Algebraic Classifiers

Дмитрий Кондрашкин

8 октября 2013 г.

Содержание

Введение
Обозначения
Определения
Новые алгоритмы
Быстрая кросс-валидация
Online обучение
Параллельное обучение
Примеры алгоритмов
Байесовский классифика-тор
Homstumps
Обобщение
Обучение и классифика-ция

Обозначения

- $o \in \mathcal{L} \times \mathcal{A}$ — объект
- $\mathcal{D}$ — пространство выборок
- $\mathcal{M}$ — множество всевозможных моделей для заданного алгоритма
- $T : \mathcal{D} \to \mathcal{M}$ — функция обучения
- $C : \mathcal{M} \times \mathcal{A} \to \mathcal{L}$ — функция классификации

Определения

Определение

Моноид — множество, на котором задана бинарная ассоциативная операция $\diamond$: $\mathcal{M} \times \mathcal{M} \to \mathcal{M}$, и в котором существует такой элемент ϵ, что для любого $m \in \mathcal{M}$ выполнено:

$$\epsilon \diamond m = m \diamond \epsilon = m.$$

Определение

Пусть $\langle \mathcal{D}, \oplus \rangle, \langle \mathcal{M}, \diamond \rangle$ — моноиды, тогда отображение $T : \mathcal{D} \to \mathcal{M}$ является *гомоморфизмом*, если для любых $d_1, d_2 \in \mathcal{D}$ выполнено:

$$T(d_1 \oplus d_2) = T(d_1) \diamond T(d_2)$$

Содержание

Допущения

- Алгоритм обучется за $O(n)$, где n — размер обучающей выборки
- Операции $\diamond, \oplus$ работают за $O(1)$
- Функция классификации работает за $O(1)$

Быстрая кросс-валидация

Схема алгоритма:

- Разобъем выборку на k подвыборок
- Обучим модель m_i на каждой подвыборке d_i
- Получим "префиксные" модели $p_i = m_i \diamond p_{i-1}, p_0 = \epsilon$
- Получим "суффиксные" модели $s_i = m_i \diamond s_{i+1}, s_{k+1} = \epsilon$

Тогда модель $p_{i-1} \diamond s_{i+1}$ обучена на всей выборке кроме объектов из d_i.
Сложность $O(n)$.

Online обучение

Online обучение: $T^O : \mathcal{M} \times \mathcal{D} \to \mathcal{M}$.
Пусть $m \in M$ уже обученная на выборке $d_m \in \mathcal{D}$ модель, $d \in \mathcal{D}$ — новые объекты. Тогда:

$$T^O(m, d) = m \diamond T(d).$$

Используем определение гомоморфизма:

$$T^O(m, d) = T(m_d) \diamond T(d) = T(d_m \oplus d).$$

Параллельное обучение

- Разбиваем выборку d на p подвыборок $d_1, \ldots, d_p$
- Каждый процессор обучает модель $m_i = T(d_i)$

Используем определение гомоморфизма:

$$T(d_1) \diamond T(d_2) \diamond \ldots \diamond T(d_p) = T(d_1 \oplus d_2 \oplus \ldots \oplus d_p).$$

Содержание

Введение
Обозначения
Определения
Новые алгоритмы
Быстрая кросс-валидация
Online обучение
Параллельное обучение
Примеры алгоритмов
Байесовский классифика-тор
Homstumps
Обобщение
Обучение и классифика-ция

Байесовский классификатор

Функция классификации:

$$C(a) = \underset{l \in \mathcal{L}}{\operatorname{argmax}} P(L = l) P(A = a | L = l).$$

Модель $\mathcal{M} = (P(L), P(A|L))$. Определим операцию $\diamond$:

$$(P_a(L), P_a(A|L)) \diamond (P_b(L), P_b(A|L)) =$$
$$(P_a(L) \odot P_b(L), P_a(A|L) \odot P_b(A|L))$$

Homstumps

Можем оценить распределения $P_l(A_i)$ для каждого класса $l \in \mathcal{L}$, и для каждого признака A_i.
Пересечение (overlap):

$$o(i) = \sum_{a \in A_i} \min_{l \in \mathcal{L}} P_l(a).$$

Разделяем по признаку с наименьшим пересечением:

$$s = \operatorname*{argmin}_{i \in \{1,\ldots,t\}} o(i).$$

Функция классификации:

$$C((a_1, \ldots, a_t)) = \operatorname*{argmin}_{l \in \mathcal{L}} P(L = l) P(A_s = a_s | L = l).$$

Содержание

Обучение и классификация

Добавим структуру для произвольных моделей:

$$\mathcal{F}(\mathcal{M}) = \{(\mathbb{Z}, \mathcal{M})\}.$$

Определим $\mathcal{F}(T) : \mathcal{F}(\mathcal{D}) \to \mathcal{F}(\mathcal{M})$:

$$\mathcal{F}(T) = \{(1, d_1), \ldots, (1, d_k)\} \mapsto \{(1, T(d_1)), \ldots, (1, T(d_k))\}.$$

Также определим функции δ и μ, чтобы следующая диаграмма была коммутативной:

$$\begin{array}{ccc} \mathcal{D} & \xrightarrow{T} & \mathcal{M} \\ \downarrow \delta & & \uparrow \mu \\ \mathcal{F}(\mathcal{D}) & \xrightarrow{\mathcal{F}(T)} & \mathcal{F}(\mathcal{M}) \end{array}$$

Обучение и классификация

Если $\mathcal{M}$ — моноид, определим функции μ и δ:

$$\delta = \{x_1, \ldots, x_n\} \mapsto \{\{(1, x_1)\}, \ldots, \{(1, x_n)\}\},$$
$$\mu = \{(1, m_1), \ldots, (1, m_k)\} \mapsto m_1 \diamond \ldots \diamond m_k.$$

Функция классификации:

$$C_{\mathcal{F}(\mathcal{M})}(m, a) = C_{\mathcal{M}}(\mu(m), a).$$

Обучение и классификация

Рассмотрим случай, когда $\mathcal{M}$ не является моноидом. Заметим, что, зафиксировав объект a, получаем функцию классификации $C_a : \mathcal{M} \to \mathcal{L}$. Рассмотрим диаграмму:

$$\begin{array}{ccccc} \mathcal{D} & \xrightarrow{T} & \mathcal{M} & \xrightarrow{C_a} & \mathcal{L} \\ \downarrow{\scriptstyle \delta} & & & & \uparrow{\scriptstyle \lambda} \\ \mathcal{F}(\mathcal{D}) & \xrightarrow{\mathcal{F}(T)} & \mathcal{F}(\mathcal{M}) & \xrightarrow{\mathcal{F}(C_a)} & \mathcal{F}(\mathcal{L}) \end{array}$$

Выберем λ, чтобы диаграмма “почти” коммутировала, можно взять функцию голосования.

Оригинальная статья:
Algebraic classifiers: a generic approach to fast cross-validation, online training, and parallel training — Michael Izbicki, 2013