

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Локтионова Оксана Геннадьевна
Должность: проректор по учебной работе
Дата подписания: 12.10.2023 13:28:42
Уникальный программный ключ:
0b817ca911e6668abb13a5d426d39e5f1c11eabbf73e943df4a4851fda56d089

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Юго-Западный государственный университет»
(ЮЗГУ)

Кафедра вычислительной техники

УТВЕРЖДАЮ
Проректор по учебной работе
О.Г. Локтионова
« 5 » 10 2023 г.



АЛГОРИТМЫ МАШИННОГО ОБУЧЕНИЯ

Методические указания по выполнению практической работы
для студентов направления подготовки 09.04.01 по дисциплине
«Системы искусственного интеллекта»

Курск 2023

СОДЕРЖАНИЕ

Теоретическая часть	4
Задание к практической работе	7
Контрольные вопросы	8

Теоретическая часть

Классические задачи искусственного интеллекта, решаемые с помощью машинного обучения:

- Классификация, как правило, выполняется с помощью обучения с учителем на этапе собственно обучения.
- Кластеризация, как правило, выполняется с помощью обучения без учителя.
- Регрессия, как правило, выполняется с помощью обучения с учителем на этапе тестирования, является частным случаем задач прогнозирования.
- Понижение размерности данных и их визуализация выполняется с помощью обучения без учителя.
- Восстановление плотности распределения вероятности по набору данных.
- Одноклассовая классификация и выявление новизны.
- Построение ранговых зависимостей.
- Обнаружение аномалий.

Типы входных данных при обучении:

- Признаковое описание объектов или матрица объекты-признаки - наиболее распространённый случай. Каждый объект описывается набором признаков.
- Матрица расстояний между объектами. Каждый объект описывается расстояниями до всех остальных объектов обучающей выборки, чаще всего отношениями попарного сходства.
- Временной ряд или сигнал. Последовательность измерений во времени, которое может представляться числом, вектором, а в общем случае - признаковым описанием в данный момент времени.
- Изображение или видеоряд.
- Обработка естественного языка.

Типы функционалов качества:

- При обучении с учителем – функционал качества может определяться как средняя ошибка ответов. Предполагается, что искомый алгоритм должен его минимизировать. Для

предотвращения переобучения в минимизируемый функционал качества часто в явном или неявном виде добавляют регуляризатор.

- При обучении без учителя – функционалы качества могут определяться по-разному, например, как отношение средних межкластерных и внутрикластерных расстояний.
- При обучении с подкреплением – функционалы качества определяются физической средой, показывающей качество приспособления агента.

Практические сферы применения

Целью машинного обучения является частичная или полная автоматизация решения сложных профессиональных задач в самых разных областях человеческой деятельности.

Машинное обучение имеет широкий спектр приложений:

- распознавание речи;
- распознавание жестов;
- распознавание рукописного ввода;
- распознавание образов;
- техническая диагностика;
- медицинская диагностика
- прогнозирование временных рядов;
- биоинформатика;
- обнаружение мошенничества;
- обнаружение спама;
- категоризация документов;
- биржевой технический анализ;
- финансовый надзор;
- кредитный скоринг;
- прогнозирование ухода клиентов;
- хемоинформатика;
- обучение ранжированию в информационном поиске.

Сфера применений машинного обучения постоянно расширяется. Повсеместная информатизация приводит к накоплению огромных объёмов данных в науке, производстве, бизнесе, транспорте, здравоохранении. Возникающие при этом задачи прогнозирования, управления и принятия решений часто сводятся к обучению по прецедентам. Раньше, когда таких

данных не было, эти задачи либо вообще не ставились, либо решались совершенно другими методами.

Кластеризация является важным процессом для различных организаций, так как она помогает в выполнении многочисленных мероприятий и задач. Предположим, что вы являетесь частью аналитической группы компании. Вам необходимо понять, как потенциальные клиенты реагируют на продукт и услуги, которые вы предоставляете. Но для этого требуются ценные данные, которые помогают понять поведение клиентов. Эти данные помогут вам получить лучший опыт и достичь бизнес-целей. Именно здесь концепция кластеризации становится существенной.

Кластеризация в машинном обучении.

Кластеризация помогает упорядочить данные по различным группам в зависимости от особенностей. Вы определяете эти особенности в соответствии с точками данных группы. Предположим, вы хотите классифицировать все типы кошек в различные группы, в зависимости от их вида. В этом случае вы разделите различные типы на различные сегменты, такие как снегоступы, персидские, японские бобины и сиамские. Таким образом, Вы сможете четко проанализировать типы кошек, находящихся под разными сегментами. Это условие также относится к различным проблемам машинного обучения. Вы можете разделить проблемы на уникальные категории в зависимости от их сходства. Вы предоставите данные алгоритму машинного обучения, что сделает его неконтролируемым вариантом обучения. Важно помнить о необходимости подачи точных данных, так как алгоритмы кластеризации группируют новые данные в зависимости от предъявляемых требований.

Алгоритмы кластеризации.

Задача кластеризации – это не требующая контроля методика машинного обучения. Ученые, занимающиеся данными, также называют эту технику кластерным анализом, так как она включает в себя аналогичный метод и рабочий механизм. При первом использовании алгоритмов кластеризации, вы должны предоставить большое количество данных в качестве входных. Эти данные не будут содержать никаких пометок. Это увеличит производительность алгоритма и создаст различные группы. Эти группы будут представлять собой кластеры данных, выровненные по сходству. Кластер включает в себя все точки данных, которые имеют общее сходство по своим характеристикам или свойствам. Они так или иначе связаны друг с другом. Вы можете использовать кластеризацию, чтобы обнаружить шаблоны, инжиниринг и формы. При генерации понимание данных, использовать кластеризацию в качестве начального процесса

ЗАДАНИЕ

«Методы кластеризации и понижения размерности».

Используя информационные источники, ответить письменно на вопросы:

- 1) В чем состоит задача кластеризации?
- 2) приведите примеры задач кластеризации.
- 3) В чем суть метода кластеризации с помощью графов?
- 4) Метод K средних.
- 5) Понижение размерности в задачах кластеризации.

«Методы классификации. Naïve Bayes. Random Forest»

Используя информационные источники, ответить письменно на вопросы:

- 1) Поясните назначение сетей Байеса.
- 2) В чем состоит удобство применения сетей Байеса?

«Операционализация моделей машинного обучения»

Используя информационные источники, ответить письменно на вопросы:

- 1) Операционализация машинного обучения.
- 2) Метапоток: основные функции.

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Перечислите алгоритмы машинного обучения.
2. В каких областях применимо машинное обучение.
3. Преимущества и недостатки машинного обучения.
4. Задачи, решаемые при помощи машинного обучения.