Методы достижения интерпретируемости алгоритмов машинного обучения

Сенько Олег Валентинович, Кунецова Анна Викторовна

ФИЦ ИУ РАН, ИБХФ им. Н.М. Эмануэля РАН

При использовании большинства методов машинного обучения процесс вычисления компьютерных решений оказывается недоступным пользователю.

Непрозрачность

- не позволяет понять причины выбора компьютерной программой предлагаемого решения
- снижает убедительность выдаваемых результатов
- затрудняет выработку оптимальных действий, связанных с предложеным решением

Непрозрачность или неинтерпретирумость свойственна следующим известным методам машинного обучения:

- Нейросетевые методы
- Метод опорных векторов

В значительной степени интерпретируемыми являются решающие или регрессионнын деревья, а также ассоциативные правила.
Обобщающая способность отдельных решающих деревья часто оказывается слабой по сранению с упомянутыми выше технологиями.

Обобщающая способность существенно возрастает при использовании ансамблей решающих деревьев, получаемых с использованием технологий

- бэггинга
- адаптивного бустинга
- градиентного бустинга

Прозрачность и интерпретируемость также в значительной мере теряются при переходе к ансамблям

Способы достижения интерпретируемости

Интерпретируемость может быть достигнута через аппроксимацию работы обученного алгоритма с помощью набора моделей, связывающих целевую переменную Y с подмножествами признаков из набора $X_1, \ldots, X_n$:

- модели должны достаточно просты для того, чтобы быть наглядно представлены в понятной для пользователей интерпретируемой форме
- модели должны быть статистически достоверны
- набор моделей должен достаточно полно описывать работу алгоритма

Способы достижения интерпретируемости.Двумерные закономерности.

Для достижения наглядности могут быть использованы, например,

- модели, связывающие Y с отдельными переменными из $X_1, \ldots, X_n$
- Бокс-плоты, демонстрирующие различия средних значений переменных из множества $X_1, \ldots, X_n$ в группах, задаваемых категориальной целевой переменной Y

Более точная аппроксимация работы обученного алгоритма может быть достигнута с помощью двумерных моделей, связывающих Y с парами переменных из набора $X_1, \ldots, X_n$.
Использование двумерных моделей позволяет сохранить высокую степень наглядности.

Двумерные модели. Проблема избыточности.

- При верификации более сложных двумерных моделей возникает дополнительная проблема избыточности
- Проблема избыточности возникает, если достоверность более сложной модели оценивать только в целом с помощью одного $p-$значения
- Двумерную модель, связывающую переменную Y с переменными X_i и X_j назовём избыточной, если статистическая достоверность связана с реальным существованием регрессионной связи только с одной из двух переменных
- Например, реальная связь существует с переменной X_i, а переменная X_j на самом деле является шумовой
- Число избыточных закономерностей может оказаться очень велико. Компьютерный анализ должен включать средства их отсева.

Двумерные модели. Проблема множественного тестирования.

При высокой исходной размерности данных существуе высокая вероятность чисто случайного возникновения закономерностей, которые могут оказаться статитически достоверными при формальном применении статистических тестов.

- Компьютерный анализ должен учитывать множественное тестирование. Иными словами, должна проводиться дополнительная коррекция целевых уровней значимости с целью достижения истинной достоверности.

В случае, если Y является непрерывной величиной, для поиска двумерных моделей, удовлетворяющих перечисленным требованиям, потенциально могли бы использоваться

- двумерные регрессионные модели
- дисперсионный анализ (Two way ANOVA)

При этом достоверными могут считаться только такие двумерные модели, у которых регрессионные коэффициенты перед обоими регрессорами оказываются значимыми. Подход является математически обоснованным при справедливости предположения о действительном существовании линейной связи с добавленной нормально распределённой шумовой составляющей. Практика показывает, что существенное отклонение от нормальности нередко приводит к ложным выводам о значимости закономерности.

Возможные подходы. ANOVA. Информационные критерии.

Метод Two way ANOVA может быть использован только, если обе независимые переменные являются категориальными. Кроме того, предположение о нормально распределённой шумовой составляющей также необходимо. В методе Two way ANOVA не имеет прямого решения проблема избыточности.
Общим способом решения проблемы избыточности могло бы являться использование информационных критериев Akaike или BIC.
Недостатком обоих подходов является оценивание ими только адекватности общей сложности моделей. Информационные критерии не оценивают значимость вклада отдельных независимых переменных.

Возможные подходы.Категориальные переменные.

В случаях, когда целевая переменная является категориальной, поиск оптимальных двумерных моделей мог бы производиться с помощью методов распознавания с визуализацией границ между классами.
Однако методы распознавания не включают статистических оценок достоверности вкладов отдельных переменных.
В случае, если категориальными являются одновременно целевая переменная и независимые переменные, для оценки значимости могут быть использованы статистические критерии:

- критерий χ^2
- G-тест

Недостатки - отсутствуют прямые способы статистической оценкии вкладов каждой из независимых переменных.

Возможные подходы.Категориальные переменные.

Для учёта множественного тестирования могут быть использованы известные методы

- Бонферрони
- Бонферрони-Холма

Однако даные методы являются чрезмерно жёсткими, отвергая на самом деле значимые модели.

Способы достижения интерпретируемости. Методы поиска двумерных закономерностей

Одним из возможных способов аппроксимации является использование двумерных моделей, выделяющих внутри совместной области допустимых значений пары переменных (X_i, X_j) подобластей, где преобладают определённые значения переменной Y.
Поиск оптимальных системы двумерных моделей может производится с помощью

Метода оптимальных достоверных разбиений (МОДР)

МОДР предназначени для поиска в данных статистически достоверных закономерностей, связанных с разбиениями интервалов допустимых значений отдельных переменных или совместных областей допустимых значений для пар переменных.

Преимущества технологии ОДР

Метод ОДР удовлетворяет перечисленным выше требованиям:

- Метод обладает высокой наглядностью.
- Вследствие использования непараметрических перестановочных тестов при верификации не требуется предположениний о типе распределений шумовых составляющих. Также отсутствуют ограничения на размер выборок.
- Метод ОДР позволяет для более сложных закономерностей оценивать значимость вклада кажого из элементов с помощью соответствующих $p-$значений. Например, для двумерных закономерностей метод оценивает значимость каждой из двух независимых переменных
- Метод включает эффективную современную процедуру учёта множественного тестирования, основанную опять же на перестановочных тестах

Метод ОДР

Метод ОДР я вляется методом интеллектуального анализа данных (ИАД). Однако в отличие от входящих в группу технологий ИАД методов кластерного анализа (включая метод k-срелних, иерархическая кластеризация SOTA и др.), ОДР предназначен для задач с выделенной целевой переменной.

Метод оптимальных достоверных разбиений. Переход к многомерности.

Основная цель метода ОДР - наглядно представить на экране компьютера по-возможности все достоверные эффекты существующие в данных. Предполагается, что таким способом можно значительно улучшить интерпретируемость и прозрачность, решений, полученных с помощью технологий машинного обучения.
Вместе с тем существует многомерная технология машинного обучения, являющаяся фактически производной от метода ОДР - метод Статистически взвешенных синдромов (СВС). Метод СВС основан на взвешенном голосовании по квадрантам разбиений, в которые попадают вновь распознаваемые объекты.
Метод СВС был успешно использован на ряде биомедицинских и экономических задач, включая задачу прогнозирования отзыва банковских лицензий.

Метод оптимальных достоверных разбиений

В МОДРоптимальные разбиения ищутся внутри фиксированных семейств различного фиксированного уровня сложности. Использовались 4 типа семейств, представленных на рисунке

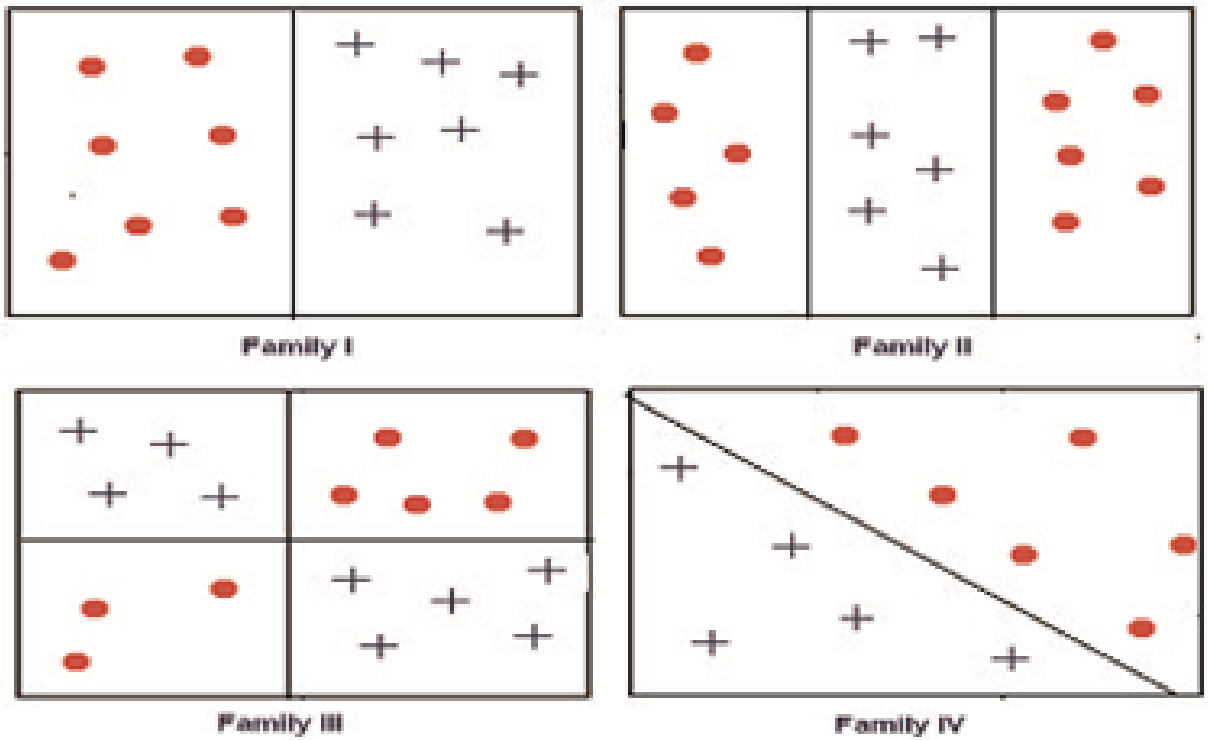


Рис.: 4 типа семейств разбиений

МОДР, семейства разбиений

- Семейство I включает все разбиения интервалов значений одиночных признаков с помощью одной граничной точки
- Семейство II включает все разбиения интервалов значений одиночных признаков с помощью двух граничных точек
- Семейство III включает все разбиения совместной области допустимых значений пары признаков на четыре подобласти с помощью двух границ, параллельных координатным осям
- Семейство IV включает все разбиения совместной области допустимых значений пар признаков на две подобласти с помощьью прямой, произвольно ориетированной относительно координатных осей

МОДР, оптимизируемый функционал

Предположим, что рабиение области допустимых значений на r подобластей индуцирует разбиение обучающей выборки $\widetilde{S}_t$ на подвыборки $\widetilde{s}_1, \ldots, \widetilde{s}_r$. Предположим, что

- $\hat{y}_i$ - среднее значение целевой переменной Y на подвыборке $\widetilde{s}_i$.
- $\hat{y}_0$ - среднее значение целевой переменной Y на всей выборке $\widetilde{S}_t$.
- m_i - размер выборки $\widetilde{s}_i$

Оптимальными считаются разбиения, для которых достигает максимума интегральный функционал

$$Q_I(\widetilde{S}_t) = \sum_{i=1}^{r} \rho(\hat{y}_i, \hat{y}_0) m_i,$$

суммирующий отклонения от средних значений для полной выборки $\widetilde{S}_t$ по всем подобластям разбиения.

МОДР, оптимизируемый функционал

Наряду с интегральным функционалом может использоваться также альтернативный локальный функционал

$$Q_L(\widetilde{S}_t) = \max_{i=1,\dots,r} [\rho(\hat{y}_i, \hat{y}_0) m_i],$$

выделяющий подобласть с максимальным отклонением от средних значений по выборке.

МОДР, примеры

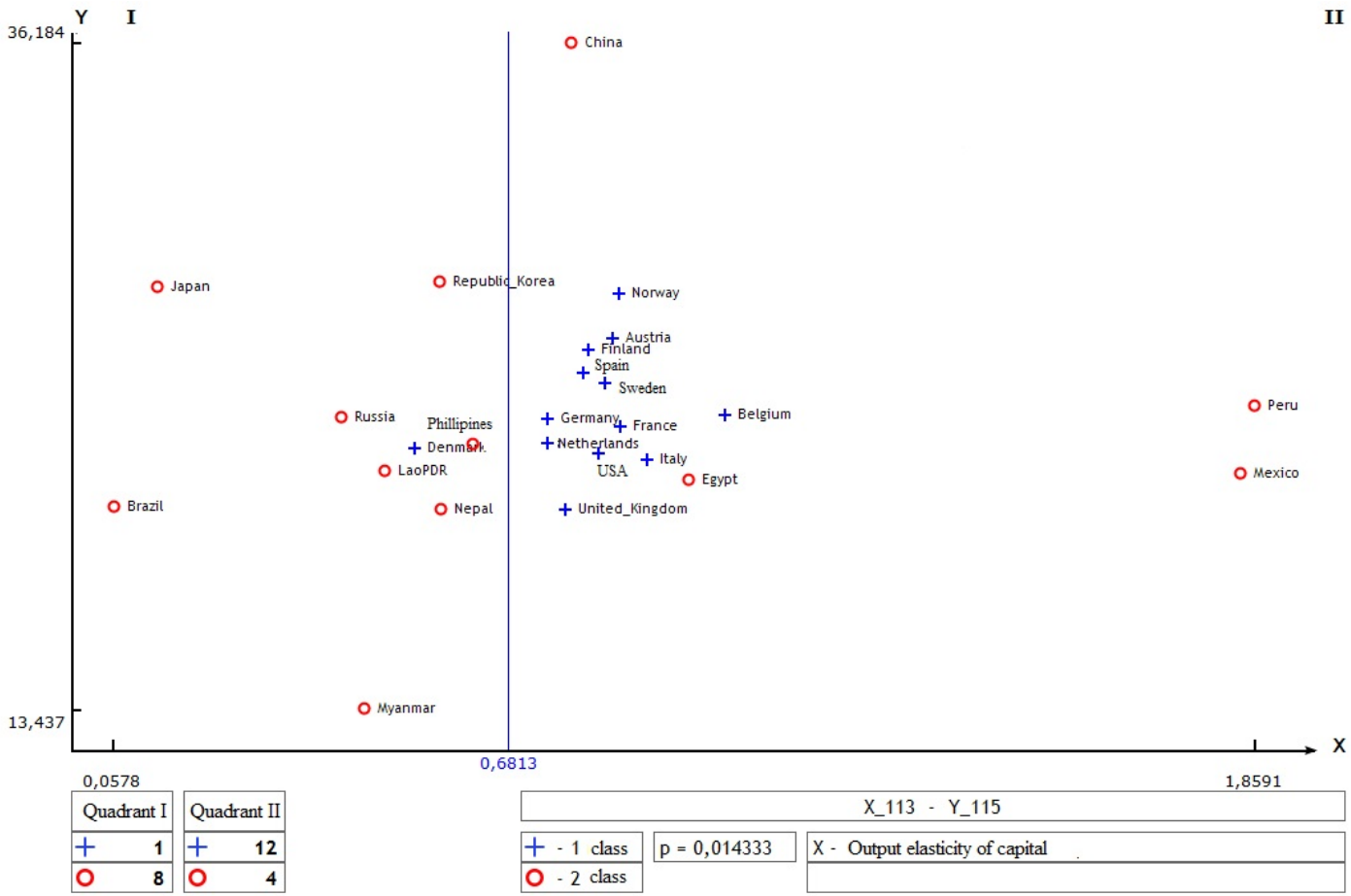


Рис.: Связь типа институциональной матрицы с эластичностью по капиталу в Модели Кобба-Дугласа

МОДР, примеры

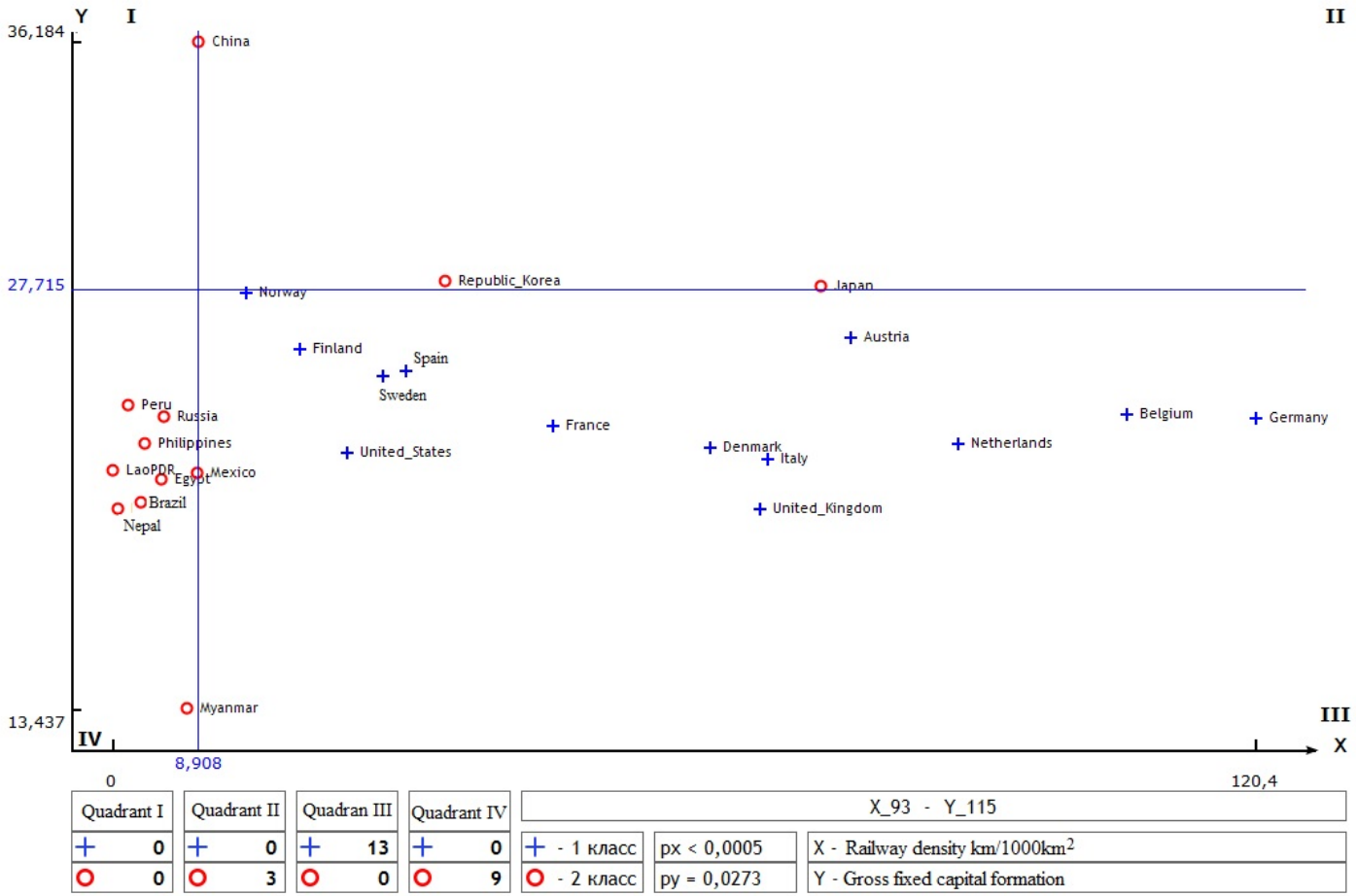


Рис.: Связь типа институциональной матрицы с густотой железнодорожной сети и валовым накоплением основного капитала

Предыдущий слайд относится к исследованиям, связанным с верификацией теории С.Г.Кирдиной о возможности выделения двух базовых типов стран

- страны типа X с редистрибутивной экономикой, унитарной политической системой и доминирующим коммунитарным сознанием
- страны типа Y с рыночной экономикой, федералистской политической системой и доминирующей индивидуалистической компонентой в общественном сознании

МОДР-Верификация

Стандартные статистические тесты, основанные на сравнении средних значений целевой переменной Y в группах, индуцируемых разбиением, правомерно использовать если

оптимальное разбиение и статистика критерия вычисляются по двум независимым выборкам

При условии невозможности генерации двух независимых выборок достаточного объёма целесообразно использовать

Перестановочные тесты

Перестановочный тест

Предположим, что требуется найти оптимальное разбиение интервала допустимых значений признака X_i внутри Семейства I по выборке

$$\widetilde{S}_t = \{(y_1, x_{1i}), \dots (y_m, x_{mi})\}$$

- На первом шаге найдём пороговое значение для признака X_i, при котором достигается максимальное значение функционала $Q_L(\widetilde{S}_t)$, которое обозначим как Q_0.
- Сгенерируем N случайных независимых перестановок $\mathbf{f}^1, \dots, \mathbf{f}^N$ натуральных чисел из множества $\{1, \dots, m\}$,где $\mathbf{f}^j = (f_1^j, \dots, f_m^j)$

Перестановочный тест

- Зафиксируем равным 0 значение счётчика ng.

Повторим следующие шаги при значениях $j = 1, \ldots, N$.

- Получим из $\widetilde{S}_t$ по перестановке $\mathbf{f}^j$ выборку $\widetilde{S}_j^p = \{(y_{f_1^j}, x_{1i}), \ldots, (y_{f_m^j}, x_{mi})\}$
- Найдём пороговое значение для признака X_i, при котором достигается максимальное значение функционала $Q_L(\widetilde{S}_j^p)$, которое обозначим как Q_j^p
- В случае выполнения неравенства $Q_j^p \geq Q_0$ значение счётчика ng увеличивается на 1

Перестановочный тест

В качестве оценки $p-$значения используется отношение ng к общему числу перестановок:

$$p = \frac{ng}{N}$$

Описанный вариант перестановочного теста проверяет нулевую гипотезу о независимости переменных Y и X и позволяет эффективно выявлять простейшие закономерности с одной граничной точкой.
Применение данного варианта для верификации более сложных двумерных моделей приводит к ошибочному выявлению частично ложных закономерностей.
Мы будем называть закономерность, характеризующую связь переменной Y с переменными X_i и X_j, частично ложной, если её статистическая достоверность целиком определяется связью Y только с одной из двух переменных: X_i или X_j.

Перестановочный тест для двумерных закономерностей.

Было предложено два подхода к верификации,направленных на исключение частично ложных закономерностей. Оба подхода предусматривают вычисление двух $p-$значений, характеризующих отдельно вклады переменных X_i и X_j. Подход I основан на сравнении функционала качества для двумерной закономерности, характеризующей совместную связь Y с переменными X_i и X_j, с фунеционалами качества для двух одномерных закономерностей, характеризующих связь Y с каждой из переменных.
Подход II основан на попытке опровержения с помощью двумерной модели возможности исчерпывающего описания данных простыми моделями.

Верификация двумерных закономерностей. Подход I

Предположим, что требуется оценить достоверность двумерной закономерности, связанной с оптимальным разбиением совместной области допустимых значений признаков $X_{i'}$ и $X_{i''}$ внутри Семейства II по выборке

$$\widetilde{S}_t = \{(y_1, x_{1i'}, x_{1i''}), \dots (y_m, x_{mi'}, x_{mi''})\}$$

- На первом этапе по $\widetilde{S}_t$ найдём оптимальную совместную пару пороговых значений для признаков $X_{i'}$ и $X_{i''}$, вычислим соответствующее значение $Q_L(X_{i'}, X_{i''}, \widetilde{S}_t)$
- Найдём оптимальные пороговые значения для признаков $X_{i'}$ и $X_{i''}$ по отдельности, вычислим соответствующие значения $Q_L(X_{i'}, \widetilde{S}_t)$,$Q_L(X_{i''}, \widetilde{S}_t)$

Верификация двумерных закономерностей. Подход I

- Вычислим величины $D'_0 = Q_L(X_{i'}, X_{i''}, \widetilde{S}_t) - Q_L(X_{i'}, \widetilde{S}_t)$
  $D''_0 = Q_L(X_{i'}, X_{i''}, \widetilde{S}_t) - Q_L(X_{i''}, \widetilde{S}_t)$
- Зафиксируем равным 0 значение счётчиков ng' и ng''
- Сгенерируем N случайных независимых перестановок $\mathbf{f}^1, \ldots, \mathbf{f}^N$ натуральных чисел из множества $\{1, \ldots, m\}$,где $\mathbf{f}^j = (f_1^j, \ldots, f_m^j)$

Повторим следующие шаги при значениях $j = 1, \ldots, N$.

- Получим из $\widetilde{S}_t$ по перестановке $\mathbf{f}^j$ выборку
  $\widetilde{S}_j^p = \{(y_{f_1^j}, x_{1i'}, x_{1i''}), \ldots, (y_{f_m^j}, x_{mi}, x_{mi''})\}$
- По $\widetilde{S}_j^p$ найдём оптимальную совместную пару пороговых значений для признаков $X_{i'}$ и $X_{i''}$, вычислим соответствующее значение $Q_L(X_{i'}, X_{i''}, \widetilde{S}_j^p)$

Верификация двумерных закономерностей. Подход I

- Найдём оптимальные пороговые значения для признаков $X_{i'}$ и $X_{i''}$ по отдельности, вычислим соответствующие значения $Q_L(X_{i'}, \widetilde{S}_j^p)$, $Q_L(X_{i''}, \widetilde{S}_j^p)$
- Вычислим величины $D'_j = Q_L(X_{i'}, X_{i''}, \widetilde{S}_j^p) - Q_L(X_{i'}, \widetilde{S}_j^p)$ $D''_j = Q_L(X_{i'}, X_{i''}, \widetilde{S}_j^p) - Q_L(X_{i''}, \widetilde{S}_j^p)$
- В случае выполнения неравенства $D'_j \geq D'_0$ значение счётчика ng' увеличивается на 1
- В случае выполнения неравенства $D''_j \geq D''_0$ значение счётчика ng'' увеличивается на 1

Статистическая значимость вклада переменной $X_{i'}$ оценивается с помощью. $p-$значения, равного отношению $p' = \frac{ng''}{N}$.
Статистическая значимость вклада переменной $X_{i''}$ оценивается с помощью $p-$значения, равного отношению $p'' = \frac{ng'}{N}$.

Верификация двумерных закономерностей. Подход I

Значимость закономерности будем оценивать как максимальное из двух рассчитанных $p-$значений:

$$p = \max\left(p', p''\right)$$

Верификация двумерных закономерностей. Подход II

Подход II основан на попытке опровержения с использованием двумерного оптимального разбиения нулевой гипотезы об исчерпывающем описании данных с помощью более простой одномерной модели.

Предположим, что оптимальное разбиение интервала значений признака $X_{i''}$ включает два подинтервала q_l и q_r.

При справедливости нулевой гипотезы целевая переменная Y очевидно не зависит от от $X_{i'}$ и $X_{i''}$ внутри каждого из подинтервалов q_l и q_r. В этом случае любые перестановки Y относительно фиксированных позиций переменных $X_{i'}$ и $X_{i''}$ внутри каждого из подинтервалов оказываются равновероятными.

Высокое значение функционала $Q_l(X_{i'}, X_{i''}, \widetilde{S}_t)$ должно свидетельствовать зависимости Y не только от $X_{i''}$ но и от $X_{i'}$

Верификация двумерных закономерностей. Подход II

На первом этапе по $\widetilde{S}_t$ найдём оптимальную совместную пару пороговых значений для признаков $X_{i'}$ и $X_{i''}$, вычислим соответствующее значение $Q_L(X_{i'}, X_{i''}, \widetilde{S}_t)$, которое обозначим Q_0
Сначала оценим значимость вклада в закономерность признака $X_{i'}$. С этой целью по граничной точке $\delta_{i''}$ для найденного двумерного разбиения сформируем две выборки:

- выборку $\widetilde{S}_l$, содержащую объекты из $\widetilde{S}_t$, удовлетворяющие условию $x_{ji''} < \delta_{i''}$
- выборку $\widetilde{S}_r$, содержащую объекты из $\widetilde{S}_t$, удовлетворяющие условию $x_{ji''} > \delta_{i''}$

Пусть

- m_l - число объектов в выборке $\widetilde{S}_l$
- m_r - число объектов в выборке $\widetilde{S}_r$

Верификация двумерных закономерностей. Подход II

- Сгенерируем N случайных независимых перестановок $\mathbf{f}^1(l), \ldots, \mathbf{f}^N(l)$ натуральных чисел из множества $\{1, \ldots, m\}$,где $\mathbf{f}^j = [f_1^j(l), \ldots, f_m^j(l)]$
- Сгенерируем N случайных независимых перестановок $\mathbf{f}^1(r), \ldots, \mathbf{f}^N(r)$ натуральных чисел из множества $\{1, \ldots, m\}$,где $\mathbf{f}^j(r) = [f_1^j(r), \ldots, f_m^j(r)]$

Повторим следующие шаги при значениях $j = 1, \ldots, N$.

- Получим из $\widetilde{S}_l$ по перестановке $\mathbf{f}^j(l)$ выборку $\widetilde{S}_j^p(l) = \{(y_{f_1^j(l)}, x_{1i'}, x_{1i''}), \ldots, (y_{f_m^j(l)}, x_{mi}, x_{mi''})\}$
- Получим из $\widetilde{S}_r$ по перестановке $\mathbf{f}^j(r)$ выборку $\widetilde{S}_j^p(r) = \{(y_{f_1^j(r)}, x_{1i'}, x_{1i''}), \ldots, (y_{f_m^j(r)}, x_{mi}, x_{mi''})\}$

Верификация двумерных закономерностей. Подход II

- По объединению $\widetilde{S}_j^p = \widetilde{S}_j^p(l) \cup \widetilde{S}_j^p(r)$ найдём оптимальную совместную пару пороговых значений для признаков $X_{i'}$ и $X_{i''}$, вычислим соответствующее значение $Q_L(X_{i'}, X_{i''}, \widetilde{S}_j^p)$, которое обозначим Q_j^p
- В случае выполнения неравенства $Q_j^p \geq Q_0$ значение счётчика ng' увеличитывается на 1

Статистическая значимость вклада переменной $X_{i'}$ оценивается с помощью $p-$значения, равного отношению $\frac{ng'}{N}$.
Аналогичным образом оценим вклад признака $X_{i''}$

Верификация двумерных закономерностей. Подход II

По граничной точки $\delta_{i'}$ для найденного двумерного разбиения сформируем две выборки:

- выборку $\widetilde{S}_l$, содержащую объекты из $\widetilde{S}_t$, удовлетворяющие условию $x_{ji'} < \delta_{i'}$
- выборку $\widetilde{S}_r$, содержащую объекты из $\widetilde{S}_t$, удовлетворяющие условию $x_{ji'} > \delta_{i'}$

Далее повторяем те же самые операции, которые использовались при верификации вклада признака $X_{i'}$. То есть генерируем по $\widetilde{S}_l$ и $\widetilde{S}_r$ случайную выборку $\widetilde{S}_j^p$, по которой вычисляем значение функционала Q_j^p. При выполнении неравенства $Q_j^p \geq Q_0$ значение счётчика ng' увеличивается на 1.

Статистическую значимость вклада переменной $X_{i'}$ оценивается с помощью. $p-$значения, равного отношению $\frac{ng''}{N}$

Примеры двумерных закономерностей. Связь корреляций роста ВВП с внутренним долгом и государственными тратами с типом институциональных матриц

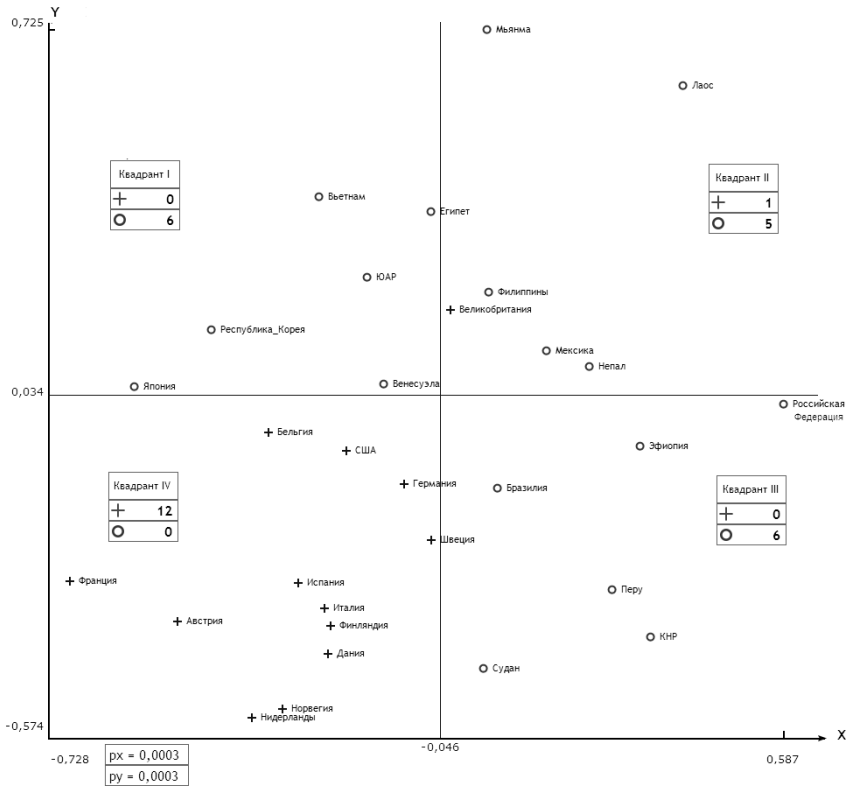


Рис.: По оси X отложен коэффициент корреляции роста ВВП с Внутренним долгом с лагом 4 года. По оси Y отложен коэффициент корреляции Государственных трат с Внутренним долгом с лагом 4 года. + страны типа Y, o - страны типа X

Примеры двумерных закономерностей. Связь поярного индекса с Совокупными активами и Себестоимостью продукции

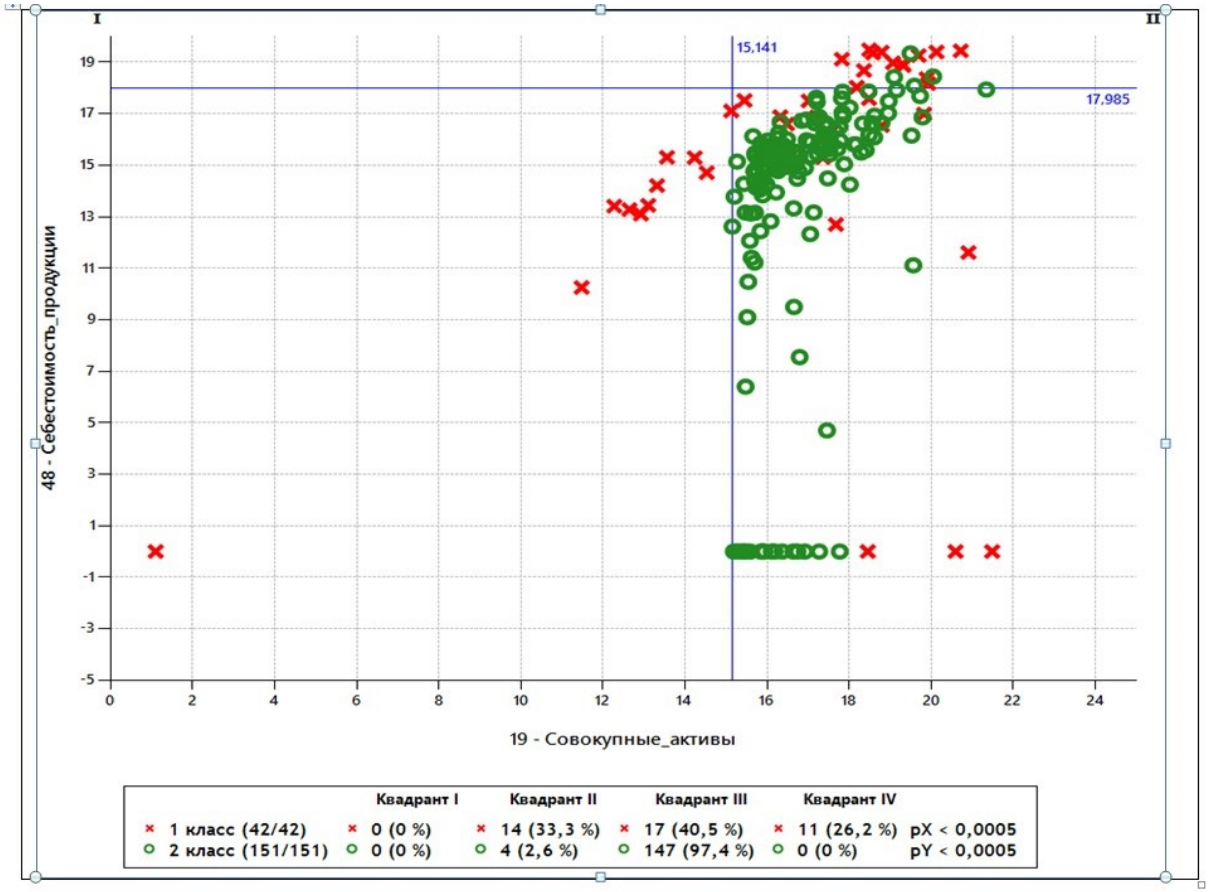
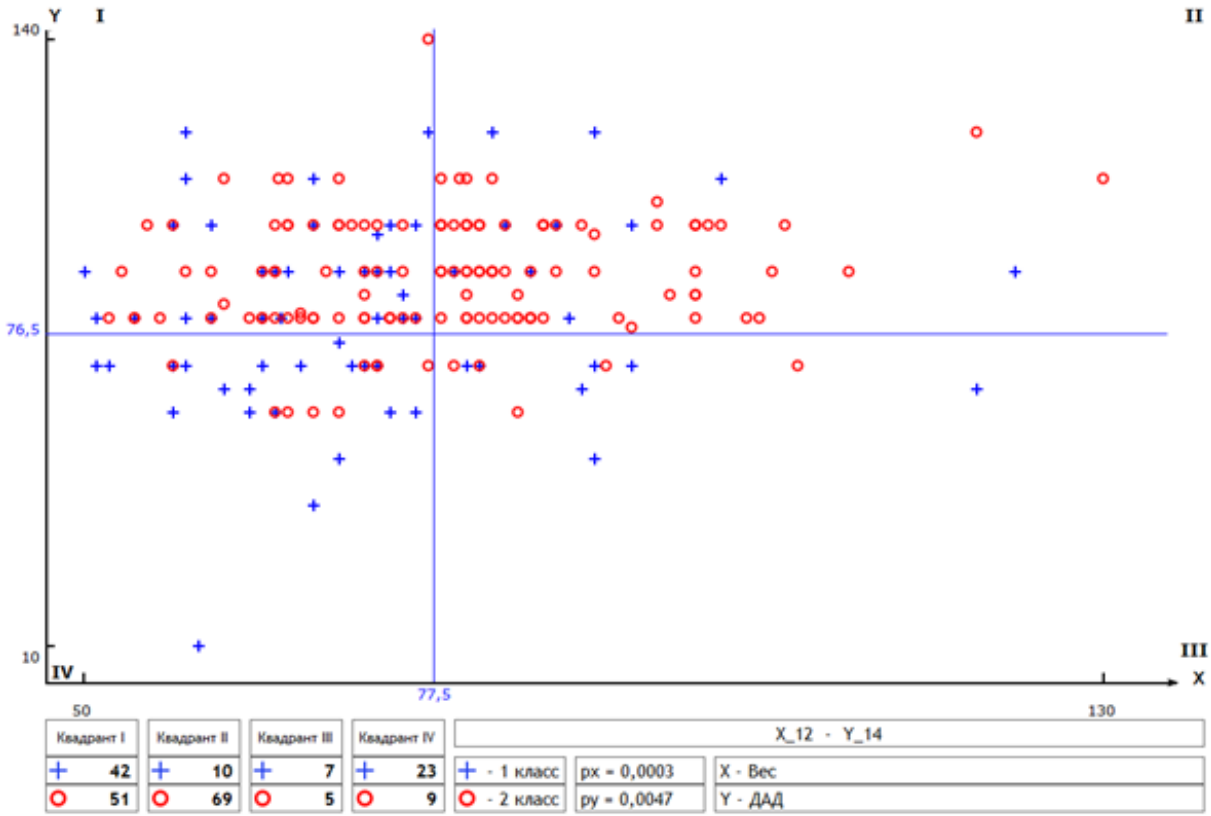


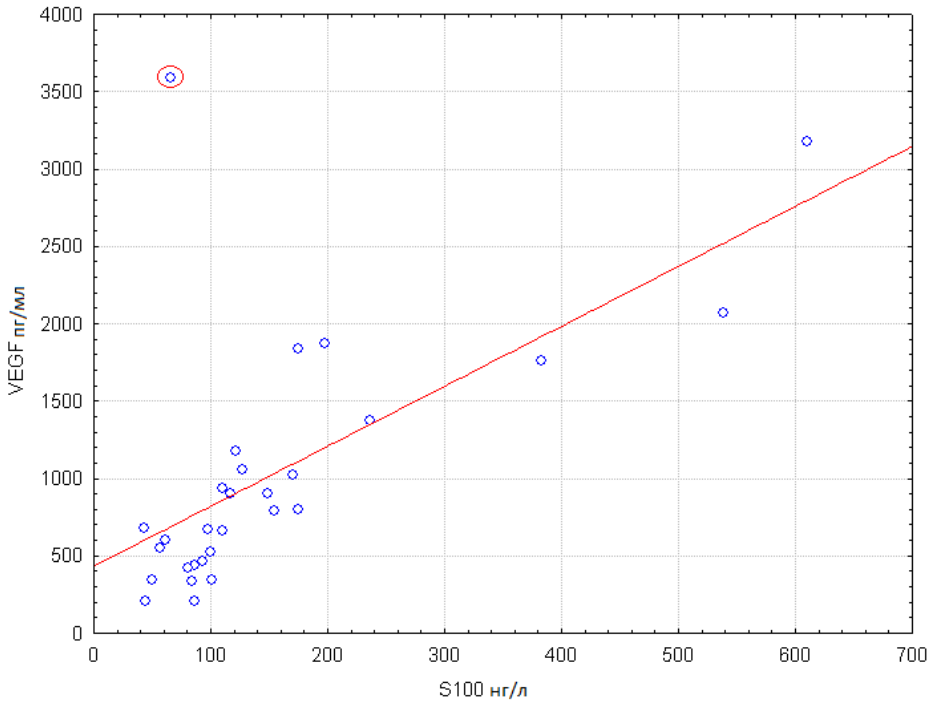
Рис.6. Двумерная диаграмма рассеяния. Ось X – логарифмы совокупных активов, ось Y – логарифмы себестоимости продукции.

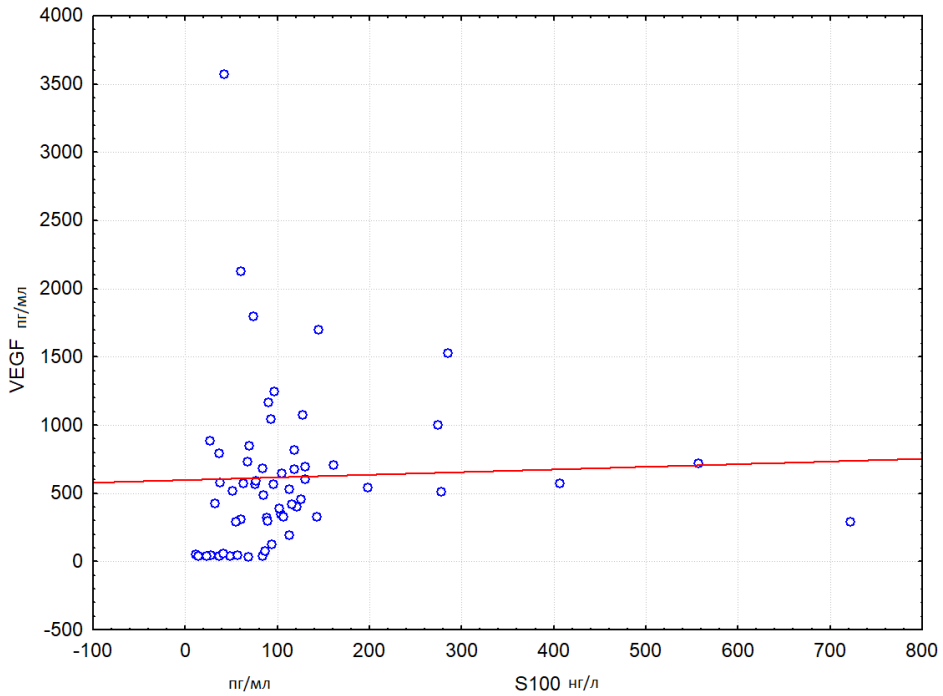
Рис.: X - организации с высоким полярным индексом, **o** - организации с низким полярным индексом

Примеры закономерностей. Медицина. Связь летальности от сердечно-сосудистых заболеваний в учреждениях уголовно-исполнительной системы с весом и диастолическим давление



Примеры закономерностей. Медицина. Влияние гипоксии на связь фактора роста сосудов с уровнем протеинов S100





Различие корреляции между S100 и VEGF в группах с $sO2 < 38.4\%$ и $sO2 > 38.4\%$.

Проблема множественного тестирования в методе ОДР

Для учёта множественного тестирования в методе ОДР разработана техника, являющаяся по сути модификацией технологии, предложенной в работе Westfall и Young в 1993 г.

- Предположим, что нам требуется найти скорректированною значимость закономерности, связанной с оптимальным разбиением области допустимых значений признаков $X_{i'}$ и $X_{i''}$ и оцениваемой $p-$значениями p' и p''. Предполагается, что $p-$значения были получены с помощью описанного ранее Подхода I. Общее $p-$значение для закономерности оценивается тогда как $p_o = \max(p', p'')$

Проблема множественного тестирования в методе ОДР

- С помощью описанной ранее методики генерируется множество случайных выборок $\widetilde{S}_p^j = \{(y_{f_1^j}, \mathbf{x_1}), \ldots, (\mathbf{y_{f_m^j}}, \mathbf{x_m})\}$, где $j = 1, \ldots, N$
- С помощью опять же Подхода I для каждой пары признаков и для каждой выборки $\widetilde{S}_p^j$ вычисляем общие $p-$значения
- для каждой выборки $\widetilde{S}_p^j$ вычисляем минимальное общее $p-$значение p_m^j
- В качестве общего скорректированного $p-$значения используется доля случайных выборок для которых $p_m^j \leq p_o$

Границы применения перестановочных тестов

Использование перестановочного теста при верификации регрессионной модели допустимо при справедливости предположения о равновероятности различных перестановок целевой переменной Y относительно фиксированных позиций векторов независимых переменных.

Для того, чтобы такое предположение оказалось справедливым, необходимо включение в нулевую гипотезу наряду с предположением о независимости Y от независимых переменных также и дополнительного предположения о независимости между собой наблюдений

Подобное предположение чаще всего является несостоятельным для показателей, входящи в многомерные временные ряды. Например, гипотеза о взаимной независимости наблюдений очевидно является несостоятельной для в случаях когда компоненты многомерных рядов соответствуют процессу случайного блуждания.

Применение методов верификации, основанных на ресэмплинге, при поиске закономерностей во временных рядах

Перестановочный тест является процедурой верификации, основанной на ресэмплинге, то есть

- Качество закономерностей, полученных на реальных данных, сравнивается с качеством закономерностей, полученных на данных, генерируемых случайным процессом.
- Для случайного процесса соблюдаются условия нулевой гипотезы.

Для верификации закономерностей, найденных по многомерным временным рядам, могут быть использованы технологии ресэмплинга, основанные, например, на бутстрэпе. Качество аппроксимации для регрессионных моделей, построенных по наблюдаемым временным рядам сравнивается с качесством аппроксимации для регрессионных моделей, построеных по временным рядам искусственно сгенерированных в соответствии с процессом случайного блуждания.