

Интеллектуальные технологии анализа
клинических данных, основанные на
универсальных методах статистической
верификации



ООО АЗФОРУС


Сенько Олег Валентинович

Кузнецова Анна Викторовна

Методы машинного обучения

Методы машинного обучения предназначены для создания алгоритмов предсказывающих значений одной из величин по известным значениям других переменных.

Генерация предсказывающих алгоритмов происходит в автоматическом режиме по обучающей информации, представляющей собой набор примеров, для которых известны значения как предсказываемой так и предсказывающих переменных.



Какие медицинские задачи могут быть решены с помощью МО?

Методы машинного обучения могут быть использованы

- для решения задач диагностики,
- для решения задачи оценивания индивидуального риска развития патологии, возникновения рецидива,
- для решения задачи оценивания риска неблагоприятного исхода

Почему МО актуально для медицины?

Использование моделей, предлагаемых «классическими» научными дисциплинами, то есть физикой, химией, биологией оказывается невозможным или весьма затруднительным из-за чрезвычайной сложности изучаемой системы - человеческого организма

Почему МО актуально для медицины?

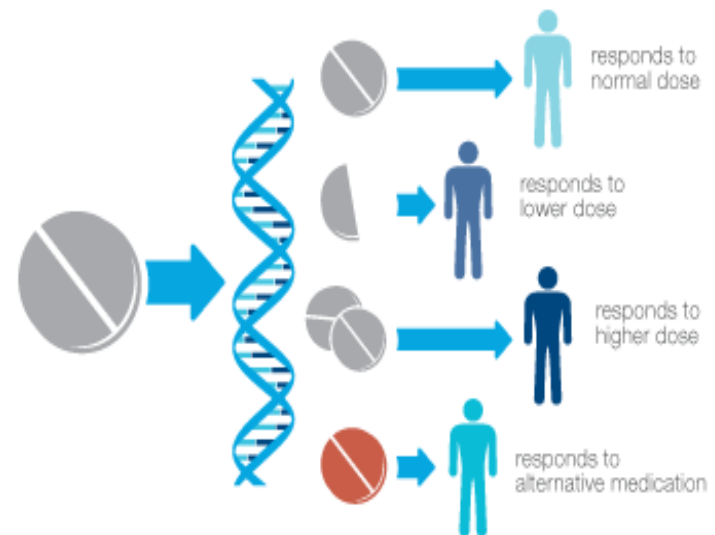
Архивы медицинских данных, находящиеся в распоряжении медицинских учреждений, содержат большое количество сведений о различных конкретных случаях заболевания, включая данные об анамнезе, ходе и тяжести течения заболевания, значения разнообразных клинических, лабораторных и инструментальных показателей.

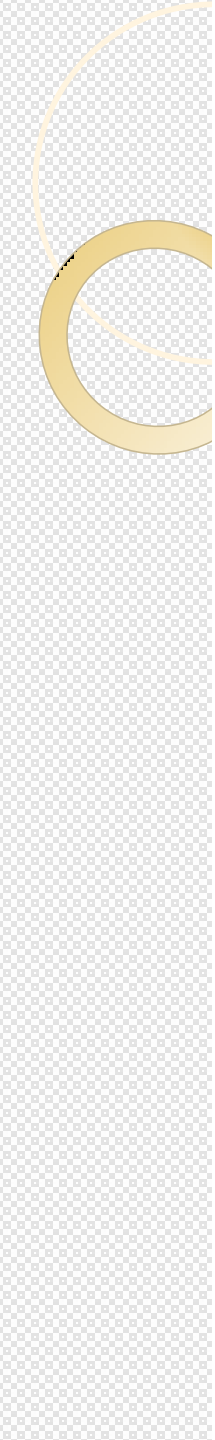
Перспективы использования МО - Персонализированная (прецизионная) медицина

Машинное обучение и ИАД являются важными инструментами развития персонализированной медицины, максимально учитывающей при диагностике и выборе лечения

индивидуальные особенности пациента

- генотип
- анамнез и т.д.





Технологии машинного обучения, используемые при решении «медицинских» задач

**Нейронные сети, включая
конволюционные и рекуррентные сети
Решающие деревья и леса
Метод опорных векторов
Логистическая регрессия
Методы, основанные на голосовании
по системам закономерностей**

Проблема чёрного-ящика

Одной из проблем, связанных с применением МО является проблема чёрного ящика



Что в чёрном ящике?

В чёрном ящике находятся закономерности, связывающие прогнозируемую величину с прогностическими факторами.

Такие закономерности могут оказаться слишком сложными для того, чтобы их можно было представить наглядно.

Однако чаще работа алгоритма сводится к композиции понятных и интерпретируемых эффектов.

Как открыть чёрный ящик?

Что такое закономерность?

Под закономерностью понимается устойчивая связь между показателями, то есть связь, которая существует не только на исследуемой выборке, но сохраняется и на других аналогичных данных.

Поиск закономерностей должен обязательно включать верификацию.

Как правило используется статистическая верификация.

Как открыть чёрный ящик?

Более простые закономерности могут найдены с помощью стандартных статистических методов

Однако простые закономерности как правило уже давно известны и используются в медицинской практике





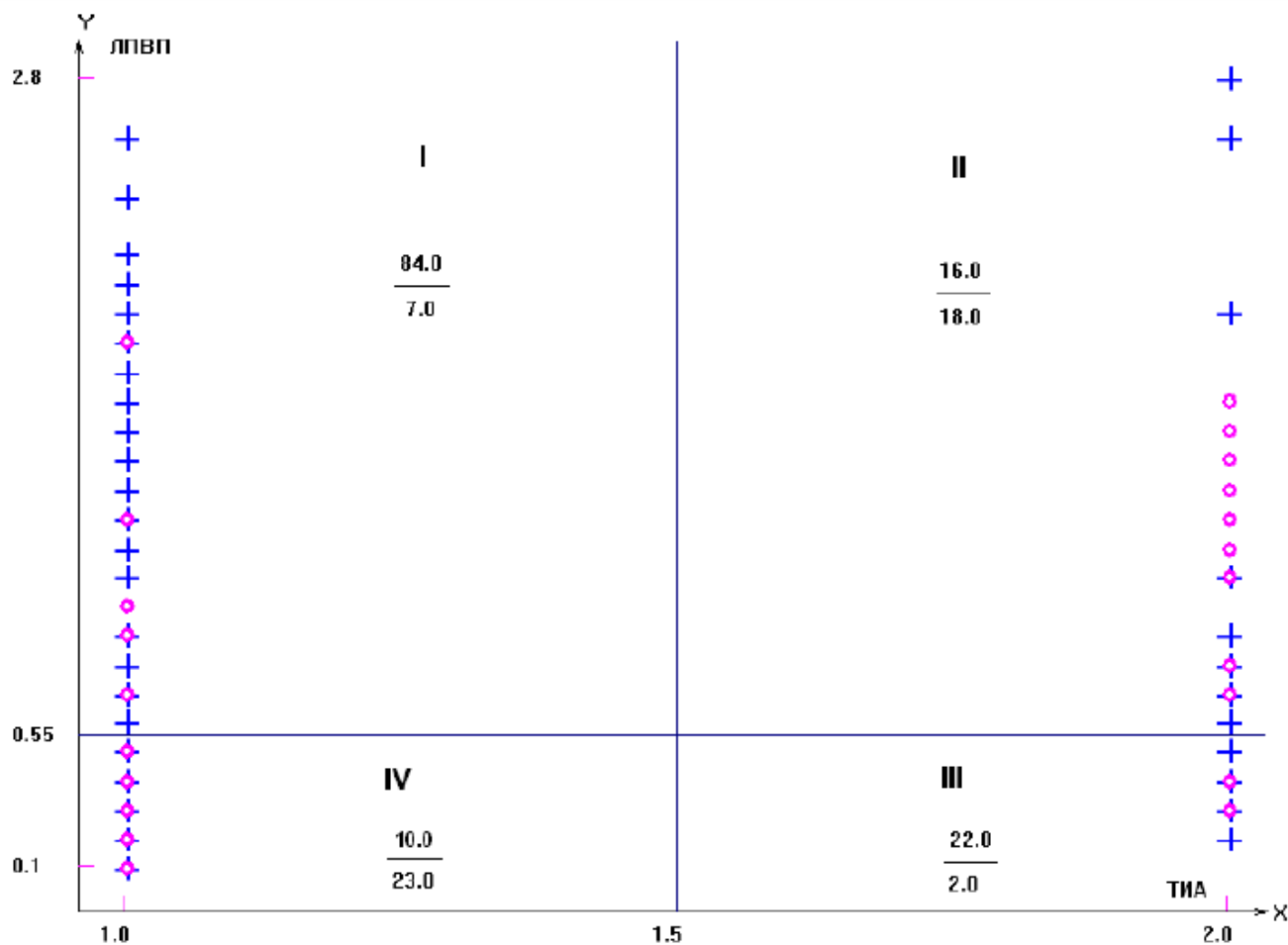
Как открыть чёрный ящик?

Для поиска более сложных закономерностей, описывающих нелинейную связь нескольких факторов, могут быть использованы как непараметрические методы моделирования, так и комбинация линейных параметрических и непараметрических методов.

Одной из возможностей является использование оптимальных разбиений совместных областей значений признаков.

Связь наличия ишемического инсульта с уровнем ЛПВП и наличием транзиторных ишемических атак

КУЗНЕЦОВА и др.





Верификация сложных закономерностей

1. При верификации сложных закономерностей должно проводиться статистическое оценивание правомерности использования каждого элемента
2. Предположения о нормальности вероятностных распределений часто оказываются несостоятельными
3. Стандартные статистические тесты не могут использоваться для оценивания статистической значимости
4. Для верификации сложных закономерностей могут эффективно использоваться технологии верификации, основанные на перестановочных тестах.



Верификация сложных закономерностей

- - Проблема множественного тестирования -- случайное возникновение конфигураций данных, формально соответствующих достоверным закономерностям, при высоком числе исходных признаков.

Поэтому требуется дополнительная коррекция значимости.

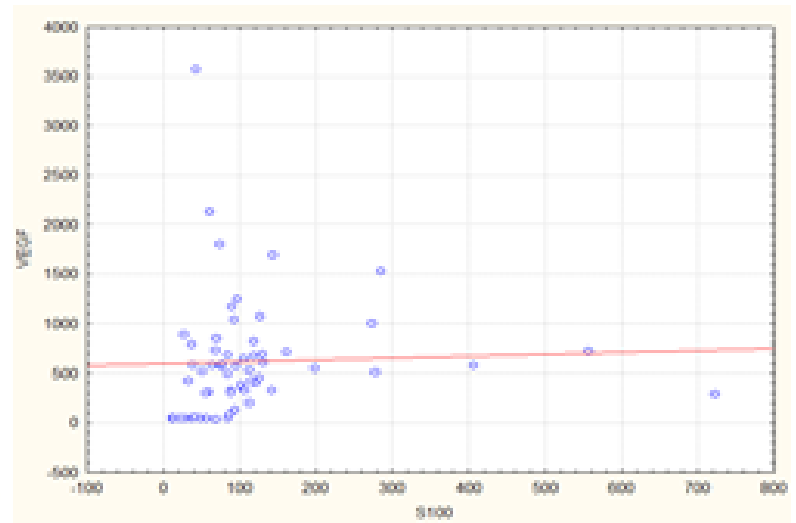
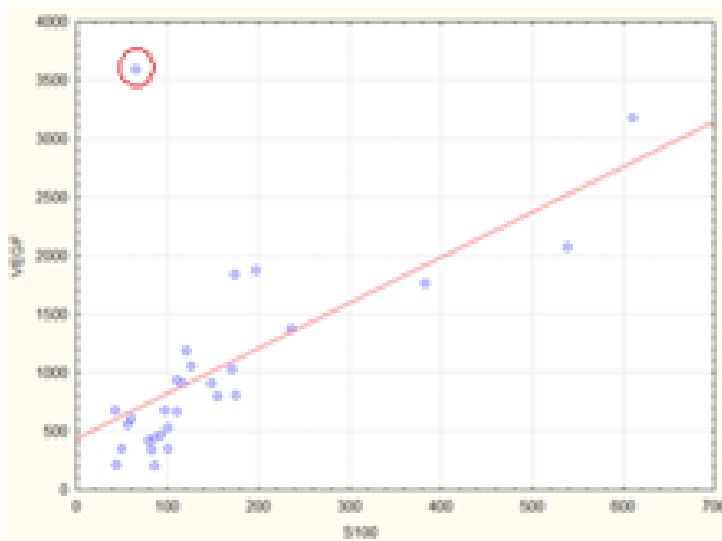
Проблема множественного тестирования становится актуальнее при верификации более сложных закономерностей.

Проблема множественного тестирования

Использование стандартных методов коррекции типа Бонферрони-Холма приводит к заниженным оценкам значимости.

Популярным в настоящее время подходом является опять же использование перестановочных технологий, позволяющих получать реалистичную коррекцию

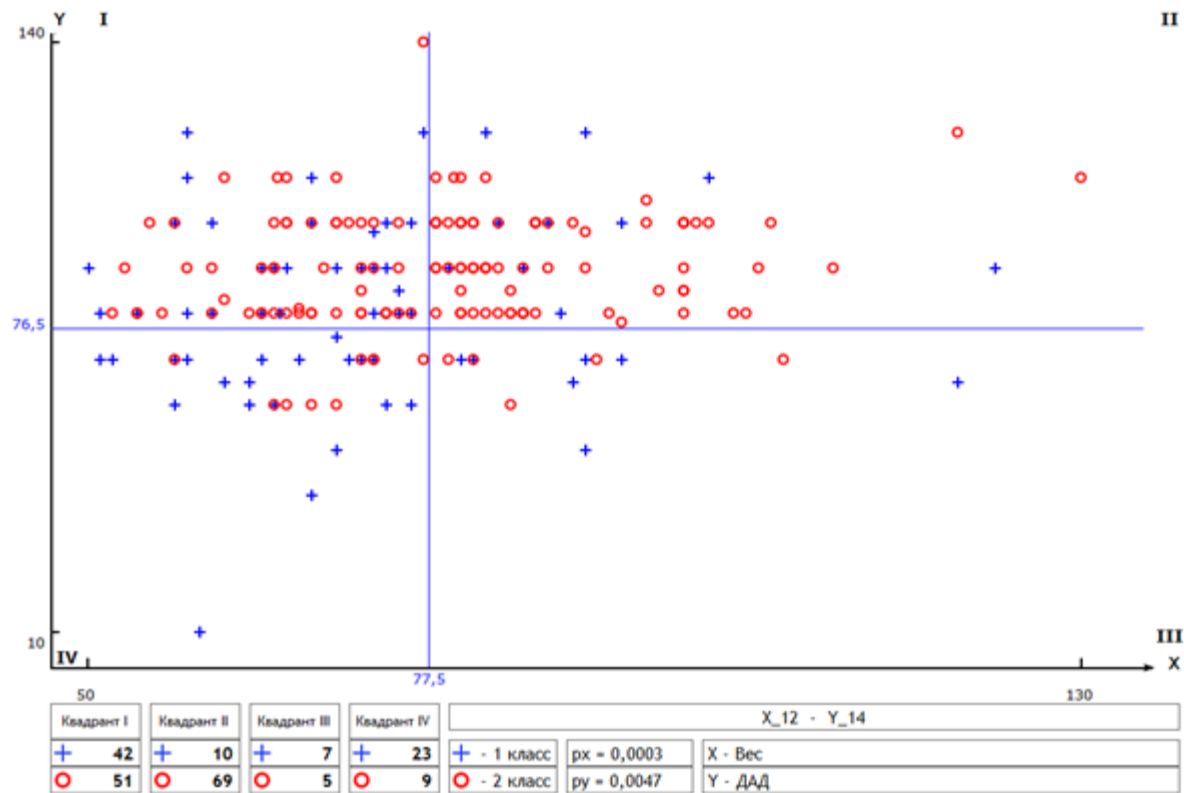
Связь уровня VEGF с показателями оксиметрии



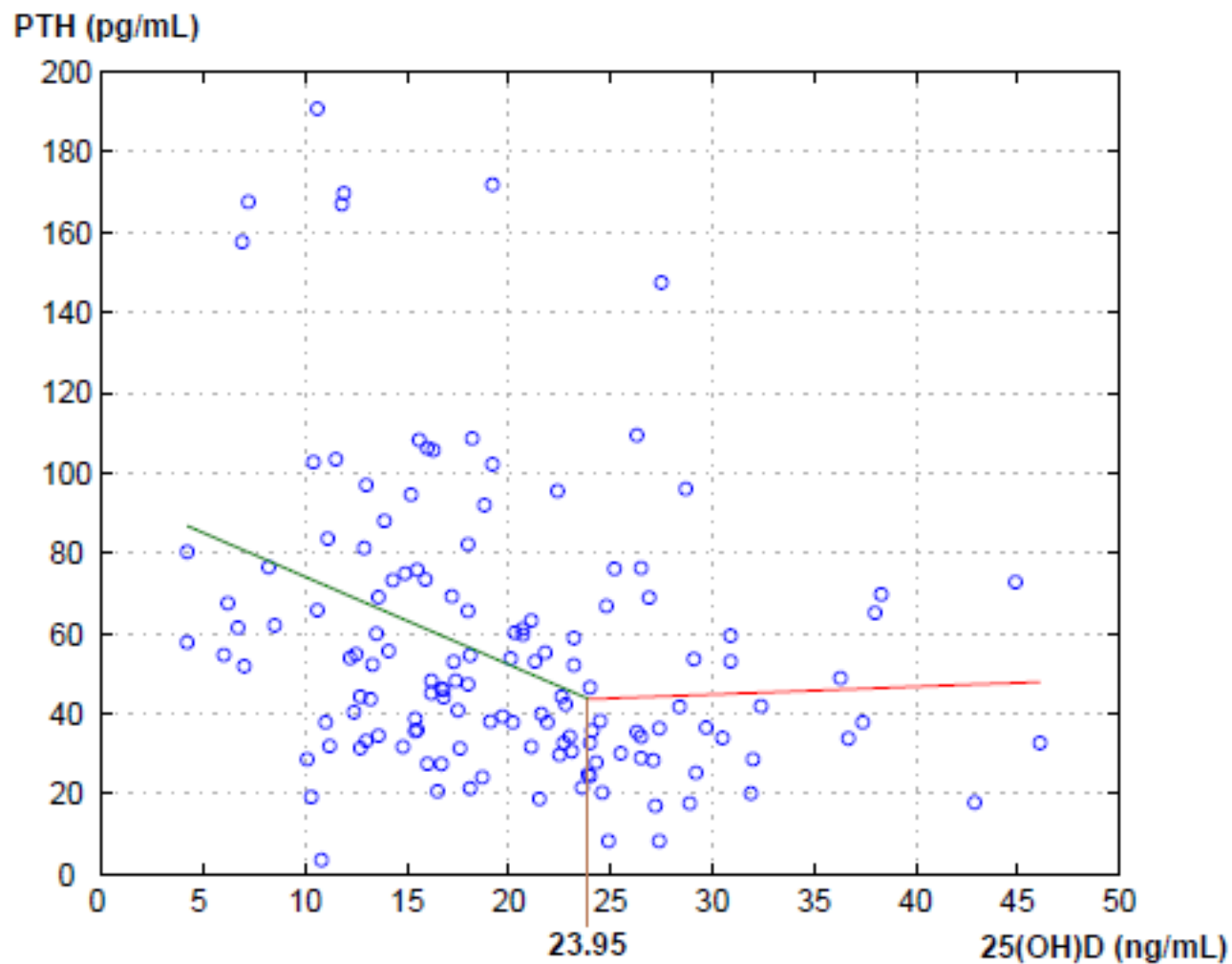
Слева изображена зависимость VEGF от S100 при значении показателя $sO_2 < 0.384$

Справа изображена зависимость VEGF от S100 при значении показателя $sO_2 > 0.384$

Связь летальности в клиниках петенциарной системы с весом и ДАД



Связь содержания ПТГ с содержанием витамина D





СПАСИБО ЗА ВНИМАНИЕ !!!