

Машинное обучение в статистическом арбитраже*

Зайцев Александр БЭК201
aazaytsev_4@edu.hse.ru

ВШЭ



О задаче

Статистический арбитраж это реализация алгоритмической торговли, суть которой стоит в получении прибыли относительно изменчивости спреда между активами (портфелями активов) при предпосылках возвращения спреда к конкретному значению или промежутку. Одно из наиболее заметных преимуществ и особенностей данной стратегии является её рыночная нейтральность, другими словами, нивелируется риск рынка, в связи с одновременным открытием длинной и короткой позиции по схожим в волатильности и изменчивости активов (портфелей активов).

Для построения статистического арбитража необходимо выполнить два шага:

- I. Отобрать активы – это может быть осуществлено на основе дополнительных предпосылок, поиска статистических взаимосвязей формирования цен активов, на основе фундамент. показателей компаний.
- II. Реализовать схему осуществления длинной и короткой позиций по выбранным активам, то есть на расширение или сужение спреда.



Эконометрика

I. Выбор пар активов — выбор одной из модели

Коинтеграцион. подход

Основным подходом является проверка коинтеграцию активов, которое заключается в том, что:

1. Каждый отдельный актив должен быть нестационарным во времени
2. Линейная комбинация при ненулевом параметре стационарна во времени

Параметр (коэф. коинтеграции) является показателем сколько необходимо первого актива чтобы полностью реплицировать второй актив.

Если активы скоинтегрированы, то у них имеется устойчивый спред, относительно которого можно получить прибыль

k-means

Не исключает выбросы, а так же располагает их в ближайший кластер из-за чего снижается общая прибыль от стратегии. В связи с чем необходимо модифицировать модель* или искать выбросы вручную. (Выбросы составляют в основном около 60% и более от активов индекса)
Не может искать кластеры сложных форм
Тенденция формирования множества небольших кластеров

Кластеризация

DBSCAN + OPTICS

Формирует кластеры на основе минимального количества информации так как выполняются следующие предпосылки:
1. Строгое решение, которое не включает выбросы
2. Отсутствие предположение о формах
3. Нет необходимости группировать все активы
Даже при большой размерности имеет тенденцию формирования одного большого кластера, в то время как остальные считаются выбросами.

AGNES

Может сформировать кластеры на основе минимального количества информации (см. предпосылки в DBSCAN+OPTICS)
Тенденция формирования множества небольших кластеров
Показывает наилучший результат по показателю Шарпа и доходности и наименьшую просадку по данным NYSE

Краткое описание результатов на данных NYSE с 1980 по 2020 год (3157 активов)

Кол-во кластеров
Кол-во активов в кластере
Кол-во выбросов
Кол-во торгующихся активов

129
1259 (39.8%)
1900* (60.2%)
330 (10.3%)

2
376 (12%)
2781 (87.95)
75 (2.5%)

188
769 (24.6%)
2388 (75.4%)
126 (4%)

Особенности подходов:

Необходимо только иметь исторические цены активов

Возможность использования фундаментальных и иных показателей как дополнительный источник для повышения доходности, коэффициента Шарпа и тд.
Отсутствие необходимости использовать только на историческое изменение цен активов
Необходимость дополнительных предпосылок в формировании пар активов для формирования четких решений

II. Реализация стратегии, модель обучения для торговли на данных S&P 500 с 1989 по 2015 год

Основные моменты:

Целей показатель вероятность акций получить большую доходность чем рынок (индекс).
Алгоритм торговли: По 10 активам с самым высоким вероятностям открываются длинные позиции, по 10 активов самой низкой вероятностью - короткие.
(Увеличение числа активов способствует увеличению неопределенности)

Модели:

RF

В рамках модели использовалось 1000 решающих деревьев при глубине 20

GB

В рамках модели использовалось 100 решающих деревьев с глубиной 3 и скорости обучения 0.1

DNN

Используется следующая архитектура нейронной сети 31-31-10-5-2 (I-H1-H2-H3-O), hidden drop-out 0.5, input dropout 0.1, L1 регулязация, ADADELTA оптимизация

Ensemble's

В рамках модели рассматривалось 3 варианта ансамбля – ср. арифм., ср гарм., и относительно веса каждого предсказания. Все они показатели примерно одинаковый результат.

MKT

(рынок)

Краткое описание результатов

Ежедневная доходность
Показатель Шарпа
Показатель Сортино
(После издержек)

0.23%
1.90
3.84

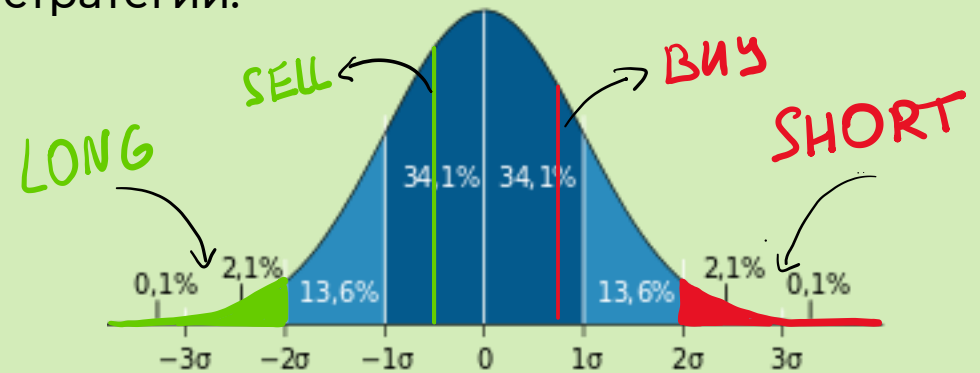
0.17%
1.23
2.04

0.13%
1.23
2.04

0.25%
1,81
3,33

0.04%
0,35
0,71

Стационарная линейная комбинация с ненулевым коэффициентом коинтеграции приводится к стандартному нормальному распределению (Z). Относительно новых точек на полученном стандартном нормальном распределении осуществляется покупка, продажа спреда. Предлагается следующая реализация стратегии:



Помимо этого необходимо иметь стоплосс, закрытие убыточной позиции после определенного количества прошедших дней.