**Datawhale 零基础入门数据挖掘-Task1 赛题理解**

**[参赛链接](https://tianchi.aliyun.com/competition/entrance/231784/information)**

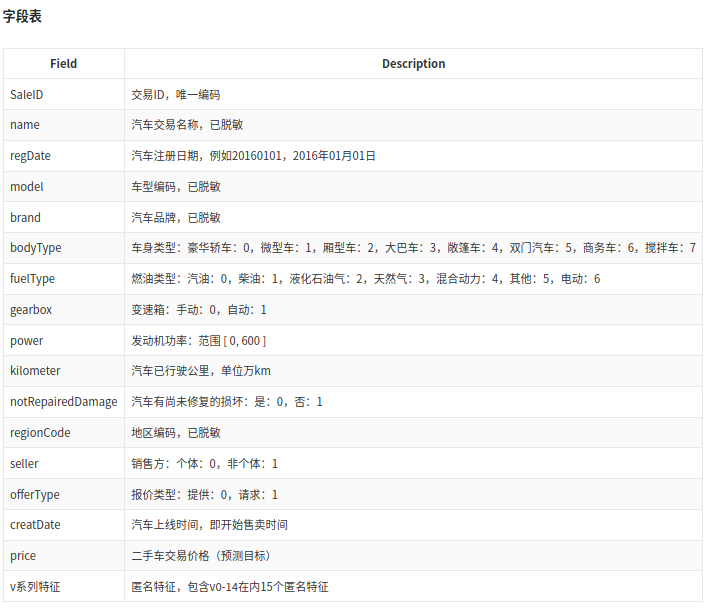
**概述**：这是第一个任务，即熟悉题意，分为以下几个方面：赛题数据

1.赛题数据

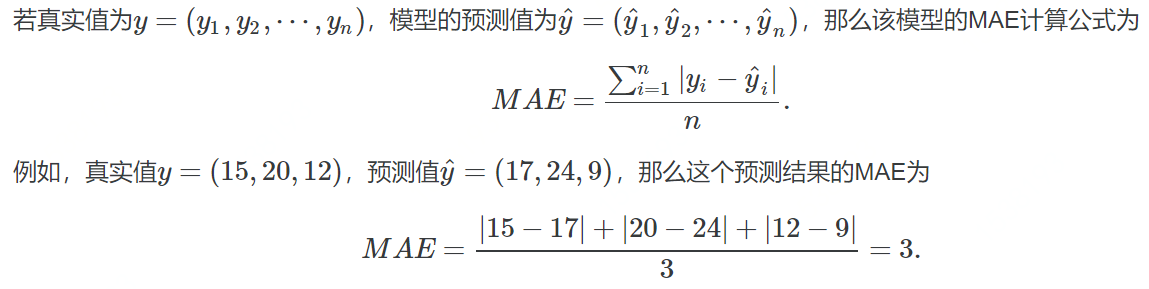
1.1数据描述

赛题以预测二手车的交易价格为任务，数据来自某交易平台的二手车交易记录，总数据量超过40w，包含31列变量信息，其中15列为匿名变量。为了保证比赛的公平性，将会从中抽取15万条作为训练集，5万条作为测试集A，5万条作为测试集B，同时会对name、model、brand和regionCode等信息进行脱敏。

1.2数据字段说明



1. 评价标准
   1. 评价标准为**MAE**(Mean Absolute Error)。



MAE越小，说明模型预测得越准确。

* 1. **总结一下各种模型评估指标**

**分类算法常见的评估指标：**

* 对于二元分类器/分类算法，评价指标主要有accuracy，[Precision，Recall，F-score，Pr曲线]，ROC-AUC曲线
* 对于多元分类器/分类算法，评价指标主要有accuracy， [宏平均和微平均，F-score]

**回归预测类常见的评估指标：**

平均绝对误差 (Mean Absolute Error，MAE)

均方误差 (Mean Squared Error，MSE)

平均绝对百分误差 (Mean Absolute Percentage Error，MAPE)

均方根误差 (Root Mean Squared Error)

R2 (R-Square)

* 1. 代码实现

[代码实现](https://github.com/kevinlee17412/usedcar/blob/master/code/assessMethods.ipynb)

1. 总结

原来模型的评估还有这么多方法，可能比赛多了，就能根据题意想到很多的关于特征提取，构建模型，甚至模型怎么优化的问题，但是我现在一片空白，接下来就是完整的跟完这个项目，熟悉流程，积累经验。

本次任务参考了该[链接](https://blog.csdn.net/DataWhale_0330/article/details/105011254?depth_1-utm_source=distribute.pc_relevant.none-task&utm_source=distribute.pc_relevant.none-task)，在此表示感谢！！！