任务：判断、（预测）市场的牛、熊、震荡

目的：

判断当前月份的牛熊，牛市时加仓可以赚钱，熊市时减仓可以规避风险，震荡市时仓位的变化不会对净值造成较大的影响。

思路：

首先是给过去某个月份加（牛、熊、震荡）的标签，这时候可以利用这个月份之前、之后月份的信息综合判断，判断的结果会更加准确。

对当前最新月份做判断时，因为无法得到未来月份的信息，所以最优的方法是利用机器学习方法，根据当前月份以及之前月份的信息，判断当前月份的状态。

时间跨度： 2006.01.01-2016.06.30 10.5年时间跨度.

标的：沪深300，000300.SH

特征：

成交额(amt)，换手率(turn)，涨跌幅(pct\_chg)，ttm PE，

PB，ROE，营业总收入（同比增长率）(yor\_tr)，营业收入（同比增长率）(yor\_or),净利润(同比增长率)( yoyprofit)，

票据转帖利率(日数据，位置：宏观-经济数据库-中国宏观-汇率利率-票据利率-票据转贴利率月息-6个月)，人民币名义有效汇率指数（位置：同票据 -汇率-人民币汇率-人民币名义有效 汇率指数-人民币有效汇率指数（月））

目前只有前四个数据可以提取出来，中间五个数据无法从wind得到。最后两个指标需要处理一下，之后会添加进去。

模型：Logistic Regression，Support Vector Machine。

预测单位：

月度：不太理想，但是方便入手。数据量太少，无法训练分类器；时间跨度较大，股市类别不易判断；月初、月末与相邻月度密切相关，按照每月1号划分月度不一定合理；月度各种指标粒度较大，不太准确。

日：目前看，相对于月，比较理想。

可以先从粗粒度的月度入手，逐步细化到日。

牛熊定义：

牛市：统计月度+前后一个月，共计3个月，月化增幅超过3%。

熊市：统计月度+前后一个月，共计3个月，月化跌幅超过3%。

震荡市：除牛熊市之外的其他情况,作为震荡市。

总共统计了126个月。

振幅阈值设为4%时，判定牛市：27，熊市：15，震荡：84

振幅阈值设为3%时，判定牛市：35，熊市：20，震荡：71

3%的月度增幅，半年化后收益大约为19.4%，年化后收益为42.6%。

工作：

提取数据：利用python接口，先记录到excel方便查看，

加标签：给月度加（牛、熊、震荡）的标签，

特征处理：给特征做一些正则化、二值化的处理。

训练模型：利用scikit-learn的包，训练模型。