

ОСОБЛИВОСТІ НАЛАШТУВАННЯ АЛГОРИТМУ РЕГРЕСІЇ ЛІСУ ПРИЙНЯТТЯ РІШЕНЬ У СЛУЖБИ МАШИННОГО НАВЧАННЯ MICROSOFT AZURE

Служба машинного навчання Azure (Azure Machine Learning) надає ряд послуг щодо розробки і розгортання рішень машинного навчання для прикладних задач. Застосування вбудованого Конструктора дозволяє спростити процес створення, навчання та оцінки моделей. Конструктор – це спеціалізований засіб візуального проектування, котрий дозволяє створювати моделі машинного навчання без єдиного рядка коду і містить ряд розроблених модулів машинного навчання. Проте спрощення побудови моделі ще не означає отримання якісної і точної моделі для рішення конкретних прикладних завдань. Адже кожна модель має ряд параметрів від налаштувань яких і залежить точність моделі.

Розглянемо особливості налаштування модулю алгоритму регресії лісу прийняття рішень. Основними етапами створення моделі машинного навчання на основі цього алгоритму з використанням конструктора студії машинного навчання є:

- Етап 1. Створення робочої області Azure.
 - Етап 2. Створення робочих ресурсів та конвеєру.
 - Етап 3. Імпортування вхідних даних для навчання.
 - Етап 4. Підготовка даних для навчання.
 - Етап 5. Налаштування та навчання моделі.
 - Етап 6. Оцінка якості моделі машинного навчання.
 - Етап 7. Розгортання моделі у вигляді кінцевого продукту прогнозування в реальному часі.
- Безпосереднє використання модуля і його налаштування відбувається на п'ятому та шостому етапах.

Причому цей процес може мати декілька ітерацій для покращення якості моделі.

Модель регресії створюється на основі ансамблів дерев прийняття рішень. Дерева прийняття рішень - це непараметричні моделі, які виконують послідовність простих тестів для кожного екземпляра, виконуючи обхід деревовидної структури двійкових даних до досягнення кінцевого вузла (рішення).

Дана модель регресії складається з сукупності дерев прийняття рішень. Кожне дерево в лісі рішення регресії виводить розподіл по Гаусу у вигляді прогнозу. Статистична обробка виконується по ансамблю дерев, щоб знайти розподіл по Гаусу, найближчий до об'єднаного розподілу для всіх дерев в моделі.

Після додавання модуля в конвеєр налаштовуються його властивості. Спочатку необхідно обрати метод для створення окремих дерев. Це може бути багтінг або реплікація.

При багтінгу (або бутстрап-агрегуванню) кожне дерево в лісі рішення регресії виводить гаусів розподіл шляхом прогнозування. Агрегування полягає в тому щоб знайти гаусіан, перші два моменти якого збігається з моментами суміші гаусових розподілів, отриманих шляхом об'єднання всіх розподілів, що повертаються окремими деревами.

При реплікації кожне дерево навчається на одних і тих же вхідних даних. Визначення того, який предикат розбиття використовується для кожного вузла дерева, залишається випадковим, і дерева будуть різними.

Модель може бути попередньо навчена у «режимі інструктора» де можна задати параметри чи їх діапазони, якщо є апріорна інформація.

Далі задаються число дерев прийняття рішень в ансамблі та максимальна глибина дерев прийняття рішень. Потім призначається число випадкових розбиттів на вузол та мінімальне число вибірок на кінцевий вузол. Лише після цього до можна переходити до навчання моделі.

Для оцінки якості моделі регресії застосовуються такі метрики:

Середнє значення «абсолютна помилка» усереднює всі помилки в моделі, де помилка означає відстань прогнозованого значення від істинного значення.

Корінь середньої квадратної помилки вимірює середнє значення квадратів помилок, а потім приймає корінь цього значення.

Відносна абсолютна помилка представляє помилку в процентах від істинного значення.

Відносне квадратичне повідомлення про помилку нормалізує загальну квадратичну помилку шляхом ділення на загальну квадратичну помилку прогнозованих значень.

Коефіцієнт визначення - це число, яке вказує, наскільки добре дані вписуються в модель. Значення 1 означає, що модель точно відповідає даним. Нульове значення означає, що дані є випадковими чи іншим чином не можуть відповідати моделі.

Тільки за рахунок правильних налаштувань можна досягти якісного навчання моделі.