

В останні роки машинне навчання (ML - machine learning) перетворилось, у великий бізнес - фірми використовують його, щоб заробити грошей, прикладні дослідження бурхливо розвиваються як в індустріальному, так і в академічному середовищі, а цікаві розробники скрізь шукають можливість підняти свій рівень володіння темою. Але виник попит, що набагато перевищує швидкість появи хороших методик для вивчення застосовуваних практично технік. Цей курс покликаний задовольнити цей попит.

Прикладне машинне навчання поєднує в собі рівні частки математичних принципів і отриманих емпіричним шляхом прийомів, - іншими словами, це справжнє мистецтво. Занадто сильна концентрація тільки на одному з цих аспектів на шкоду іншому – програшна стратегія. Тут важливий баланс.

Довгий час найліпшим і єдиним способом розуміння машинного навчання було здобуття наукового ступеня в одній з областей, які (здебільшого незалежно один від одного) розвивали статистичні методи та техніки оптимізації. Основний акцент ці програми робили на ключові алгоритми, у тому числі на їх теоретичні властивості і обмеження, а також на характерні особливості тих задач, що відносяться до даної сфери. Втім, паралельно не менш цінні знання накопичувалися неофіційним чином - у процесі неформального спілкування на конференціях, обміну інформацією та сценаріями обробки даних між колегами з дослідницьких лабораторій. Саме ці знання, за великим рахунком, і дозволили встановити, які алгоритми найбільше підходять у кожній ситуації, як обробляти дані на кожному етапі і як зв'язати між собою різні етапи робочого процесу.

Зараз ми живемо в епоху відкритого вихідного коду, з готовими до використання високоякісними алгоритмами, доступними на сайті GitHub, і універсальними, добре спроектованими фреймворками, що дозволяють пов'язати усі фрагменти один з одним. Але навіть серед цього достатку неофіційні практичні знання незмінно залишаються недоступними. В цьому курсі ця інформація зводиться воедино саме цього фрагменту не вистачає для перетворення машинного навчання з академічної дисципліни у набір знань та навичок із проектування програмного забезпечення.

Варто наголосити на ще одному моменті. Більшість широко використаних методів машинного навчання далекі від досконалості і містять ряд прогалин.

Сучасні методи і мають високі вимоги до даних. Вони, за великим рахунком, раді забезпечити нас надмірно впевненими передбаченнями і, якщо не прийняти необхідних заходів. І невеликі зміни вхідних даних можуть призвести до великих і дивних змін у виявлених шаблонах.

Отримані результати часом складно інтерпретувати і дослідження сучасної інженерії машинного навчання може розглядатись як вправа на згладжування цих (та інших) гострих кутів у методах оптимізації та статистичного навчання.