 空气自行车
比普通自行车轻65%
- 巨无霸自行车
重1吨多的特大自行车
- 凤凰自行车
著名品牌的自行车
- 宝马自行车
比轿车还贵的豪华单车

展开 ▾

相关品牌

- 兰博基尼
举世难得的艺术品
- 迪卡侬
体育用品零售商
- 奥迪
大众旗下豪华汽车品牌
- 永久
上海自行车品牌

展开 ▾

关于自行车, 网友正在热搜

广告

► 展示广告

例：Google搜索广告

广告

与“mountain bik...”匹配的 Google 购物搜索结果

2018 TREK MARLIN 4... NZ\$494.99 Evo Cycles NZ 免费送货	26 / 27.5 inch carbon fiber... NZ\$1,103.42 AliExpress.com 免费送货	2017 GT Aggressor Comp NZ\$489.00 Evo Cycles NZ 免费送货	Kalosse 26*4.0 inch snow bike... NZ\$614.72 AliExpress.com 免费送货	2018 TREK MARLIN 5 RED NZ\$584.99 Evo Cycles NZ 免费送货

NZ Dirt Jump Nationals - Freestyle Bmx & Mountain Bikes
www.narrowspark.co.nz/Dirt-Jump/Nationals ▾ 07-843 6862
 Competitors And Spectators Welcome To Our Open Event At Narrows Park
 Highlights: Serving Since 1953, Non Profit Organisation
[Upcoming Events](#) [Contact Us](#)
[Book Now](#)

Mountain bikes | Trek Bikes
https://www.trekbikes.com/us/en_US/bikes/mountain-bikes/c/B300/ ▾ 翻译此页
 Trek is the world leader in **mountain bike** technology. No surprise that our **mountain bikes** are the most technologically advanced on the market. Here, innovations are not limited to only the highest-end MTB models. Every Trek **mountain bike** is loaded with features and details that will make any ride, on any trail, better.

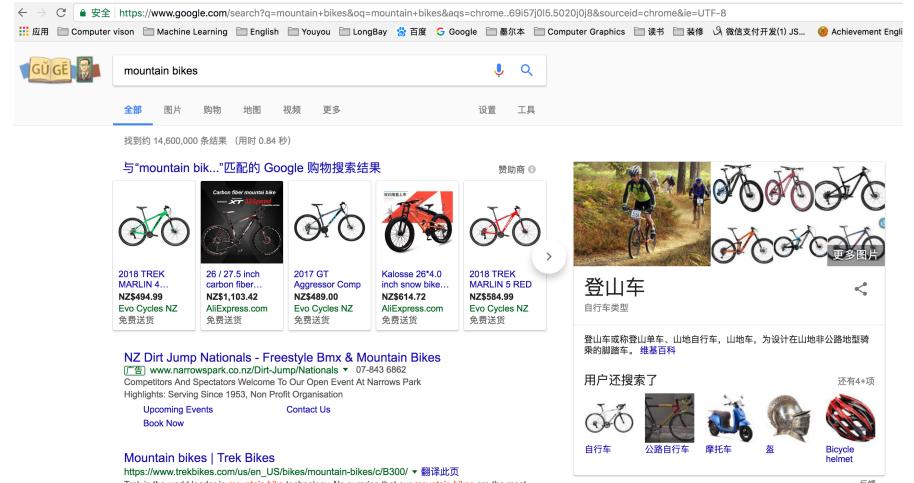
登山车
 自行车类型
 登山车或称登山单车、山地自行车，山地车，为设计在山地非公路地型骑乘的脚踏车。[维基百科](#)

用户还搜索了
 自行车 公路自行车 摩托车 盔 Bicycle helmet

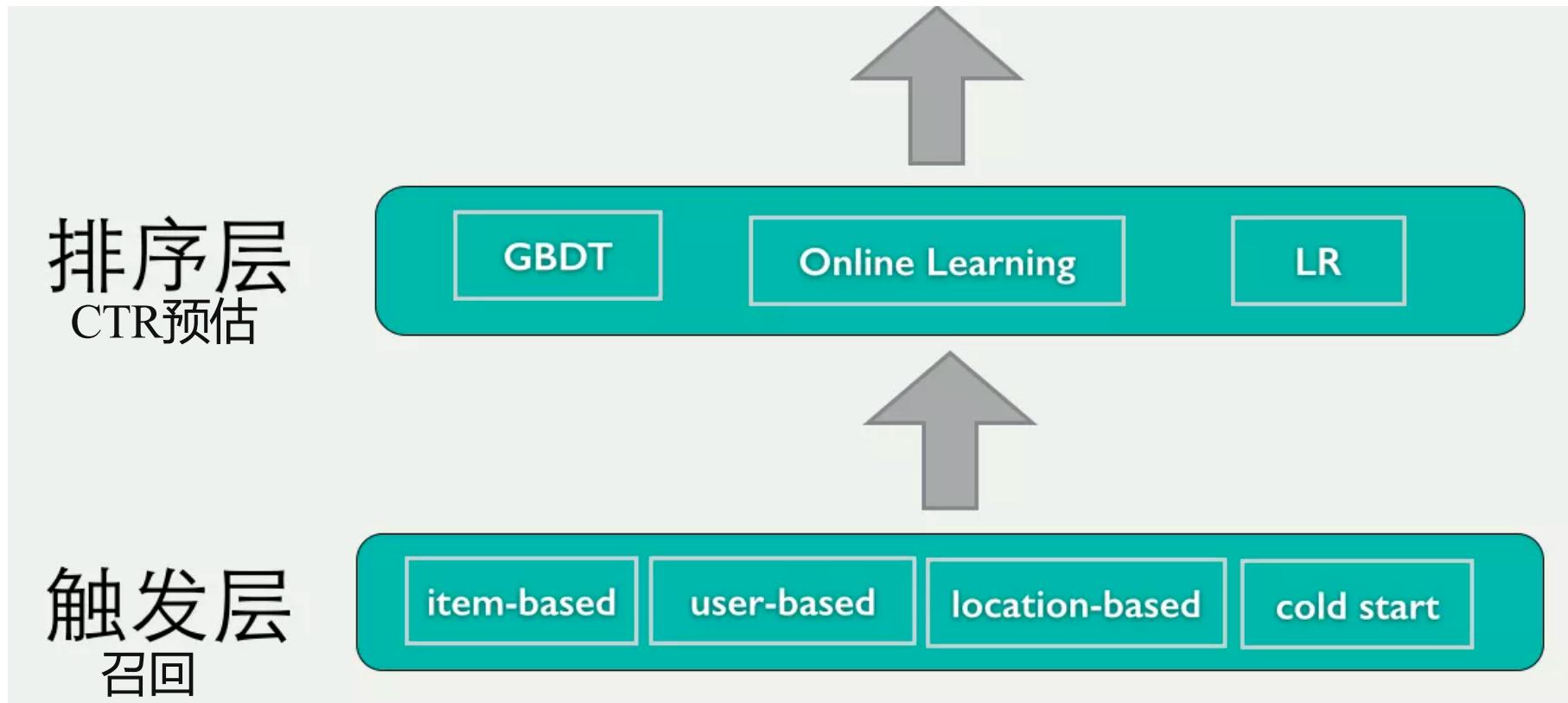


► 展示广告

- 只显示相关的广告
 - 用户喜欢相关的广告
 - 只有用户点击时，
 - 广告平台才能收到广告费
- 所以预测一个广告是否会被点击很关键
 - 预估点击率 (predict Click Through Rate, pCTR)



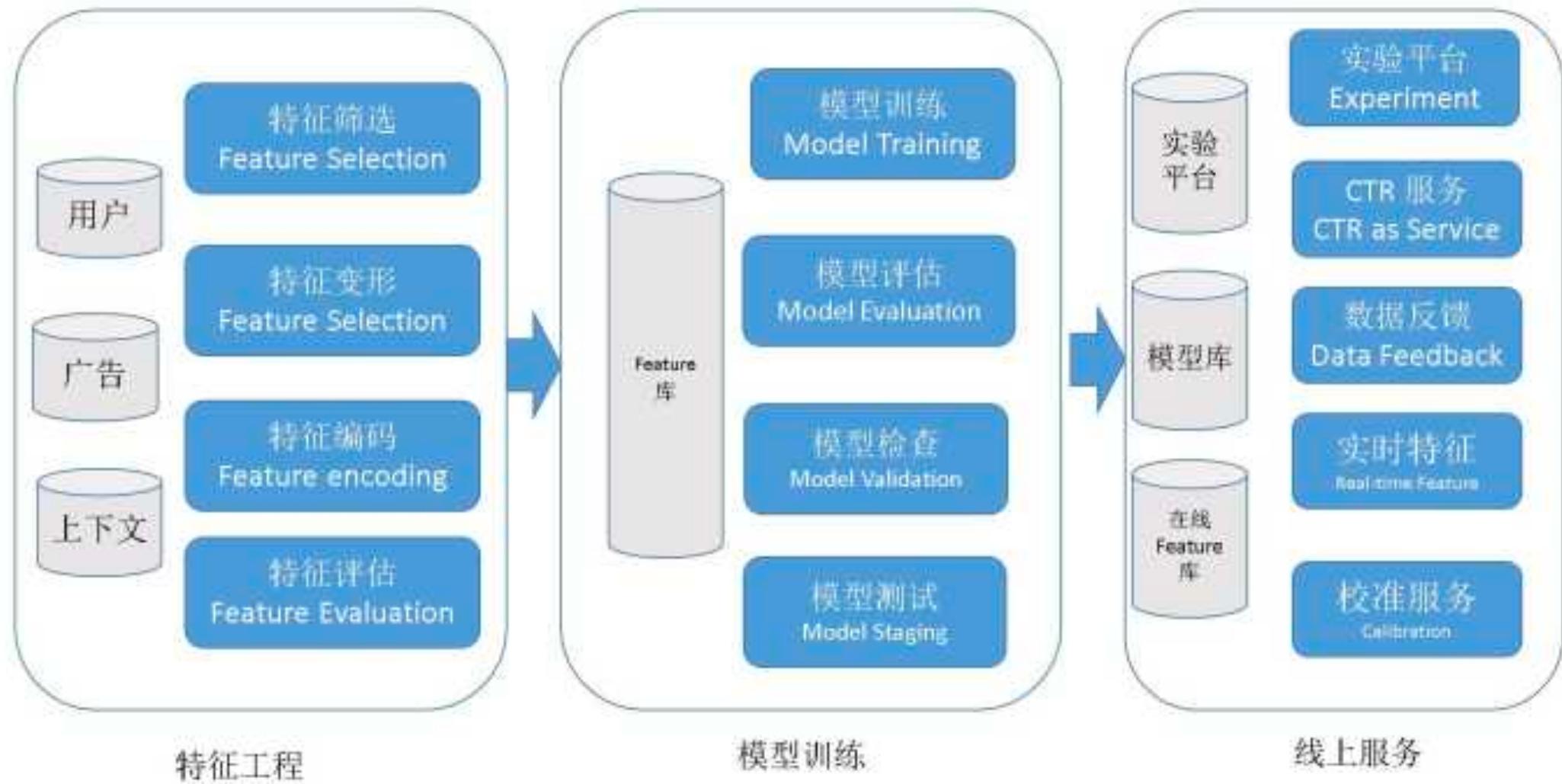
► 推荐系统 vs. 点击率预估



► CTR的挑战

- 用最少的资源在大量的数据上训练大量的模型
 - 上亿的特征数目（模型系数的数目）
 - 每天有上亿的流量（提供上亿次预测服务）
 - 上亿的训练样本数目
- 其他
 - 正负样本数目不均衡
 - 面向稀有事件的 Logistic Regression 模型校准
(<https://vividfree.github.io/%E6%9C%BA%E5%99%A8%E5%AD%A6%E4%B9%A0/2015/12/15/model-calibration-for-logistic-regression-in-rare-events-data>)
 - 特征稀疏（OneHot编码）

CTR 模型训练的三大步骤



特征工程

模型训练

线上服务



► CTR预估

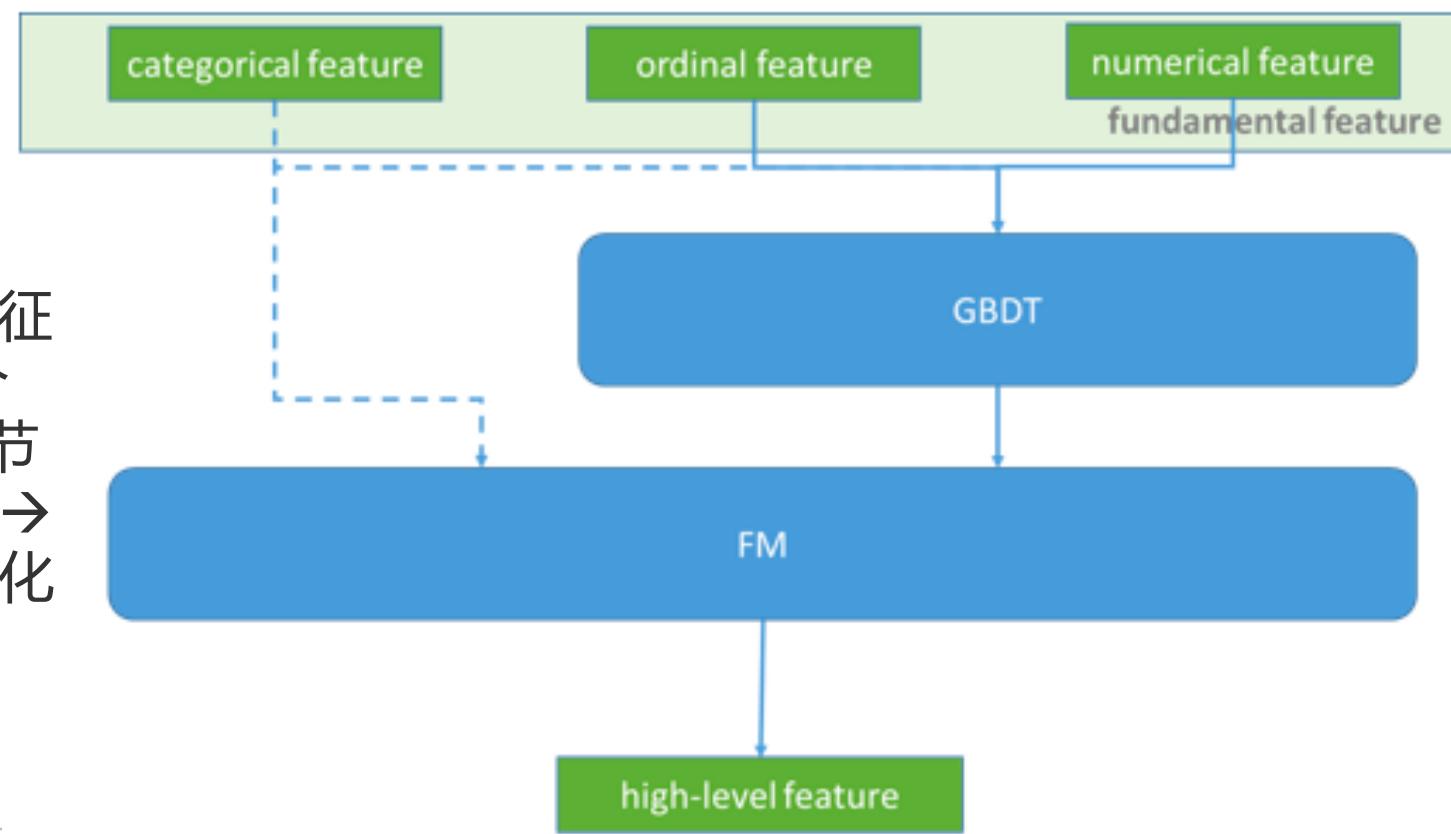
- 预测模型：预估广告是否被点击，是一个两类分类问题
- 特征：用户、广告、场景、媒体
- 性能评估：准确率/召回率、AUC

► CTR —— 基础特征

- 展示广告：在某场景下，通过某媒体向用户展示广告
 - 场景：当时场景，如何时何地，使用何种设备，使用什么浏览器等
 - 广告 - 广告主特征、广告自身的特征如campaign、创意、类型，是否重定向等
 - 媒体 - 媒体（网页、app等）的特征、广告位的特征等
 - 用户 - 用户画像、用户浏览历史、检索关键词等

► CTR —— 特征工程

- 特征离散化
- 特征交叉
 - FM/FFM：类别型特征
OneHot编码后再组合
 - GBDT：用树的叶子节点索引作为样本特征→特征组合并特征离散化
- 特征选择：嵌入到模型



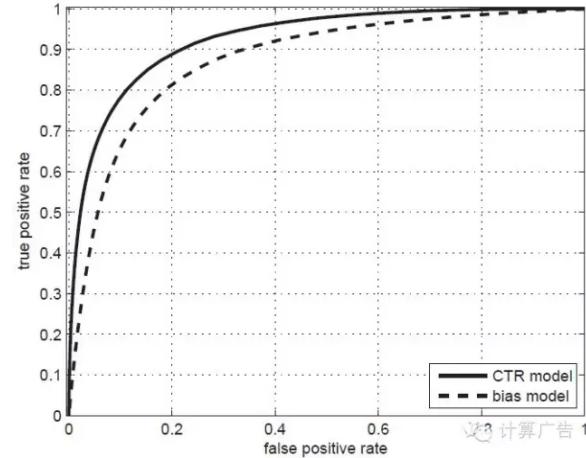
► 预测模型

- Logistic回归 (LR)：
 - 特征工程很重要
 - LR-FTRL : Google在2010年提出了一些理论基础，2013年给出了Paper，并且带有FTRL的实现伪代码。在此之后，FTRL才大规模应用在工业界。
- FM(Factorization Machines)
 - Steffen Rendle于2010年提出Factorization Machines (FM)，并发布开源工具libFM。凭借这单个模型，他在KDD Cup 2012上，取得Track1的第2名和Track2的第3名。本质上可看作是：高效特征交叉 + LR。
- 深度学习DNN
 - 特征工程工作量少，但 10^{11} 级别的原始广告特征、以及特征的稀疏性对DNN还是巨大挑战

► CTR模型评估

- 原则上分类性能评价指标均可用于CTR模型评估
 - Logloss、 PR、 ...
- 最常用的评估指标：ROC曲线下的面积（Area Under Curve， AUC）
 - 随机分类模型AUC为0.5
 - 越接近1模型的效果越好

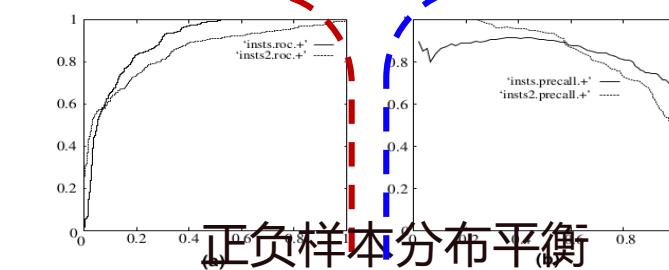
当随机挑选一个正样本以及一个负样本，当前的分类算法根据计算得到的Score值将这个正样本排在负样本前面的概率就是AUC值。 [1]



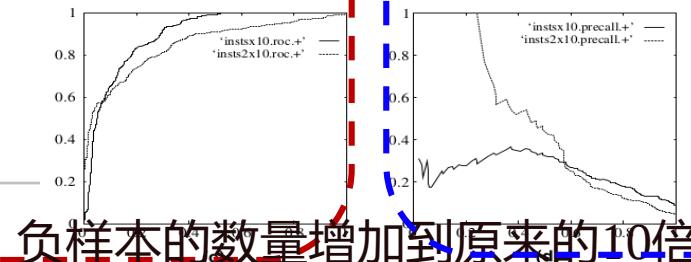
[1] Tom Fawcett , An introduction to ROC analysis, Pattern Recognition Letters, 27, 861-874. 2006

- 当测试集中的正负样本的分布变化的时候，ROC曲线能够保持不变 → AUC不变
- CTR场景下，测试数据中的正负样本的分布可能随着时间变化而变化

ROC曲线
基本保持
不变



PR曲线变
化很大





不止于代码



THANK YOU



AI100

