**赛题背景：**盐城是江苏省最大的乘用车制造基地，汽车产业是盐城市第一支柱产业，已形成了以整车制造为主，集汽车零部件及配件制造、汽车运输、汽车销售、售后服务等行业为一体的较为完整的产业体系。盐城正积极打造集汽车制造、汽车科技、汽车商贸、汽车文化、汽车旅游为一体的中国现代汽车名城。同时汽车产业的发展，城市汽车保有量的增长对城市的建设、道路规划提出挑战。预测盐城总体汽车增长情况，对于政府的道路规划起到一定的指导意义。  
  
  
**赛题描述：**本次赛题需要选手利用历史某3年的汽车日上牌数据，预测某2年每天的汽车上牌数。初赛将挑选出5个汽车品牌，给出这些品牌每天的上牌数，当天是星期几，来预测5个汽车品牌未来每天的上牌总数。复赛将挑选出10个汽车品牌，仍旧给出品牌每天的上牌数，当天是星期几，来预测10个汽车品牌未来每天的上牌数。  
  
  
**数据说明：**   
  
1、数据分成训练数据（train.txt）和测试数据（test.txt）。其中，

|  |  |  |
| --- | --- | --- |
| date | int | 日期，经过脱敏，用数字来表示 |
| day\_of\_week | int | 表示星期几 |
| brand | int | 汽车品牌 |
| cnt | int | 上牌数 |

前3个字段是特征变量，"cnt"是目标变量。  
  
2、数据经过严格脱敏，所以选手看到的”cnt”并非真值；字段”date”, “brand”用数字代替；字段”day\_of\_week”是真实的数据。  
3、排名结果依据预测结果的MSE（mean square error）。

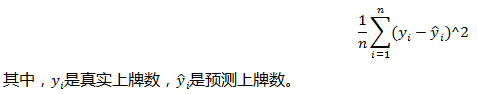
4、补充说明（12月26日更新）：初赛和复赛A榜阶段结束后，将补充A榜答案，请选手注意更新下载！

**初赛提交结果：**  
  
选手需要提交测试数据的预测结果，共2列：

|  |  |  |
| --- | --- | --- |
| date | int | 日期 |
| cnt | int | 上牌总数 |

参考"sample.txt"。  
  
  
**复赛提交结果：**   
  
选手需要提交测试数据的预测结果，共2列：

|  |  |  |
| --- | --- | --- |
| date | int | 日期 |
| brand | int | 品牌 |
| cnt | int | 上牌总数 |

参考“sample.txt”。  
  
  
**结果评估标准：**  
  
1）选手代码跑出的结果与提交结果一致。  
2）预测结果以mean square error作为评判标准，具体公式如下：  
  
  
  
 **注意事项：**  
  
1、 本次比赛允许使用外部数据，及使用开源的已有算法。  
2、 如果抽查发现参赛队伍有造假和作弊行为，将取消该队伍的参赛资格。