18 ноября 2021 г.

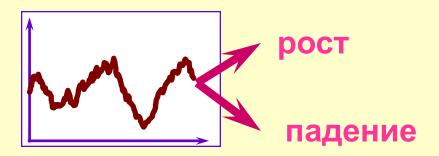
КРУГЛЫЙ СТОЛ МАШИННОЕ ОБУЧЕНИЕ: ЭФФЕКТИВНЫЙ ИНСТРУМЕНТ ДЛЯ ПОЗНАНИЯ ФИЗИЧЕСКОЙ КАРТИНЫ МИРА

Искусственный интеллект, машинное обучение и обработка данных: полезные ископаемые, магнитные бури и электронный нос

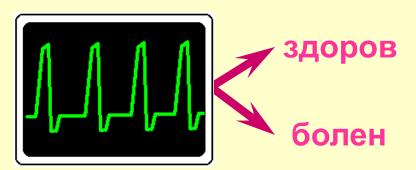
Сергей Анатольевич ДОЛЕНКО
Лаборатория адаптивных методов обработки данных
НИИЯФ МГУ

ПЛОХО ФОРМАЛИЗУЕМЫЕ ЗАДАЧИ

• финансы



• медицинская диагностика



 распознавание речи и текста



• распознавание изображений

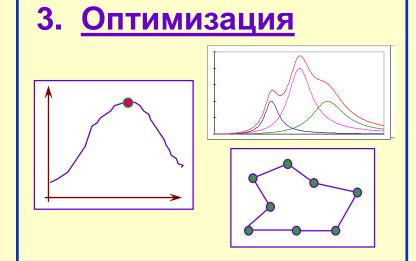


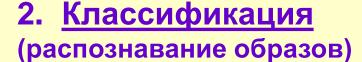
ОСНОВНЫЕ ТИПЫ ЗАДАЧ ОБРАБОТКИ ДАННЫХ

1. Регрессия

- прогнозирование
- оценка
- поиск скрытых закономерностей в данных
- аппроксимация зависимостей

• моделирование
$$y = x_1^3 - 4x_3^2x_8^5 + x_3^2$$

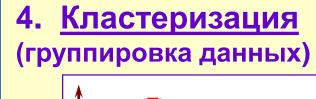


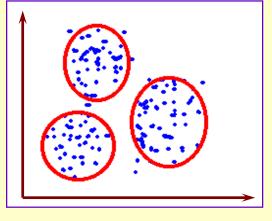


АБВГД









АППРОКСИМАЦИОНЫЕ МЕТОДЫ И ИХ МЕСТО

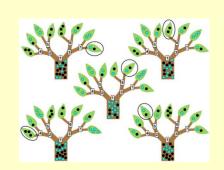
- Большинство рассматриваемых задач представляют собой задачи построения функционального отображения
- Наилучший вариант когда функцию можно задать аналитически на основе физической модели объекта
- Если аналитическое решение невозможно пытаемся решить задачу численно на основе математической модели
- Если содержательных моделей нет остается построение аппроксимационной модели на основе имеющихся примеров
- Аппроксимационная модель = разложение искомого отображения
- Таким образом, «обучение на примерах» = построение аппроксимационной модели, т.е. «черного ящика»
- Если есть возможность, ящик лучше покрасить, т.е. использовать имеющуюся априорную информацию

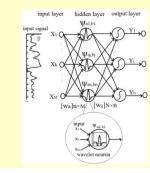
АЛГОРИТМЫ (МЕТОДЫ) МАШИННОГО ОБУЧЕНИЯ

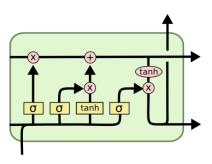
Некоторые наиболее распространенные и наиболее эффективные методы машинного обучения

- Линейная и нелинейная регрессия
- Алгоритмы на основе деревьев решений (случайный лес, градиентный бустинг и др.)
- Искусственные нейронные сети (многослойные, глубокие, свёрточные, рекуррентные, генеративные, вейвлет-нейронные и др.)
- Нейро-нечёткие системы
- Комитеты (ансамбли) алгоритмов

Могут требовать существенных вычислительных затрат при обучении и применении







ЗАКОН МУРА

Гордон Эрл Мур (Gordon Earle Moore), р.1929 "Количество элементов на кристаллах электронных микросхем будет и далее удваиваться каждый год« (1965)

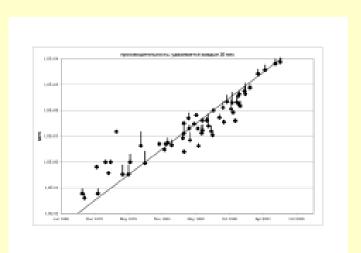


Это эмпирическая закономерность

"Доступная вычислительная мощность удваивается каждые 18 (20) месяцев"

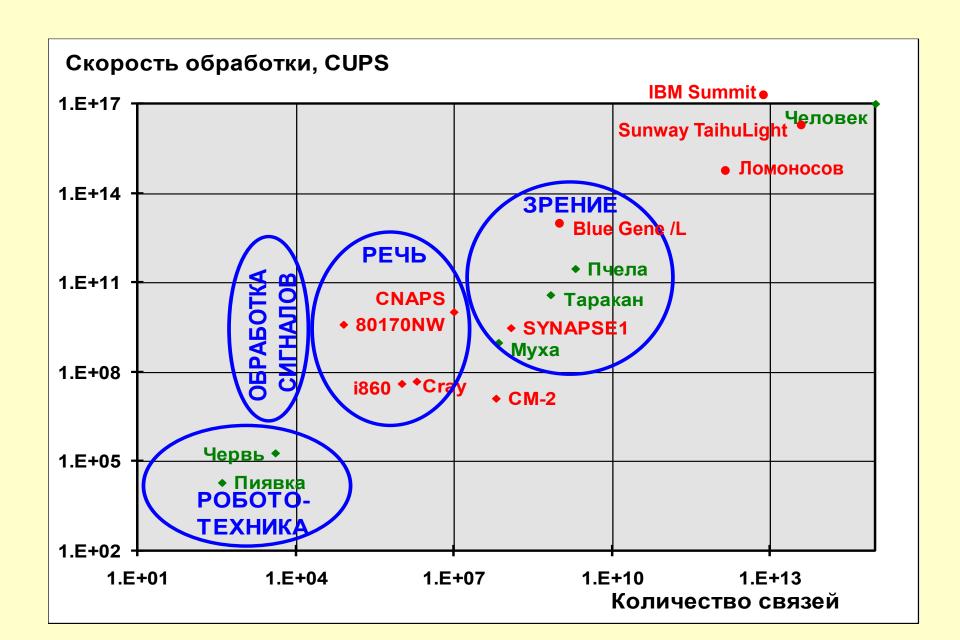
Увеличение производительности на порядок ~ каждые 5 лет

С некоторыми оговорками, закон Мура выполняется уже более 55 лет!



За это время произошло несколько технологических революций

ТЕХНИЧЕСКИЕ ТРЕБОВАНИЯ К ИНС



УРОВНИ ВЗАИМОДЕЙСТВИЯ ИНТЕЛЛЕКТА С ОКРУЖАЮЩИМ МИРОМ

- 1. Персептивный (лат. *perceptio* восприятие)
 - Восприятие окружающего мира
 - Осуществление простейших ответных действий реакция на внешние раздражители
- 2. Когнитивный (лат. cognitio познание)
 - Познание окружающего мира
 - Установление закономерностей
 - Обучение действиям в ситуациях, схожих с известными
- 3. Креативный (лат. creatio творчество)
 - Творческое осмысление и освоение окружающего мира
 - Изобретение нового
 - Умение адекватно действовать в незнакомых ситуациях

На повестке дня – переход с когнитивного на креативный уровень. Узенькие мостики через пропасть уже перекинуты, и они всё шире... Мы уже всерьёз говорим о создании искусственного интеллекта (ИИ)

МАШИННОЕ ОБУЧЕНИЕ И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

Интелле́кт (от лат. intellectus «восприятие»; «разумение», «понимание»; «понятие», «рассудок») или ум — качество психики, состоящее из способности осознавать новые ситуации, способности к обучению и запоминанию на основе опыта, пониманию и применению абстрактных концепций, и использованию своих знаний для управления окружающей средой. Общая способность к познанию и решению проблем, которая объединяет познавательные способности: ощущение, восприятие, память, представление, мышление, воображение.

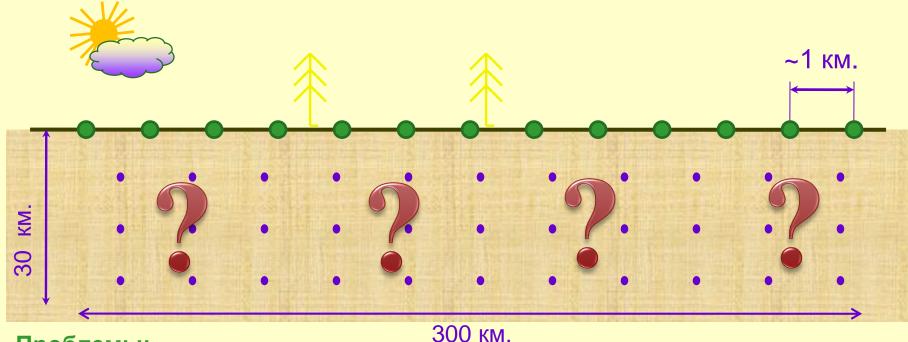
Машинное обучение (МО) обеспечивает «слабый» искусственный интеллект (ИИ), способный решать одну или ограниченный набор задач, которым его обучили. Можно считать, что МО – важнейшая часть ИИ.

Сильный (общий) ИИ – должен обладать креативностью, т.е. широкими когнитивными способностями и возможностью решения задач в нескольких контекстах.

Проблемы сильного ИИ – технологические, этические, юридические...

ПРИМЕР 1: ПОЛЕЗНЫЕ ИСКОПАЕМЫЕ

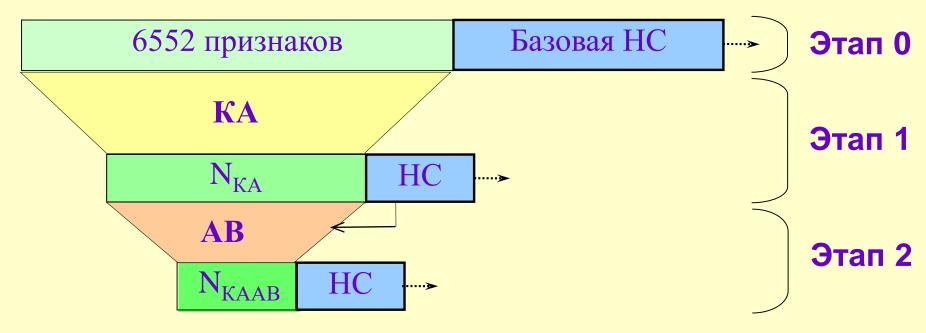
Обратная задача электроразведки: определение распределения проводимости в толще земли на основе данных об электрическом и магнитном поле на поверхности земли.



Проблемы:

- Исключительно высокая размерность входных данных 126 пикетов * 13 частот * 4 компоненты поля = 6552 входа
- Очень высокая размерность выходных данных (233-336 выходов)
- Сложность разработки физически осмысленных классов
- Высокая вычислительная стоимость решения прямой задачи

РЕШЕНИЕ ОБРАТНОЙ ЗАДАЧИ ЭЛЕКТРОРАЗВЕДКИ: КОМПРЕССИЯ ВХОДНЫХ ДАННЫХ

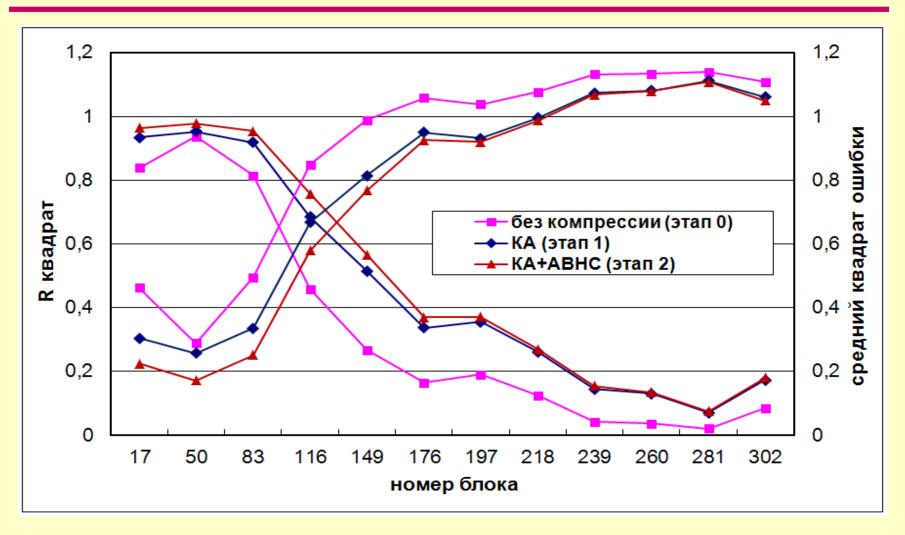


Общая схема алгоритма с корреляционным анализом (КА)

Компрессия (сжатие) входных данных осуществляется в два этапа

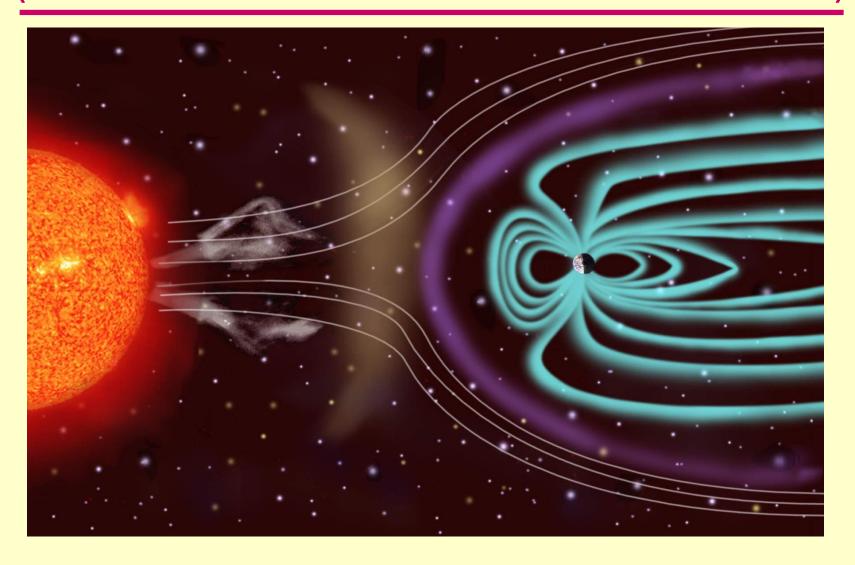
НС – нейронная сеть, АВ – анализ весов нейронной сети

РЕШЕНИЕ ОБРАТНОЙ ЗАДАЧИ ЭЛЕКТРОРАЗВЕДКИ: АЛГОРИТМЫ КОМПРЕССИИ



Результаты последовательных этапов компрессии для алгоритма с корреляционным анализом

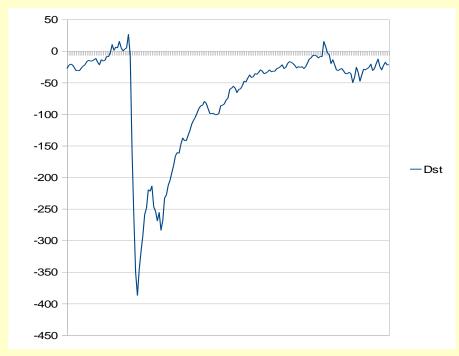
ПРИМЕР 2: МАГНИТНЫЕ БУРИ (ПРОГНОЗ СОСТОЯНИЯ МАГНИТОСФЕРЫ ЗЕМЛИ)

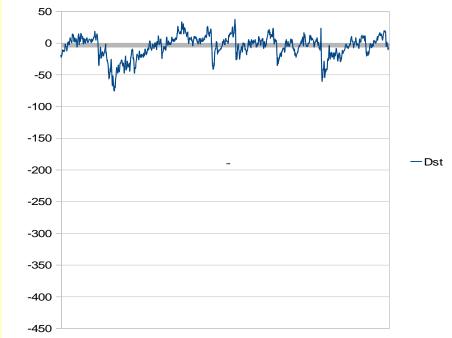


Воздействие солнечного ветра на магнитосферу Земли

ИНДЕКС Dst

Индекс Dst – максимальное отклонение вариации магнитного поля от спокойного уровня, усредненное по значениям, измеренным на контрольной цепочке расположенных в низких широтах магнитных станций





Характерное развитие магнитной бури (3 суток)

Характерное поведение индекса Dst за месяц

Прогнозируется среднечасовое значение индекса Dst на 1-6 часов вперёд

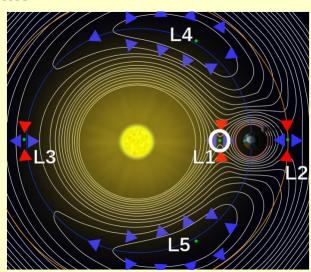
источники и подготовка данных

Используются временные ряды (BP) среднечасовых значений следующих величин:

- <u>Параметры вектора межпланетного магнитного поля (ММП)</u> в точке Лагранжа L1 между Землёй и Солнцем:
 - B_x, B_y, B_z (x-, y- и z-компоненты ММП);
 - амплитуда В (модуль ММП) (нТл);
- Параметры солнечного ветра (СВ)

в той же точке Лагранжа L1:

- скорость CB V (км/с);
- плотность протонов в СВ n_p (см⁻³);
- температура СВ Т (К);
- <u>Геомагнитные индексы</u>:
 - экваториальный геомагнитный индекс Dst (нТл);
 - глобальный геомагнитный индекс Кр (безразмерный);
- ✓ Учёт предыстории всех BP на глубину в 24 часа.

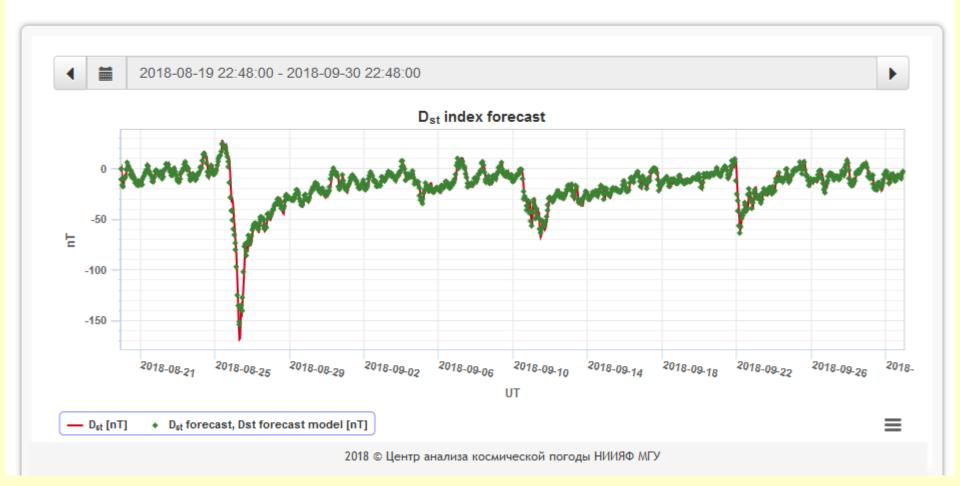


ПРИМЕР ПРОГНОЗИРОВАНИЯ: ГЕОМАГНИТНАЯ БУРЯ



ОНЛАЙН ПРОГНОЗИРОВАНИЕ Dst ИНДЕКСА



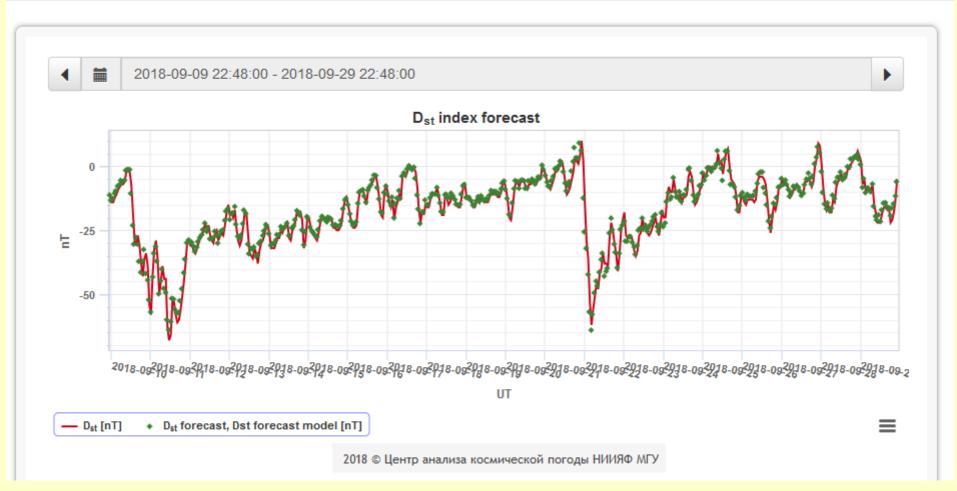


http://swx.sinp.msu.ru

Горизонт прогноза: 1 час

ОНЛАЙН ПРОГНОЗИРОВАНИЕ Dst ИНДЕКСА (2)



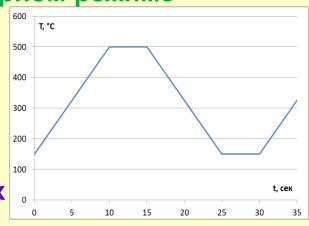


http://swx.sinp.msu.ru

Горизонт прогноза: 1 час

ПРИМЕР 3: ЭЛЕКТРОННЫЙ НОС

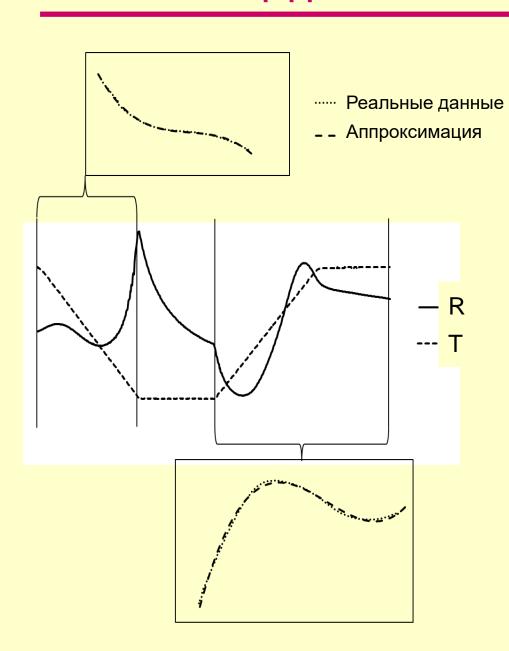
- Классификация измерений полупроводникового газового сенсора
- Сопротивление сенсора R зависит от температуры газа Т и от концентраций примесей в воздухе
- Селективное определение СО, H₂ и их смесей в воздухе (моделирует тление на ранней стадии возгорания)
- Единственный сенсор на основе оксида металла SnO₂ (МОХ-сенсор)
- Концентрация СО от 1 ppm до 50 ppm
- Обычный воздух с улицы мегаполиса Москвы летом
- Сенсор работает в динамическом температурном режиме
- Температура сенсора изменяется
 в диапазоне от 150 до 500 °С,
 линейный нагрев и охлаждение,
 изотермические плато длиной 5 сек при
 максимальной и минимальной температурах



ХАРАКТЕРИСТИКИ ДАННЫХ

- 30 циклов измерений сопротивления сенсора для каждой концентрации
- Первые 10 циклов отбрасывались как переходные между концентрациями, использовались 20 циклов
- Около 600 измеренных значений на каждый цикл измерений,
 что соответствует 600 исходным входным признакам
- Это слишком много: количество весовых коэффициентов сравнимо с количеством примеров в тренировочном наборе
- Следовательно, необходимо эффективное понижение входной размерности данных
- Предлагаемая процедура извлечения признаков позволила описать каждый цикл с помощью 8 коэффициентов

ПРОЦЕДУРА ИЗВЛЕЧЕНИЯ ПРИЗНАКОВ



- Стадии охлаждения и нагрева обрабатывались раздельно
- Использовались функция log10 и простая нормализация для приведения X и Y к одинаковым диапазонам
- Затем использовалась аппроксимация полиномами 3-й степени
- Результаты аппроксимации:
 вверху стадия охлаждения,
 внизу стадии нагрева и
 высокотемпературного плато
- Точки наблюдаемые данные,
 пунктир аппроксимация
- В качестве извлечённых признаков использовались коэффициенты полиномов (всего 8 коэффициентов)

НАИЛУЧШИЕ РЕЗУЛЬТАТЫ КЛАССИФИКАЦИИ

Алгоритм	Точность классификации
Линейная модель	78 %
Многослойный персептрон	91 %
Нейро-нечёткая система	89 %

В таблице представлены наилучшие результаты, полученные в процессе численных экспериментов с различными значениями параметров алгоритмов.

Нейро-нечёткая система – перспективный подход для решения задач классификации с высокой входной размерностью, высокой вариабельностью данных и высоким уровнем шумов

КУРСЫ ПОВЫШЕНИЯ КВАЛИФИКАЦИИ ПРИ МГУ

- Тема курса: «Машинное обучение. Искусственные нейронные сети и генетические алгоритмы»
- Курс проводится с конца февраля по начало июня
- 2 занятия в неделю в вечернее время с 19 до 22
- Лекции и практические занятия, зачёт
- Возможность онлайн-участия. Возможность очного участия (в МГУ) будет зависеть от ситуации
- Окончившим курсы слушателям выдаётся удостоверение установленного образца с гербовой печатью МГУ имени М.В.Ломоносова
- http://kpk-nnga.sinp.msu.ru/
- kpk_nnga@sinp.msu.ru



Спасибо за внимание!

