

Машинное обучение: начало

И. Куралёнок

СПб, 2017

Что нужно, чтобы понять?

- ТВ и МС
- Линейная алгебра
- Язык программирования

Как отчитываться?

- Экзамен как в универе
- Домашняя работа, которая дает бонусы на экзамене
- Статья с “хорошей” конференции/журнала

Какие у нас цели?

- Уметь сформулировать задачу в терминах ML
- Найти подходящий класс решающих алгоритмов по формулировке
- Ориентироваться в области и знать “где посмотреть” существующие решения
- Понимать границы применимости

Что почитать?

- Википедия (лучше en)
- Т. Hastie, R. Tibshirani, J. Friedman “The elements of Statistical Learning”
- Т. Mitchell “Machine Learning”
- Труды конференций: NIPS, ICML, CIKM, KDD, etc.
- Журналы: JML, JMLR, JIS, NC, etc.
- Видео курсы (нет им числа)

Чтобы решить задачу обучения

Нужно:

Чтобы решить задачу обучения

Нужно:

- 1 Скачать данные
- 2 Придумать метрику
- 3 Запустить XGBoost

Чтобы решить задачу обучения

Нужно:

- 1 Придумать цель и найти данные
- 2 Обеспечить наблюдаемость цели в рамках описания
- 3 Найти класс решающих функций подходящий для задачи
- 4 Составить целевую функцию, близкую к предметной цели и “простую в оптимизации”
- 5 Провести оптимизацию и убедиться в том, что ответ достигает поставленных целей

Чего (пока) не умеет ML

Зачем мы все еще нужны :)

- 1 Ставить задачу и разбивать ее на этапы
- 2 Моделировать ненаблюдаемые ситуации
- 3 Переиспользовать накопленные знания для решения новых задач
- 4 Понять полезность результата

Data Mining vs. Machine Learning

	Data Mining	Machine Learning
Цель дисциплины	Выявление “скрытых данных”	Оптимизация целевой функции
Исследования	Больше про данные	Больше про методы
Типичный результат	“Мы применили такой метод и получили клевые результаты на таких стандартных данных”	“Предложили новый метод, который работает круче чем другие на нескольких датасетах (возможно даже синтетика)”
Где почитать	SIGIR, WSDM, WWWC, ...	ICML, CIKM, ...

Artificial Intelligence vs. Machine Learning

	Artificial Intelligence	Machine Learning
Цель дисциплины	Рациональное поведение умных машин	Оптимизация целевой функции
Исследования	Больше про мат. моделирование	Больше про методы
Типичный результат	“Мы придумали как формализовать задачу игры в шахматы, применили такие методы и обыграли человека”	“Предложили новый метод, который работает круче чем другие на нескольких датасетах (возможно даже синтетика)”
Где почитать	AAAI, IJCAI, ...	ICML, CIKM, ...

Машинное обучение: определения (I)

A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E .

Tom M. Mitchell

Машинное обучение: определения (II)

Machine learning — the ability of a machine to improve its performance based on previous results.

Webster

Машинное обучение: определения (III)

Машинное обучение — обширный подраздел искусственного интеллекта, изучающий методы построения алгоритмов, способных обучаться.

ru.wikipedia.org

Машинное обучение: немного истории

- 50-70гг базы знаний, полнотекстовый поиск, распознавание образов, нейронные сети
- 70-80гг символьный вывод, Quinlan ID3 деревья, разумные практические результаты, VC-оценки
- 80-90гг первые конференции, много практического применения, активное применение кластеризации в анализе
- 90-00гг повторное сэмплирование в ML, SVM, применение в IR, ML \neq DM, LASSO, bootstrap, bagging, boosting
- 00-10гг Compressed sensing и прочие восстановления сигналов, царство деревьев, развитие ансамблей, ...
- 10- Deep learning ...

Основные понятия

- Область работы = Universe = Γ .
- Решающая функция = Decision Function = $\hat{F} \in F$
– класс решающих функций.
- Опыт = Data Set = $D = X \times Y$.
- Целевая функция = Target = $T(y, F(x))$.

Задача обучения

В ML оптимизация часто проводится в одних условиях, а эксплуатация в других.

$$\arg \max_{F, B: \hat{F}=B(F)} A(\Gamma, \hat{F})$$

- A — цели эксплуатации (например деньги) на всей области работы Γ
- B — способ оптимизации, который реализуем

Как устроена эксплуатация

На самом деле, как она _может_ быть устроена

Будем считать, что работа на разных элементах Γ независима. Тогда эксплуатацию A можно представить так:

$$A = \mu_{x \sim U(\Gamma)} T_A(\hat{F}(x))$$

Как устроена оптимизация

На самом деле, как она _может_ быть устроена

Способ оптимизации B , в свою очередь, такая штука:

$$\hat{F} = B(F, D) = \arg \max_F T_B(D, F) = \arg \max_F T_B(Y, F(X));$$

В рамках этой лекции, мы не будем рассматривать разницу между T_A и T_B , поэтому будем считать, что они одинаковы T .

Классификация машинного обучения

ML можно делить по:

- способу получения опыта;
- виду целевой функции;
- классу решающих функций.

Классификация машинного обучения

ML можно делить по:

- **способу получения опыта;**
- виду целевой функции;
- классу решающих функций.

Классификация машинного обучения: ОПЫТ

- Transductive learning
- Обычное обучение
- Стохастическая оптимизация (stochastic optimization)¹
- Активное обучение (active learning)
- Обучение с бюджетом (budget learning)
- Интерактивное обучение (online learning)
- Многорукие бандиты (multi-armed bandits)
- Обучение с подкреплением (reinforcement learning)

¹По сути это не обучение, но очень похоже.

Transductive learning

- 1 Все множество, на котором работаем задано перечислением Γ_0
- 2 Для части данных известен ответ, и они формируют D

$$\hat{F} = B(F, D, \Gamma_0)$$
$$A(\Gamma_0, \hat{F})$$

Обычное обучение

- 1 Фиксируем множество примеров X
- 2 Определяем генеральную совокупность Γ
- 3 Обучаемся на доступных примерах, используя информацию о всех

$$\hat{F} = B(F, D)$$
$$A(\Gamma, \hat{F})$$

Стохастическая оптимизация²

- 1 Определяем генеральную совокупность Γ
- 2 Обучаемся на следующем примере пока не надоест (до момента T)

$$D_0 = \emptyset$$

$$D_{t+1} = D_t \cup (x_t \sim U(\Gamma))$$

$$\hat{F}_t = B(F, D_t)$$

$$A(\hat{F}_T, \Gamma)$$

²Одна из возможных версий предмета :).

Активное обучение

- 1 Определяем генеральную совокупность Γ
- 2 Обучаемся на всех доступных примерах
- 3 Пополняем множество примеров по просьбе алгоритма \mathcal{A} и переходим к п. 2, если \mathcal{A} не требует больше данных останавливаемся в T

$$D_{t+1} = D_t \cup \{\mathcal{A}(D_t, \hat{F}_t)\}$$

$$\hat{F}_t = B(F, D_t)$$

$$A(\hat{F}_T, \Gamma)$$

Обучение с бюджетом

- 1 Введем стоимость получения информации по точке $c(d)$ и бюджет обучения B
- 2 Определяем генеральную совокупность Γ
- 3 Обучаемся на всех доступных примерах
- 4 Пополняем множество примеров по просьбе алгоритма \mathcal{A} , пока не закончился бюджет $\sum_t c(x_t, y_t) < C$ и переходим к п. 3

$$D_{t+1} = D_t \cup \{\mathcal{A}(D_t, \hat{F}_t)\}$$
$$\hat{F}_t = B(F, D_t)$$
$$A(\hat{F}_T, \Gamma)$$

Интерактивное обучение (online)

- 1 Определяем генеральную совокупность Γ
- 2 Обучаемся на всех доступных примерах
- 3 Получаем следующую точку, исходя из работы решающей функции и переходим к п. 2

$$A(\hat{F}, \Gamma) = \sum_t A(\hat{F}_t, x_t)$$

$$D_0 = \emptyset$$

$$D_{t+1} = D_t \cup \{(x_{t+1})\}$$

$$\hat{F}_t = B(F, D_t)$$

Многорукие бандиты

- 1 Фиксируем множество возможных действий M , $m \in M$.
- 2 Каждое действие ведет к ответу $y \sim \xi(x)^*$.
- 3 Получаем следующую точку x_t
- 4 Обновляем модель $\hat{F}_t = \{\xi_{tm}, m \in M$.
- 5 Выбираем следующее действие исходя из алгоритма $A(\hat{F}_t, x)$.
- 6 Получаем по действию ответ среды
- 7 Повторяем с п.3 пока не надоест.

$$\begin{aligned}A(\hat{F}, \Gamma) &= \sum_t A(y_t = y \sim \xi_{m_t=A(\hat{F}_t, x_t)}, x_t) \\D_0 &= \emptyset \\D_{t+1} &= D_t \cup \{((x_{t+1}, y_{t+1}), m_{t+1})\} \\ \hat{F}_t &= B(F, D_t)\end{aligned}$$

Обучение с подкреплением³

- 1 Фиксируем множество возможных действий M , “поле” X , множество возможных ответов среды $S = S^o \cup S^n$.
- 2 Получим наблюдаемую часть ответа среды на текущее состояние.
- 3 Обновим модель предсказания $\hat{F}_t = S(m)$
- 4 Сделаем следующий ход исходя из алгоритма $\mathcal{A}(\hat{F}_t, s^o)$.
- 5 Повторяем с п.2 пока не надоест или среда нас не убьет.

$$A(\hat{F}, M, X, S) = \sum_t A(S_t(m_t = \mathcal{A}(\hat{F}_t, x_t)), S_{i_1}^t)$$

$$D_0 = \emptyset$$

$$D_{t+1} = D_t \cup \{((x_{t+1}, y_{t+1} \in S^o), m_{t+1})\}$$

$$\hat{F}_t = B(F, D_t)$$

³На самом деле, это уже AI.

Классификация машинного обучения

ML можно делить по:

- способу получения опыта;
- **виду целевой функции;**
- классу решающих функций.

Классификация машинного обучения: цель

- С учителем
 - классификация (classification);
 - аппроксимация (regression);
 - отношение порядка (learning to rank);
 - обучение метрики (metric learning).
- Без учителя:
 - кластеризация (cluster analysis);
 - уменьшение размерности (dimensionality reduction);
 - обучение отображению (representation learning).
- Смешанные:
 - условная кластеризация;
 - transfer learning.

Обучение с учителем I

Классификация: в качестве метрики LL, KL, CE, etc.

$$y \in \{-1, 1\}, F : \Gamma \rightarrow [0, 1]$$

$$y \in \{1, \dots, m\}, F : \Gamma \rightarrow [0, 1]^m$$

$$y \in \{0, 1, \dots, m\}, F : \Gamma \rightarrow [0, 1]^{m+1}$$

Аппроксимация: варианты MSE. $y \in \mathbb{R}, F : \Gamma \rightarrow \mathbb{R}$

Обучение с учителем II

Отношения порядка: цель зависит от модели, но почти всегда хотим найти $F : \Gamma \rightarrow \mathbb{R}$:

- pointwise: $x \in \Gamma, y \in \mathbb{R}, T$ – MSE;
- pairwise: $x \in \Gamma^2, y \in \{<, =, >\}, T$ – см. классификацию;
- listwise: $x \in \Gamma^n, y \in ?, T$ – специфичен для конкретного применения.

Обучение с учителем III

Обучение метрики: хотим построить такую функцию от пары чтобы:

- Она отражала заданную семантику

$$X = \Gamma^2,$$

$$Y = \{(a, b, c, d) : m(a, b) < m(c, d), a, b, c, d \in \Gamma\},$$

$$F : \Gamma^2 \rightarrow \mathbb{R}$$

- По возможности была метрикой

Целевая функция обычно основывается на классификации

Обучение без учителя I

Уменьшение размерности: надо отобразить исходное пространство в пространство меньшей размерности, максимально сохранив заданные свойства.

- расстояние: $x, y \in \Gamma, \|F(x) - F(y)\| - \|x - y\|$;
- статистику: $\{x_i\}_1^n \in \Gamma^n, \Psi(\{x_i\}_1^n) - \Psi(\{F(x_i)\}_1^n)$;

Кластеризация: это такое уменьшение размерности до “упора”, в качестве статистики, которую надо оставить выступает “чувство прекрасного”. Например:

$$\Psi = \begin{cases} 0, & \|x - y\| < \epsilon \\ 1 & \end{cases}$$

Обучение представлению тоже уменьшение размерности, но ограничения накладываются уже на то как видимы результаты.

Классификация машинного обучения

ML можно делить по:

- способу получения опыта;
- виду целевой функции;
- **классу решающих функций.**

Основные классы решающих функций

- Линейные решения
- Параметрические семейства функций
- Графы
- Нейронные сети (ANN)
- Instance based learning (kNN)
- Предикаты
- Ансамбли

Деление по решающей функции (I)

- Линейные решения
 - Линейная регрессия, логистическая регрессия
 - Скрытый дискриминантный анализ (LDA/QDA*)
 - LASSO
 - SVM
 - LSI*

Деление по решающей функции (II)

- Графы
 - Марковские модели (цепи, HMM)
 - Графические модели
 - Conditional Random Fields
- Нейронные сети (ANN)
 - Персептронные сети
 - Сети Хопфилда++
 - Сети Кохоннена

Деление по решающей функции (III)

- Параметрические семейства функций
 - Сэмплирование
 - Генетические алгоритмы
 - PLSI/LDA (Latent Dirichlet Allocation)/прочие модели с распределениями (несть им числа)
- Instance based learning (kNN)

Деление по решающей функции (IV)

- Предикаты
 - Логические выражения
 - Деревья решений
- Ансамбли
 - Просто ансамбли
 - Bagging
 - Boosting
 - BagBoo/BooBag

Вопросы?

Дедуктивные/индуктивные методы

Индуктивные	Дедуктивные
Полагаются на статистику	Полагаются на prior knowledge
Используют классы элементарных функций	Решающая функция следует из предполагаемой структуры
Работают в любой области	Привязаны к области
Знание области отражается на составлении target	Понимание области меняет решающую функцию
Логистическая регрессия	LDA
<i>Для вхождения в область, при больших размерностях</i>	<i>Небольшие размерности, «давно тут сидим»</i>

Применение методов ML

- Практически везде (дайте задачу, я попробую придумать применение)
- Есть два больших класса работ

	Академические	Практические
Цели	Существуют ситуации, когда работает хорошо	Обеспечивает измеряемое качество на множестве примеров
Искать	Красивые идеи, хорошую математику	Работающие вещи, много грязных приемов
Смотреть	Конференции	Соревнования