

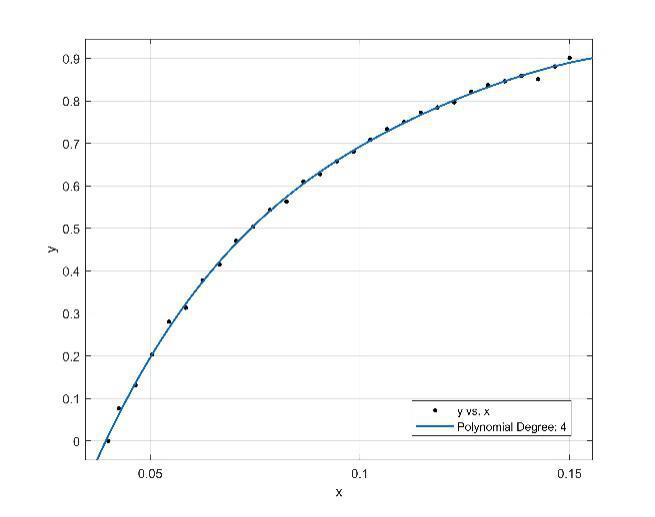
建模的同学提出我们需要提取数据的以下的特征来对数据进行分析，每家公司每个月的销项金额、销项价税合计，进项金额、进项价税合计。企业利润，企业现金流量，企业增值税，企业利润率，过往违约记录，作废发票比例，负数发票比例，行业前景指数。

企业的现金流量等于销项价税合计 与 进项价税合计之和。利润等于销项金额与进项金额之差，因此秩序统计出每家公司每个月的指标之和，再对每月的数据进行二次处理即可得到我们想要的评价指标。

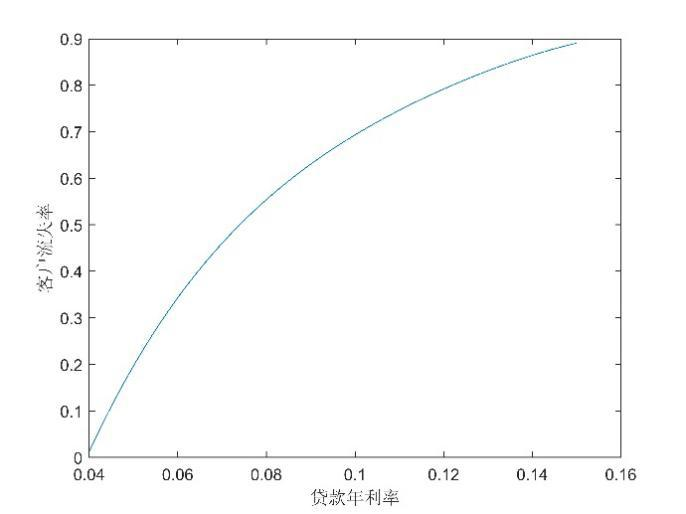
在进行月平均利润的计算时，

对附件一中的公司的进项和出项的情况进行单独的统计，得到表“进出项统计”，利用SPSS的聚类分析（使用卡方测量计数）以后，发现聚类的效果十分符合我们的预期，类别分为两大类，分别是利润率非常高的企业以及利润率处于正常水平的企业，大致以利润率3为分界线，因此我将利润率等于3作为两种企业的分界依据，两种企业需要以不同的指标加以评价。利用这一指标对数据集进行了拆分

利用excel作出各信誉评级下贷款年利率与客户流失率的关系图，发现ABC 3个评级的客户之间趋势变化相近，因此将三者均值代表贷款年利率与企业客户流失率关系。利用MATLAB的拟合工具箱对该函数进行拟合，得到4次函数下最优的拟合效果。



四次函数拟合图。

四次函数图像。

信贷风险指数转换为极小型指标,令max = 1，max-x作为函数的值去计算上述的四次函数，得到银行应该给出的贷款利率rate。

这样的定价机制可理解为信誉评级越高的企业,银行对其贷款风险越小,也即可以采用更低的贷款利率以挽留该企业用户。

Solve.py中利用python的sympy库对方程进行了求解，同时注意给出的贷款利率范围应该在0.04-0.15之间，对于超出范围的解需要将其变为其最为接近的边界值0.04或0.15。