Машинное обучение: Лекция 1

Гаврилов Даниил

2017-2018

olavursky@gmail.com
Санкт-Петербургский Государственный Университет
Факультет Прикладной Математики и Процессов Управления

Что из себя представляет машинное обучение

Что из себя представляет машинное обучение

Мы можем собрать очень много разных данных $\{x\}_i^n$ (например большой корпус о продажах квартир в крупных городах, фото кошек, etc).

Что из себя представляет машинное обучение

Мы можем собрать очень много разных данных $\{x\}_i^n$ (например большой корпус о продажах квартир в крупных городах, фото кошек, etc).
Мы можем сказать, что существует некоторый скрытый от нашего взгляда закон, который описывает устройство этих данных. И, более того, мы говорим, что этот закон принимает вид функции вероятности $p(x)$

Что из себя представляет машинное обучение

Мы можем собрать очень много разных данных $\{x\}_i^n$ (например большой корпус о продажах квартир в крупных городах, фото кошек, etc).
Мы можем сказать, что существует некоторый скрытый от нашего взгляда закон, который описывает устройство этих данных. И, более того, мы говорим, что этот закон принимает вид функции вероятности $p(x)$
Задача машинного обучения заключается в том, чтобы по конечному числу наблюдений $\{x\}_i^n$ восстановить этот закон $p(x)$.

Немного о другом

Немного о другом

Представим, что мы решаем одну задачу – нам дано множество чисел

$$C = \{1, .., 100\}$$

Немного о другом

Представим, что мы решаем одну задачу – нам дано множество чисел

$$C = \{1, .., 100\}$$

Мы говорим, что существует некоторое $A \subseteq C$, которое мы назовем положительным классом.

Немного о другом

Представим, что мы решаем одну задачу – нам дано множество чисел

$$C = \{1, .., 100\}$$

Мы говорим, что существует некоторое $A \subseteq C$, которое мы назовем положительным классом.
Нам показывают $D \subseteq A$. Задача заключается в том, что нам необходимо сказать, какие еще числа, помимо тех, что принадлежат D, находятся в A

Немного о другом

Представим, что мы решаем одну задачу – нам дано множество чисел

$$C = \{1, .., 100\}$$

Мы говорим, что существует некоторое $A \subseteq C$, которое мы назовем положительным классом.

Нам показывают $D \subseteq A$. Задача заключается в том, что нам необходимо сказать, какие еще числа, помимо тех, что принадлежат D, находятся в A

Пусть, например, $D = \{1, 2, 16, 64\}$.

Множество гипотез

Введем множество гипотез H, которое будет содержать всевозможные решения задачи (например, гипотеза, что A – это все числа от 1 до 100, только четные числа, или, например, все числа больше 50).

Множество гипотез

Введем множество гипотез H, которое будет содержать всевозможные решения задачи (например, гипотеза, что A – это все числа от 1 до 100, только четные числа, или, например, все числа больше 50).
Подмножество H, в котором будут только те гипотезы, которые не противоречат данным, будем называть множеством вариантов V (В нашем случае, очевидно, гипотеза "все числа больше 50 противоречит нашим данным, а потому нет даже смысла думать о ней).

Правдоподобие

Правдоподобие

Возьмем для рассмотрения две гипотезы "Степени числа 2"и "Четные числа".

Правдоподобие

Возьмем для рассмотрения две гипотезы "Степени числа 2"и "Четные числа".
Заметим, что в контексте множества $D = \{1, 2, 16, 64\}$, первая кажется нам более вероятной, несмотря на то, что обе они принадлежат множеству вариантов.

Правдоподобие

Возьмем для рассмотрения две гипотезы "Степени числа 2"и "Четные числа".
Заметим, что в контексте множества $D = \{1, 2, 16, 64\}$, первая кажется нам более вероятной, несмотря на то, что обе они принадлежат множеству вариантов.
Причина проста – мы пытаемся избежать подозрительных совпадений: если А – это четные числа, то почему мы с вами видим только степени двух?

Чувство прекрасного

Чувство прекрасного

Однако, если мы с вами попытаемся найти $h_{ML} = argmax_h p(D|h)$, то мы обнаружим, что $h_{ML} = D$

Чувство прекрасного

Однако, если мы с вами попытаемся найти $h_{ML} = argmax_h p(D|h)$, то мы обнаружим, что $h_{ML} = D$

Несмотря на правдоподобность этого решения, мы вряд ли будем согласны с тем, что оно верное, поскольку у нас с вами есть чувство прекрасного, заключающееся в том, что $p(h_{ML})$ существенно меньше, чем $p(h_{two})$.

Решение, которое бы нас удовлетворяло, должно находиться в балансе между правдоподобием и приорным распределением на гипотезы.

Баланс

Очевидно, нам нужно искать некий баланс между $p(D|h)$ и $p(h)$.

Баланс

Очевидно, нам нужно искать некий баланс между $p(D|h)$ и $p(h)$.
$p(h|D) = \frac{p(D|h)p(h)}{p(D)}$ является этим балансом.

Баланс

Очевидно, нам нужно искать некий баланс между $p(D|h)$ и $p(h)$.
$p(h|D) = \frac{p(D|h)p(h)}{p(D)}$ является этим балансом.
Однако проблема в том, что мы все еще не решили нашу задачу, поскольку нам нужно выдать $p(x \in A|D)$ для каждого x из C.

Вывод

Мы хотим найти $p(x \in A|D)$.
Собирая всё то, что мы смогли получить по ходу решения задачи, у нас есть:

$p(h|D)$ – Вероятность того, что гипотеза h верна, при условии наблюдения нами данных D.

Вывод

Мы хотим найти $p(x \in A|D)$.
Собирая всё то, что мы смогли получить по ходу решения задачи, у нас есть:

$p(h|D)$ – Вероятность того, что гипотеза h верна, при условии наблюдения нами данных D.
$p(x \in A|h)$ – Вероятность того, что x принадлежит положительному классу A, при условии верности гипотезы h.

Вывод

- Bayes model averaging

$$p(x \in A|D) = \sum_{\hat{h}} p(x \in A|\hat{h})p(\hat{h}|D)$$

Вывод

- Bayes model averaging

$$p(x \in A|D) = \sum_{\hat{h}} p(x \in A|\hat{h})p(\hat{h}|D)$$

- Maximum a posteriori estimation (MAP)

$$p(x \in A|D) = p(x \in A|h_{MAP})$$

$$h_{MAP} = argmax_h p(h|D)$$

Итог

Рассмотрев эту задачу, мы обозначили фреймворк того, как мы будем пытаться использовать опыт из некоторых данных, чтобы выявить закон, согласно которым эти данные появились. Теперь можно перейти к машинному обучению.

Теперь у нас с вами есть множество, в общем случае многомерных, данных

$$X = \{x_i\}_i^n$$

Мы хотим получить закон $p(x)$, согласно которому они были сгенерированы.

Теперь у нас с вами есть множество, в общем случае многомерных, данных

$$X = \{x_i\}_i^n$$

Мы хотим получить закон $p(x)$, согласно которому они были сгенерированы. Введем параметризацию этого закона θ (частный случай гипотезы h) – тогда

$$p(\theta|X) = \frac{p(X|\theta)p(\theta)}{\int_\theta p(X|\theta)p(\theta)d\theta}$$

$$p(x|X) = \int_\theta p(x|\theta)p(\theta|X)d\theta$$

Однако, проблема в том, что зачастую $p(\theta|X)$ не может быть представлен аналитически.
Более того, $\int_\theta p(x|\theta)p(\theta|X)d\theta$ зачастую не может быть взят и посчитан численно, поскольку θ имеет достаточно большую размерность даже в простых задачах. Потому, для упрощения жизни, будем пока делать MAP (когда-то потом поговорим о том как все-таки аппроксимировать этот интеграл для более точного вывода).

$$p(x|X) = p(X|\theta_{MAP})$$

$$\theta_{MAP} = argmax_\theta \frac{p(X|\theta)p(\theta)}{\int_\theta p(X|\theta)p(\theta)d\theta} =$$

$$= argmax_\theta p(X|\theta)p(\theta)$$

Заключение

На этой лекции мы разобрали фреймворк, которым мы будем с вами пытаться решать задачи машинного обучения.
Его достаточно общая форма позволяет нам применять его практически везде, описывая им различные алгоритмы.
В следующий раз мы начнем говорить о том, как все-таки вводится параметризация на различные законы, и даже посмотрим как работают простейшие алгоритмы машинного обучения на примере.