

# SVM & HS300

## k\_svm.py

1. 利用沪深300的日行情数据：开高低收、交易量，构建五个特征来刻画K线的形态：

$$high\_low = \frac{high}{low} - 1$$

$$high\_close = \frac{high}{close} - 1$$

$$close\_low = \frac{close}{low} - 1$$

$$close\_open = \frac{close}{open} - 1$$

$$vol\_pct = \frac{vol_i}{vol_{i-1}} - 1$$

以前3天的特征数据，利用基于高斯核的支持向量机分类器，预测后3天的市场涨跌方向；

2. 以10年的数据为训练集，采用网格搜索和交叉验证优化参数，其中  $K - Fold$  的  $K$  设为10，下一年的数据为测试集，依次往后滚动每隔1年构建一个新的分类器；
3. 根据预测进行多空交易，收益率为预测期间第三天的收盘价减去第一天的开盘价，并扣除一定的滑点比例  
单方向的滑点为0.0002

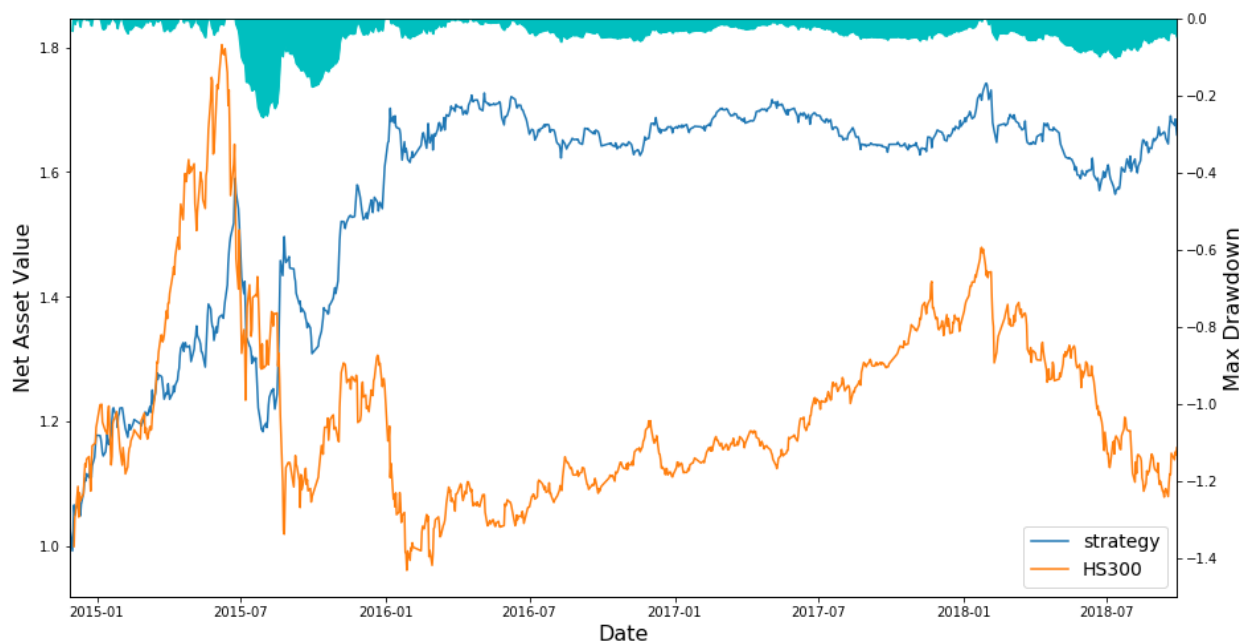
如果预测正确：

$$profit = |Close_3 - Open_1| - slippage$$

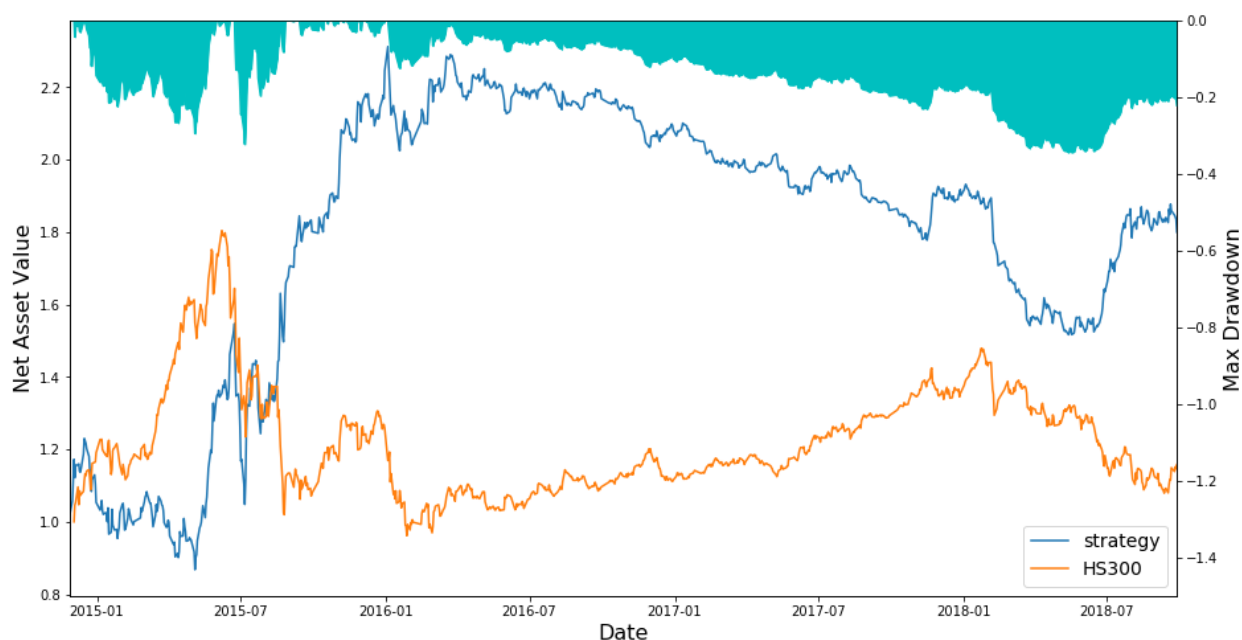
如果预测失误：

$$loss = -|Close_3 - Open_1| - slippage$$

#### 4. 收益情况



2015年1月至2018年10月考虑滑点的情况下的总回报，是1.66，最高值1.75左右，最大回撤26%，沪深300的总回报1.11，最大回撤47%。



如果仅仅是预测后1天的涨跌，总回报是1.80，最高值为2.31，最大回撤34%；

5. 可以看到策略主要是在2015年完成收益率的积累，后续表现乏力，说明在趋势阶段预测效果较好，其他阶段则不如人意。

## 6. 预测正确率

在2016年和2017年的预测正确率均低于50%，可能原因是，特征的相关性很高，有效信息更新太慢。

period	C	gamma	train score	average train score (k=10)	test score
2014.11- 2015.11	8	0.02	0.6592	0.5650	0.5333
2015.11-2016.11	6	0.02	0.6329	0.5529	0.4750
2016.11-2017.11	5	0.06	0.7063	0.5517	0.5167
2017.11-2018.10	8	0.02	0.6383	0.5579	0.4566

## 7. reference

[优矿，基于SVM的大盘预测](#)

## v\_svm.py

1. 利用沪深300的日行情数据：开高低收、交易量，构建波动、动量和趋势相关的七个特征：

$$high\_low = \frac{high - open}{open}$$

$$close\_open = \frac{close}{open} - 1$$

$$vol\_pct = \frac{vol_i}{vol_{i-1}} - 1$$

$$pct\_m = \frac{close_i}{open_{i-m}} - 1, m = 5$$

$$high\_v = \frac{high_i}{\max\{high_t, t=i, i-1, \dots, i-v\}} - 1, v = 20$$

$$low\_v = \frac{low_i}{\max\{low_t, t=i, i-1, \dots, i-v\}} - 1, v = 20$$

$$\sigma = \sqrt{\sum_{i=1}^{21} \frac{(r_i - \bar{r})^2}{20}} \times \frac{240}{20}$$

分别表示当天的振幅（最大可能收益），当天实现的收益，交易量的变化幅度，m天的累计收益率，当天最高价与20日最高价的差距，当天最低价与20日最低价的差距，20日的波动率。

## 2. 特征说明

- (1) 振幅、当天收益率，交易量的变化，可以反映当天的市场波动情况；
- (2) m天的累计收益率，反映了当前的市场位置和周线的情况；
- (3) 后三者反映了中期月度频率下的市场位置和波动情况

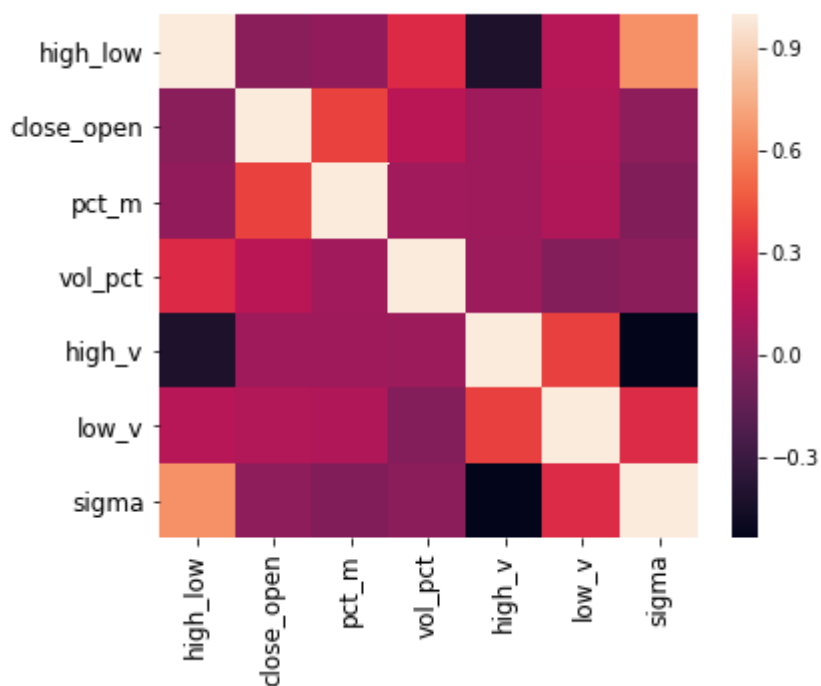
利用这七个特征和基于高斯核的支持向量机预测后一天的市场涨跌，并进行相应的多空交易。

## 3. 特征之间的相关性较小

相关性最强的是 $\sigma$ 和 $high\_low$ ，相关系数为0.65；

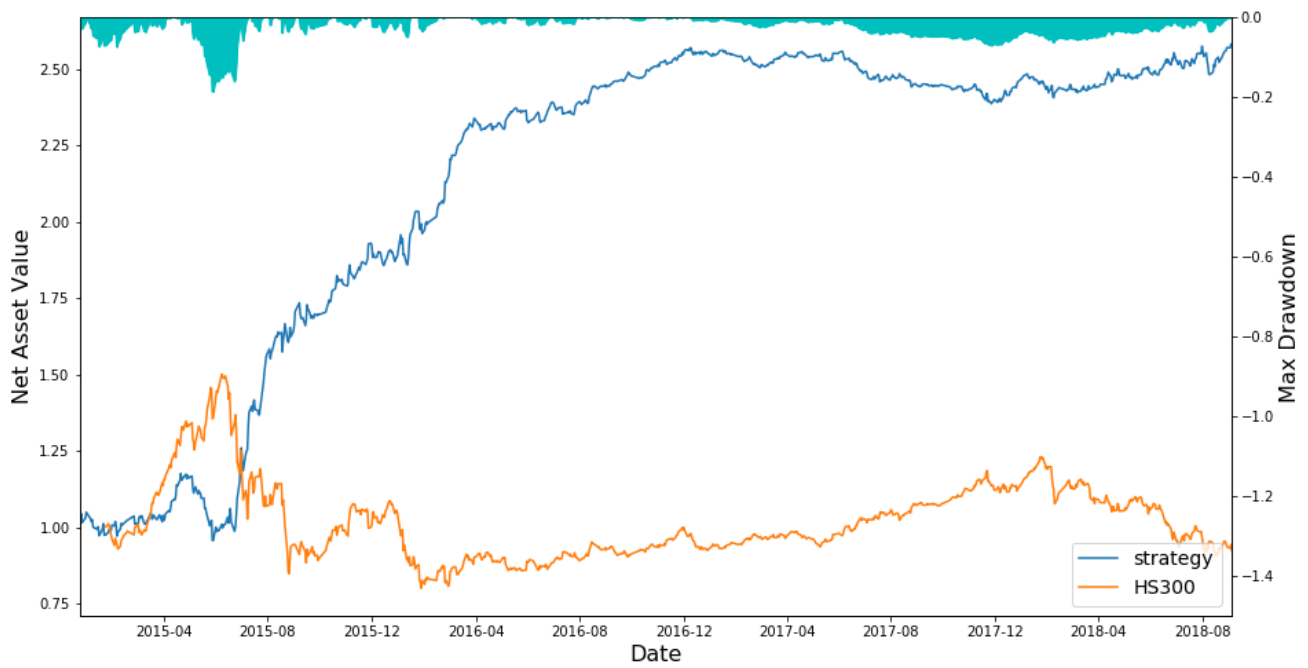
其次是 $\sigma$ 和 $high\_v$ ，相关系数为-0.54；

其余的相关系数的绝对值都小于0.5



## 4. 收益情况

总回报是2.58，最大回撤18.67%，calmar比率1.43，远好于指数的表现。



## 5. 预测正确率

可以看到特征的预测准确率是比较高的，除了2016年外，均达到了60%以上。

period	C	gamma	train score	average train score (k=10)	test score
2014.11- 2015.11	1	0.006	0.6483	0.6413	0.6392
2015.11- 2016.11	2	0.006	0.6521	0.6404	0.6792
2016.11- 2017.11	1	0.012	0.6508	0.6454	0.5083
2017.11- 2018.10	3	0.008	0.6396	0.6267	0.6162