携程机票航班延误预测算法大赛

背景：随着国内民航的不断发展，航空出行已经成为人们比较普遍的出行方式，但是航班延误却成为旅客们比较头疼的问题。台风，雾霾或飞机故障等因素都有可能导致大面积航班延误的情况。大面积延误给旅客出行带来很多不便，如何在计划起飞前2小时预测航班延误情况，让出行旅客更好的规划出行方式，成为一个重大课题。

要求：提前2小时（航班计划起飞时间前2小时），预测航班是否会延误3小时以上（给出延误3小时以上的概率）

回归问题

Online dataset path:

../input/ctrip001/data/train\_data/airport\_city.csv

../input/ctrip001/data/train\_data/flight\_information.csv

../input/ctrip001/data/train\_data/weather.csv

../input/ctrip001/data/train\_data/special\_news.csv

数据集：

* 航班动态历史数据表

出发机场 到达机场 航班编号 计划起飞时间 计划到达时间 实际起飞时间实际到达时间 飞机编号 航班是否取消

* 城市天气表

城市名 天气 当天最低温度 当天最高温度 日期

* 机场城市对应表

机场编码 城市名称

* 机场特情表

特情机场 收集时间 开始时间 结束时间 特情内容

**Feature**

航班基本信息

1. 航班编号'
2. 计划起飞时间
3. 计划到达时间
4. 飞机编号
5. 飞行时间(flightTime)

航班历史延误信息

1. 平均延误时间
2. 延误时间标准差
3. 最大延误时间
4. 延误时间中位数

飞机编号、日期 ===> 前序航班

1. 前序航班（lastFlight）
2. 前序航班延误时间（timeLastFlightDelay）
3. 若本航班不延误，准备本次飞行的剩余时间（本次飞行预计起飞时间 - 上次飞行实际到达时间 timePrepareThisFlightRemain）
4. 计划准备本次飞行的时间（本次飞行预计起飞时间 - 上次飞行计划到达时间timePrepareThisFlightPlan）

机场 – 城市 – 天气

1. 出发机场最低气温
2. 出发机场最高气温
3. 出发机场天气（weatherVecFrom）
4. 到达机场最低气温
5. 到达机场最高气温
6. 到达机场天气（weatherVecTo）

机场特情

1. 出发机场或到达机场是否有特情

**Label**

1. 计划起飞时间、实际起飞时间 ===> 航班延误时间，是否延误三小时以上

**抽样** 随机抽取正样本，数量为负样本数量的2倍