

# 价格预测一 实验报告

## 1. 算法思想

**1.1 问题描述** 通过一段时间内的期货交易数据，预测未来一段时间内期货价格的涨/跌情况。

**1.2 问题分析** 预测价格上涨、保持平稳、下跌是一个分类问题，可以提取特征训练线性分类器，比如SVM进行价格预测。

## 2. 实现思路

### 2.1 数据处理

1. 每20s的数据作为一个训练样本，前10s作为输入，后10s进行标注。
2. `read_data.py` 将数据读入，筛选出交易数据以字典方式存入文件
3. `label_data.py` 标注数据，每组输入为 $[a_1, a_2, \dots, a_8]$ ，其中 $a_1, a_2, \dots, a_5$ 为每2s内价格按照成交量加权的均值减去这10s内的均价， $a_6$ 为成交量， $a_7, a_8$ 为10s内的最高价、最低价减去均价。即 $a_i = \frac{\sum_{2s} price_t * volume_t}{\sum_{2s} volume_t} - mean, (i = 1, 2, 3, 4, 5)$   $a_6 = \sum volume$   
 $a_7 = maxprice - mean$   $a_8 = minprice - mean$  通过这样的特征提取，一方面达到降维的目的，一方面平滑了数据，降低了段时间剧烈变化的价格对趋势预测的干扰

### 2.2 训练模型

1. 使用 `sklearn` 的 `linear_model.SGDClassifier` 模型，此方法构造的分类器实现了SVM模型的训练和预测算法
2. 由于使用原始数据进行训练，绝大部分数据均为“保持平稳”的类别，无法训练出有实际价值的可以给出“上涨”或者“下跌”判断的模型，故进行训练数据的平衡，随机从训练集中丢弃标记为“保持平稳”的数据，使得数据中上涨、保持平稳、下跌三类的数据量较为均衡。

## 3. 结果分析

我们将样本集划分为训练集(约占10%)和测试集(约占90%)，具体划分方式设定自`config.py`中的`train_sample`和`test_sample`。(user可以通过更改`train_sample`和`test_sample`来观察不同样本集/训练集下的正确率)

通过测试，我们得到模型对测试集的预测成果。其中，由实际投资盈利的途径，我们对预测正确率定义如下：

1. 预测正确的情形：预测下跌，实际下跌(可通过做空盈利)；或预测上涨，实际上涨(可通过做多盈利)。
2. 预测错误的情形：预测下跌，实际上涨(将导致亏损)；或预测上涨，实际下跌(将导致亏损)。
3. 其余情况均不列入正确率的计算范围。

因此，正确率 = 情况1 / (情况1 + 情况2)

由`predict.py`的输出，我们得出四个模型的预测正确率分别为：

Model A1 accuracy : 0.6163890139576768

Model A3 accuracy : 0.7141213689962611

Model B2 accuracy : 0.9657342037633296

Model B3 accuracy : 0.6575596202824728

可见，线性分类机能够在价格即将出现波动时，作出超过50%正确率的预测，能起到一定的效果。其中，对B2单品而言，分类机效果非常好。

但是，从另一面看，这里的正确率计算是排除了价格不变时的情况(我们认为这些情况不会影响投资者收益)，也就是说，我们只检测了市场上涨或者下跌时，模型作出的上涨或下跌预测是否正确，而模型对市场是否会保持平稳的判断不会被考察到。若算上这些情况，线性分类机准确预测价格波动趋势的正确率变动至：

Model A1 accuracy : 0.6664058875435517

Model A3 accuracy : 0.24415194613870111

Model B2 accuracy : 0.32040398803958414

Model B3 accuracy : 0.1893834120776591

可以看出，A1单品的模型仍然能保持较高的预测成功率，但模型A3, B2, B3的准确率甚至不及三分之一。几乎可以说，除了A1之外，其他三个单品的模型完全不能预测市场是否会保持稳定。

综合以上结果，我们发现，除了A1之外，实现的分类器能够对波动市场的涨或跌作比较好的二元的判断，但并不能判断市场是否降保持稳定。而对于A1模型而言，这个分类器已经能进行准确率超过60%的市场预测，对市场涨跌，是否稳定都有较好的效果。

## 4. 分工情况

---

钮泽平：数据的读取、数据预处理和标注

周沁泓：分类器的训练和预测、数据库和模型库的建立

其中，本次作业思路全部由两人共同讨论完成，以上为实现部分的分工。