

УДК 336.763

ПРИМЕНЕНИЕ КЛАСТЕРНОГО АНАЛИЗА ДЛЯ ОПТИМИЗАЦИИ ПОРТФЕЛЬНОГО ИНВЕСТИРОВАНИЯ

Коцюбинская С.А.

студент

*Институт сферы обслуживания и предпринимательства (филиал) ДГТУ в
г. Шахты*

Россия, Шахты

Аннотация: Формирование эффективного инвестиционного портфеля требует выявления наиболее привлекательных для инвестора активов. Применение кластерного анализа позволяет рационально подойти к этому процессу, повысить его эффективность. Рассмотренный в статье метод кластеризации k-means направлен непосредственно на установление принадлежности каждой акции к конкретному кластеру и позволяет сделать выводы об эффективности отдельных активов, основываясь на их исторических данных.

Ключевые слова: оптимизация инвестиционного портфеля, метод k-means, Elbow Curve, отбор акций.

THE USE OF CLUSTER ANALYSIS TO PORTFOLIO OPTIMIZATION

Kotsyubinskaya S.A.

student

*Institute of Service and Entrepreneurship (branch) of Don State Technical University,
Shakhty*

Russia, Shakhty

Abstract: Efficient portfolio creation requires identification of the most attractive assets for investor. The use of cluster analysis allows a rational approach to this process

to increase its efficiency. Considered in the article K-means clustering is aimed at establishing the belonging of each stock to a specific cluster. It allows to make rational decisions about the efficiency of individual assets, based on their historical data.

Keywords: portfolio optimization, k-means clustering, Elbow Curve, stock selection.

Эффективность инвестиционного портфеля, его качественная оценка, основывается на нахождении компромисса между доходностью и риском. Однако рассмотрение акций, входящих в его состав со стороны объединения двух основополагающих параметров в один количественный показатель, отражающий сходство или различие группируемых объектов, позволяет отобрать акции наиболее рационально, основываясь на их собственном соотношении доходность-риск и сделать вывод об эффективности каждой отдельной акции.

Подобное разбиение на группы, называемые классами, лежит в основе кластерного анализа, который в свою очередь базируется на принципе максимизации внутриклассового сходства и его межклассовой минимизации.

1. Кластеризация

Кластеризация – это задача разбиения множества объектов на группы, называемые кластерами, не заданными изначально, а определяемыми в процессе работы алгоритма, состоящего из следующих этапов:

- Отбор объектов кластеризации;
- Задание переменных для оценки объектов выборки, их нормализация;
- Вычисление меры сходства между объектами;
- Создание кластеров;
- Представление результатов.

В общем виде задача кластеризации представлена следующим образом: пусть X – множество объектов, подвергающихся кластеризации, тогда Y – множество кластеров, $p(x, x')$ – функция расстояния между объектами. Дневник науки | www.dnevniknauki.ru | СМИ Эл № ФС 77-68405 ISSN 2541-8327

$X^m = \{x_1, \dots, x_m\} \in X$ – конечная обучающая выборка объектов. Требуется разбить выборку на непересекающиеся подмножества, так чтобы каждый кластер состоял из объектов, близких по метрике p , а объекты, входящие в разные кластеры, существенно отличались. При этом каждому объекту $x_i \in X^m$ соответствует номер кластера y_i , а алгоритм кластеризации — это функция $F: X \rightarrow Y$, ставящаяся в соответствие любому объекту $x \in X$ номер кластера $y \in Y$ [].

2. Метод k-средних (k-means)

k-means - один из алгоритмов кластеризации. Являясь нечетким, он ставит в соответствие каждому объекту набор вещественных значений, указывающих на вероятность принадлежности к кластеру. Данный метод разделяет объекты на k число кластеров и включает в себе следующие этапы:

- 1) Начальная стадия: разделение объектов на k кластеров случайным образом
- 2) Этап повторения:
 - Вычислить центр каждого кластера как среднее значение данных
 - Рассчитать расстояние от каждого объекта к каждому кластеру (например, Евклидово расстояние)
 - Рассчитать целевую функцию, известную как квадратичная функция ошибок:

$$SF = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - C_j\|^2$$

где, K – количество центров кластеров, n - число данных в кластере i , $\|x_i^{(j)} - C_j\|^2$ - Евклидово расстояние каждого объекта от его центра C_j .

- 3) Стадия улучшения: присвоить каждый объект кластеру с ближайшим центром
- 4) Стадия остановки: алгоритм функционирует до тех пор, пока ни один объект не переместится в кластер, либо целевая функция не будет уменьшена.

Однако данный алгоритм неприменим без начального задания числа кластеров для последующего анализа, следовательно, ставится задача определения их оптимального количества.

3. Кривая локтя (Elbow Curve)

Используется с целью выявления связи между количеством выбранных кластеров и суммарной квадратической ошибкой, полученной в результате работы заданного числа кластеров, т.е. для получения оптимального числа кластеров необходима своего рода проверка результатов кластеризации. При этом важно учитывать компромисс между ними возникающий: нацеленность на минимизацию ошибки и в то же время на сокращение числа кластеров, проявляющийся в эффекте «локтя».

Таким образом, «Elbow Curve» — это метод интерпретации и оценки соответствия в рамках кластерного анализа, предназначенный для поиска соответствующего числа кластеров в некотором наборе данных.

4. Экспериментальные исследования

Кластеризация, являясь методом группировки объектов в соответствие с некоторой заданной метрикой находит свое отражение в различных областях человеческой жизни, в том числе и в сфере инвестиций. Целесообразность применения кластерного анализа при формировании эффективного инвестиционного портфеля обуславливается потребностью рационального отбора акций, входящих в его состав, т.е. выявления наиболее пригодных для инвестора активов исходя из его склонности к риску и его нацеленности на ожидаемую прибыль.

Таким образом, в качестве объектов кластеризации выступают акции, входящие в индекс Московской Биржи (ИМОЕХ) по состоянию на декабрь 2018 года и с доступными историческими данными за 2 года, а оценка объектов выборки производится на основе двух показателей: ожидаемой доходности и

риске, которые в виду рассмотрения акций в ежегодном интервале подвергаются нормированию (умножение на количество рабочих дней равное 255).

Выявление числа кластеров основывается на методе Elbow Curve, автоматизация которого как и работа с акциями реализовывалась посредством использования объектно-ориентированного языка python и ряда его библиотек: pandas, numpy, scipy, sklearn, matplotlib, направленных на работу с данными, их представление, математические вычисления. Определения количества кластеров осуществлялось за счет их формирования в числе, лежащем в интервале от 2 до 20 и оценке пригодности полученных групп, выраженной суммарной квадратической ошибкой. Графическое представление метода Elbow Curve представлено на рисунке ниже:

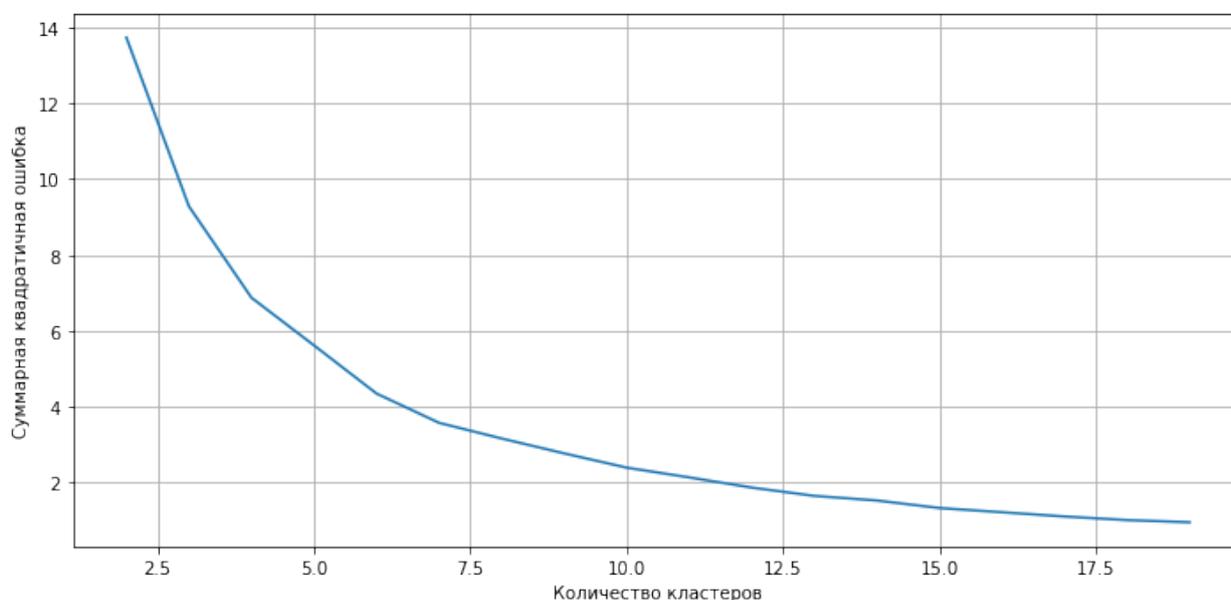


Рис. 1 - Elbow Curve

По мере достижения количеством кластеров значения 5 происходит замедление падения показателей суммарной квадратической ошибки для каждого последующего увеличения числа кластеров, что свидетельствует о достижении компромисса между минимизацией двух параметров: числом кластеров и ошибке их применения, следовательно, оптимально разбиение на пять групп.

Последующий анализ основывается на методе k-means, направлен непосредственно на установление принадлежности каждой акции к конкретному кластеру и, как указывалось выше, реализован посредством языка python. Полученные результаты, а именно группировка акций представлены на рисунке 2.

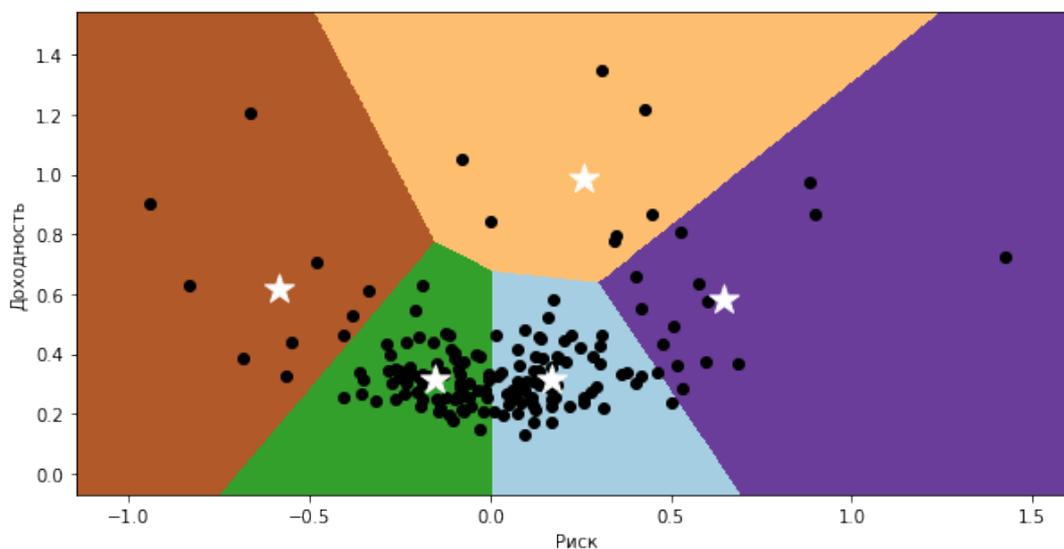


Рис. 2 – Кластерный анализ методом k-means. Центры кластеров обозначены белыми звездами, а сами кластеры отражены посредством цветового разнообразия

Кластерный анализ ориентирован на объединение объектов максимально сходных по показателям, поэтому целесообразно отразить на графике отклонение акций, входящих в состав кластера от значения среднего показателя группы посредством рассмотрения изменения кумулятивной величины их доходности со временем. Что в свою очередь позволяет сделать выводы об эффективности отдельных активов, основываясь на их исторических данных.

Согласно рис. 2 привлекательными для инвестора акциями станут, активы входящие в состав кластеров, обозначенных голубым и фиолетовым цветами, что также отражается на рис.3: кластеры № 1 и № 3 соответственно. Именно им характерен стабильный рост на протяжении всего рассматриваемого периода.

Красной линией обозначен средний временной показатель доходности кластера, указывающий на общность значений сгруппированных акций.

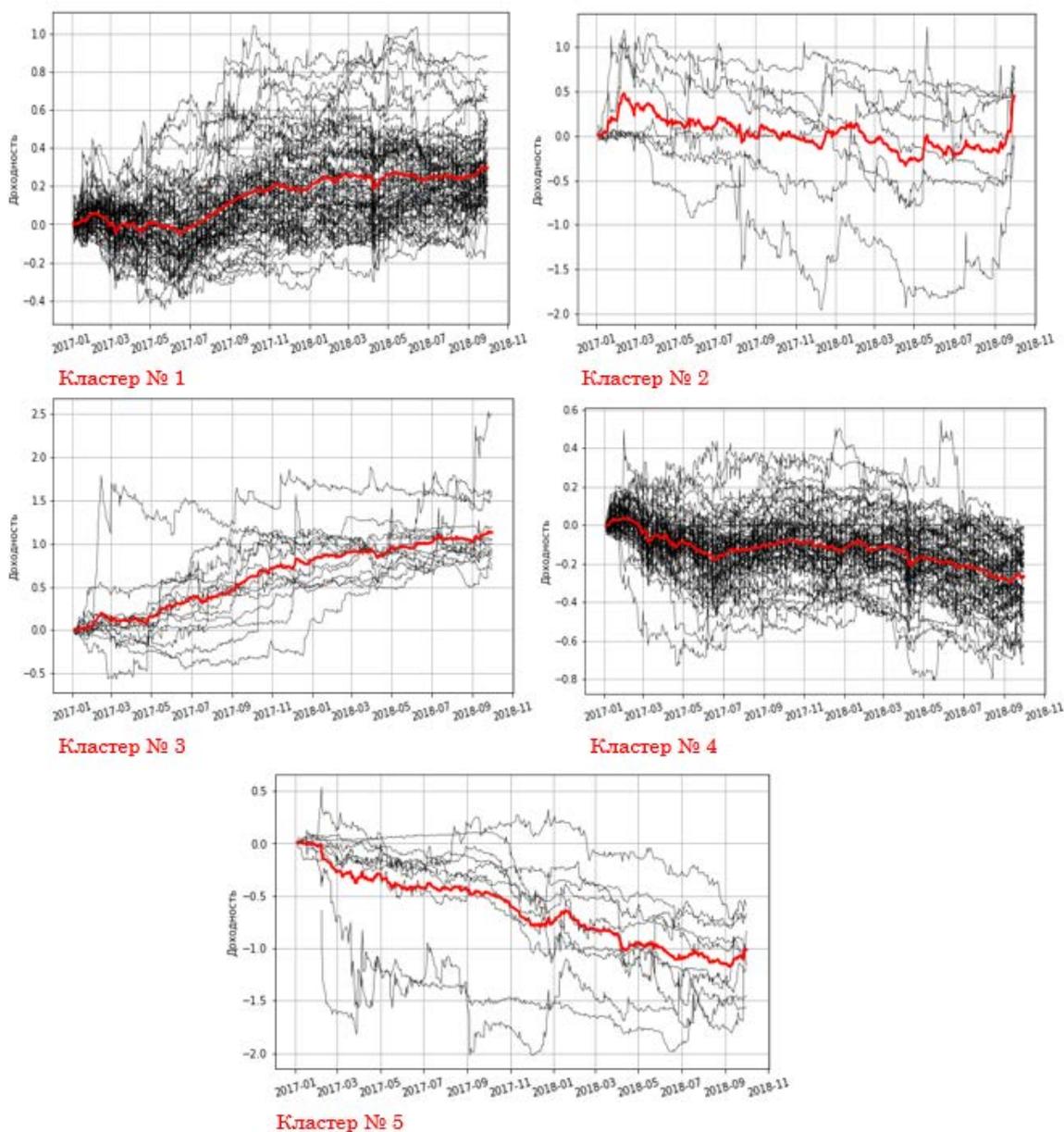


Рис. 3 – Полученные кластеры

Инвестор нацелен минимизировать риск и максимизировать доходность, поэтому очевидна целесообразность исключения ряда акций, обладающих отрицательной или же незначительной доходностью, а также акций,

соответствующих высокому риску, из рассмотрения при формировании портфеля.

Кластерный анализ позволяет значительно быстро исключить неэффективные акции, не анализируя их по отдельности, а выявляя сходство в их поведении и отбрасывая не пригодные для инвестирования группой с общими отрицательными показателями, тем самым, обеспечивая фокусировку на оптимальных активах.

Однако оптимальная работа алгоритма предполагает обновление данных со временем (например, раз в квартал), для этого может проводится либо повторный запуск, либо же добавление (удаление) акций, переходящих из одного кластера в другой.

Выводы. Цель кластерного анализа - разбиение выборки на группы схожих объектов для упрощения последующей обработки данных и дальнейшего принятия решения. В виду этого выделение риска и ожидаемой доходности в качестве параметров кластеризации направлено не на формирование инвестиционного портфеля согласно соотношению доходность-риск, а на отбор акций, его формирующих и их системного рассмотрения в дальнейшем.

Библиографический список:

1. Иванюк В.А., Андропов К.Н., Егорова Н.Е. МЕТОДЫ ОПТИМИЗАЦИИ ИНВЕСТИЦИОННОЙ СТРАТЕГИИ / В.А. Иванюк, К.Н. Андропов, Н.Е. Егорова // Фундаментальные исследования. - 2016. - № 3-1.- С. 159-163.
2. Индекс МосБиржи IMOEX [Электронный ресурс]. Режим доступа: <https://www.moex.com/ru/index/IMOEX/constituents/>
3. Lemieux V., Rahmdel P., Walker R., Wong B.W., Flood M. Clustering Techniques and Their Effect on Portfolio Formation and Risk Analysis / V. Lemieux,

P. Rahmdel, R. Walker, B.W. Wong, M. Flood // Staff Discussion Papers, Office of Financial Research. - 2015. - № 2015-01.

4. Markowitz H. Portfolio Selection / H. Markowitz // Journal of Finance. - 1952 - № 1. - pp. 71-91.

Оригинальность 93%