

МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ



Сенько Олег Валентинович

Методы машинного обучения

Методы машинного обучения предназначены для создания алгоритмов предсказывающих значений одной из переменных 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.

Методы машинного обучения

- Y – предсказываемая переменная
- X – переменные, по значениям которых производится предсказывание

Обучающая информация

$$\tilde{S}_t = \{(y_1, \mathbf{x}_1), \dots, (y_m, \mathbf{x}_m)\}$$

y_j - значение целевой переменной Y ,

\mathbf{x}_j - вектор значений X – переменных

Виды задач машинного обучения

В зависимости от характера предсказываемой переменной выделяются несколько видов задач машинного обучения.

Если нам требуется предсказать значение категориальной переменной Y которая принимает значения из некоторого конечного множества категорий (классов), то говорится о решении **задачи распознавания**

Виды задач машинного обучения

Если нам требуется предсказать значение числовой переменной Y принимающей значения из некоторого интервала числовой оси, то говорится о решении **задачи регрессионного анализа**

Если нам требуется предсказать вид кривой, описывающей вероятность того, что некоторый объект просуществует до заданного момента времени, то говорится о решении **задачи выживаемости**

Интеллектуальный анализ данных (ИАД)

Наряду с задачами машинного обучения можно выделить тесно связанные с ними задачи интеллектуального анализа данных.

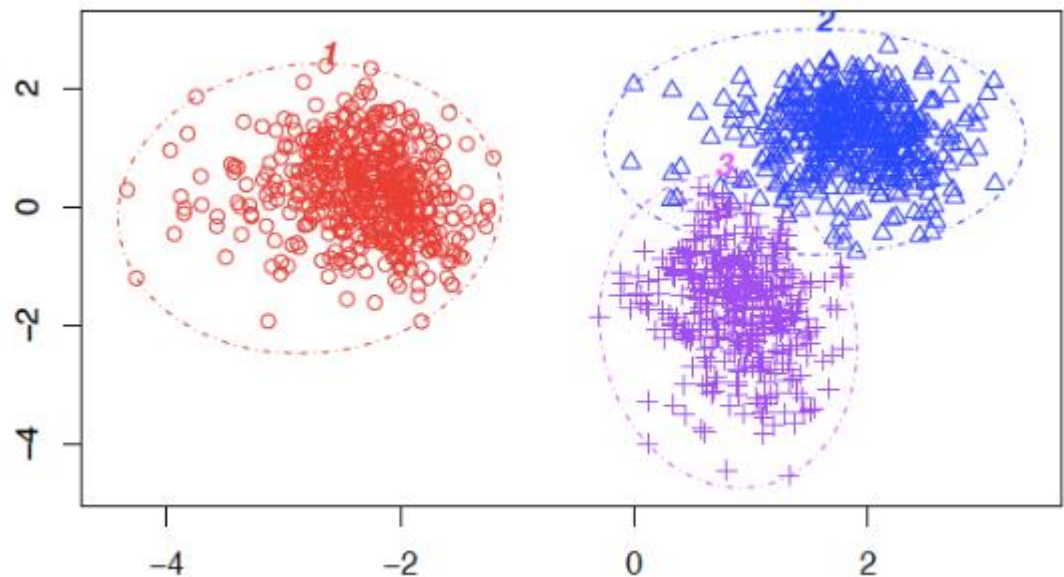
Часто задачи ИАД носят вспомогательный характер при решении задач машинного обучения.

Однако они могут иметь также и самостоятельное значение.

ИАД – кластерный анализа

Под кластерным анализом понимается выделение в данных групп похожих объектов.

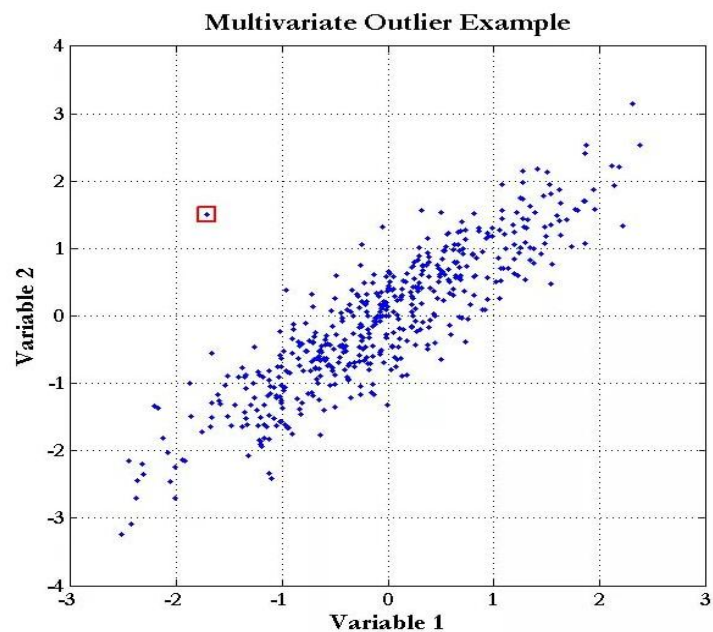
При этом не выделяется какая-либо переменная, которую нужно предсказывать.



ИАД – поиск выпадающих наблюдений

Одним из важных инструментов ИАД является поиск выпадающих объектов (outliers) то есть наблюдений, отклоняющихся от основной части данных или от основной закономерности, существующей в данных

Выпадающие объекты нередко могут являться следствием ошибок, возникающих при заполнении баз данных.

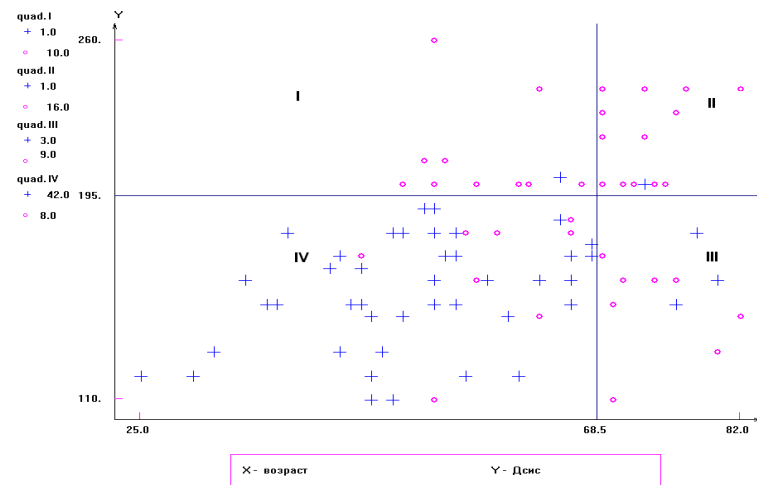


ИАД – поиск информативных предсказывающих переменных

Выбор действительно информативных переменных является важнейшей составляющей успеха при решении задач распознавания и регрессионного анализа. При этом новые переменные могут искаться не только среди существующих переменных, но также и заново рассчитываться по последним.

Поиск статистически достоверных закономерностей в данных

- Поиск статистически достоверных закономерностей, связывающих предсказываемую переменную Y с отдельными X переменными или с их сочетаниями .



Компьютерное зрение

Методы машинного обучения в настоящее время являются важнейшим инструментом, используемым при создании систем компьютерного зрения.

МЕДИЦИНА

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

МЕДИЦИНА

Методы машинного обучения могут быть использованы

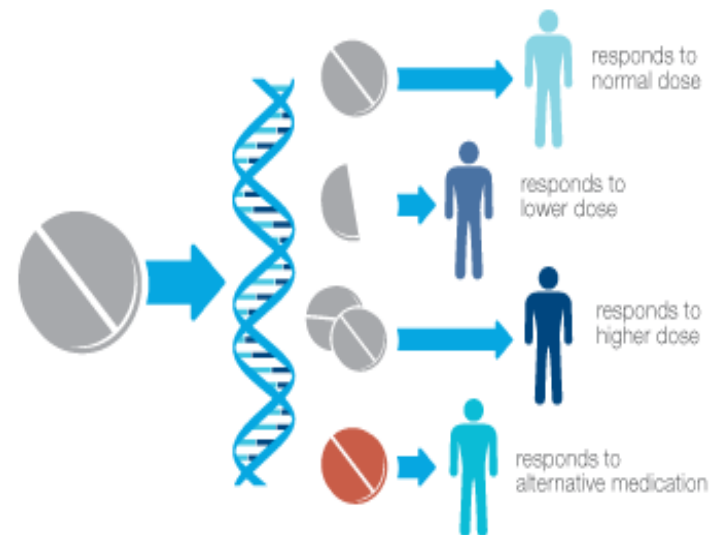
- для решения задач диагностики
- для решения задачи оценивания индивидуального риска развития патологии , возникновения рецидива
- для решения задачи оценивания риска неблагоприятного исхода

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.



ФИЗИКА

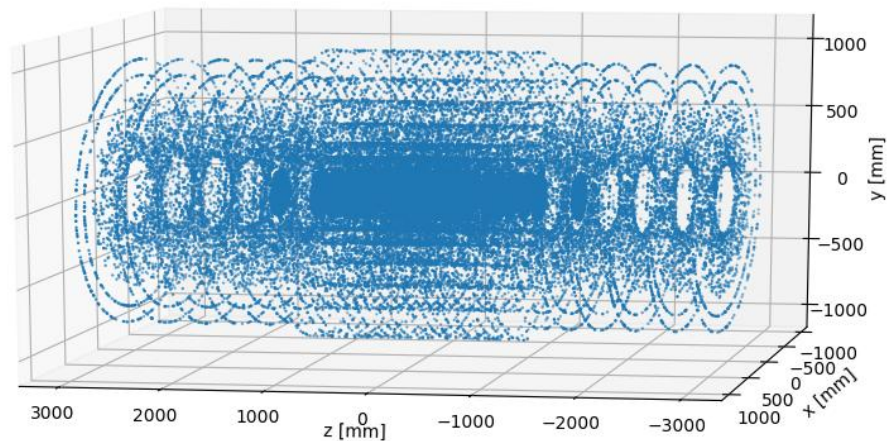
Современные крупные физические эксперименты связаны накоплением больших объёмов информации.

Методы машинного обучения могут быть использованы

- для решения задач выбора наиболее адекватных физических моделей;
- для решения задач планирования эксперимента, выбора его оптимальных параметров;

Физика высоких энергий

Одной из основных задач машинного обучения при обработке результатов экспериментов на ускорителях элементарных частиц является выделения и классификация траекторий заряженных частиц в массиве точечных измерений.



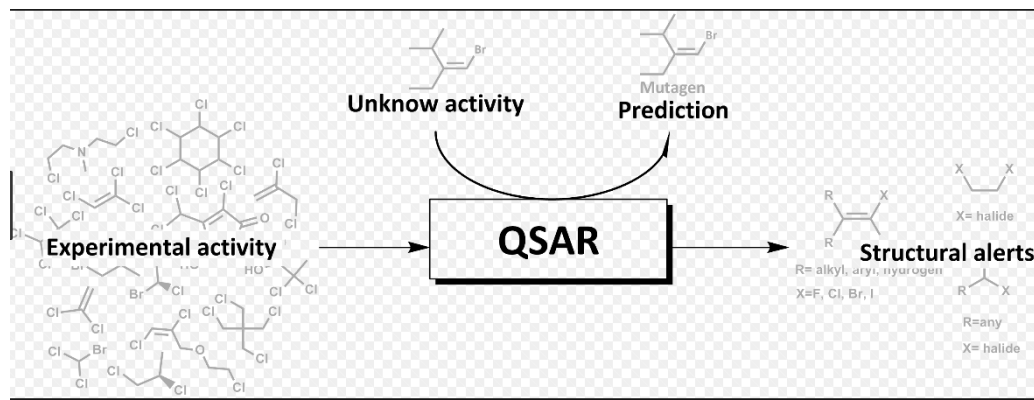
ХИМИЯ

Накоплена информация о свойствах большого числа химических соединений. Данная информация может быть использована для прогнозирования свойств новых ещё не синтезированных соединений по их химическим или структурным формулам.

Наличие эффективных способов прогнозирования позволяет значительно ускорить поиск соединений с заданными свойствами.

Прогноз биологической активности химических соединений

Важным направлением прогнозирования свойств химических соединений является прогнозирование биологической активности. За моделями, позволяющими прогнозировать количественные характеристики биологической активности, исторически закрепилось англоязычное название - Quantitative Structure-Activity Relationship (QSAR).



Маркетинговые исследования

Предсказание поведения потребителей в зависимости от меняющихся условий, оценка предпочтений отдельных потребителей по совокупности показателей, включая предшествующую историю покупательской активности.

Финансы. Банковское дело

Задача банковского кредитного скоринга: ранжирование предполагаемых заёмщиков по их кредитоспособности. Выбор оптимальной (максимизирующей доход) стратегии выдачи кредитов по результатам ранжирования.

Финансы. Банковское дело

Распознавание случаев мошенничества с кредитными картами. Действия мошенников распознаются на основе анализа изменений характера платёжной активности владельца карты.

Горнодобывающая промышленность

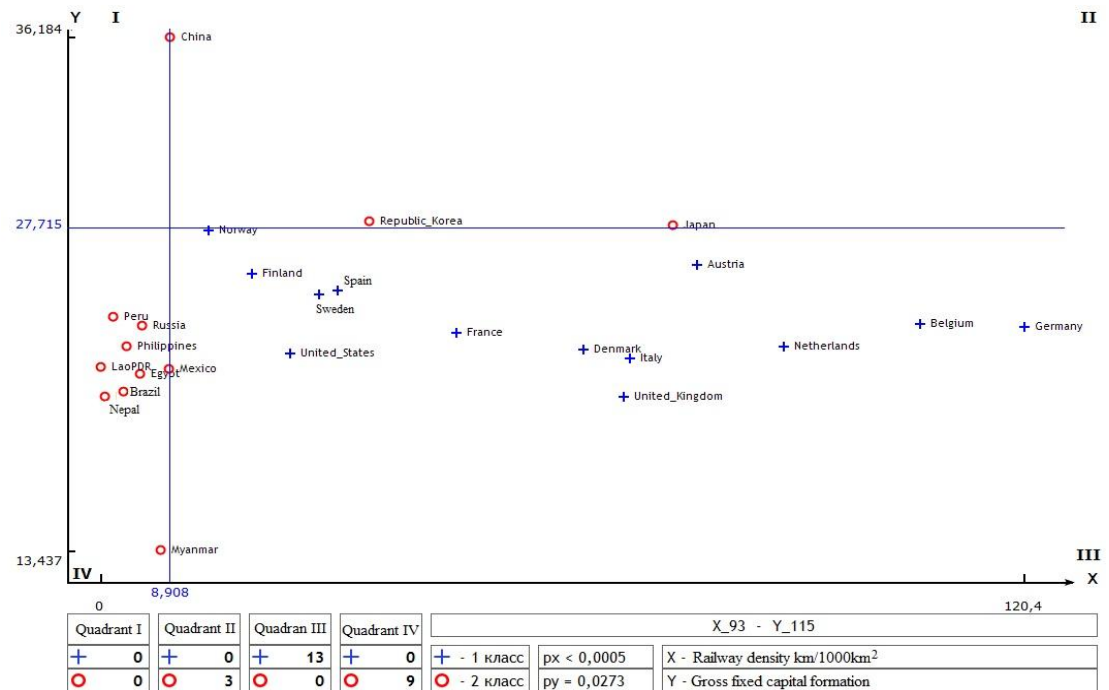
- диагностика состояния горнодобывающего оборудования;
- оценка состояния горных пород

Нефтедобывающая промышленность

- прогноз производительности нефтяных скважин

Социология.

Массивы социологической информации используются для исследования влияния различных факторов на поведение людей, выявить связь между устойчивыми социальными институтами, характером экономического развития, климатическими факторами.



Антропология и этнография

Накоплена информация о встречаемости мифологических сюжетов в фольклорных традициях по всему миру. Использование современных методов анализ данных в приложении

к этой информации позволяет выявить

связь между различными

этнoсами, реконструировать

исторический процесс расселения

народов.

