Алгоритмическая теория обучения

Глеб Пособин

9 ноября 2016 г.

Базовые определения

X — пространство примеров, обычно $\{0,1\}^n$ или $\mathbb{R}^n$.

Концепт — подмножество X. Обычно концепты удобно рассматривать как отображения $X \to \{0,1\}$.

Класс концептов $\mathcal{C}$ — множество концептов.

Пример для концепта $h \in 2^X$ — пара $(x, h(x)) \in X \times \{0,1\}$.

На X задано распределение $\mathcal{D}$. Тогда возникает естественное расстояние между концептами: $dist(a,b) := \mathbf{Pr}_{x \in \mathcal{D}}[a(x) \neq b(x)]$.

Некоторый выделенный концепт c называют целевым. Расстояние между c и произвольным концептом h называют ошибкой: $error(h) := dist(c,h)$

Оракул $EX(c,\mathcal{D})$ за единичное время генерирует объект x из $\mathcal{D}$ и возвращает пару $(x, c(x))$.

PAC модель

Класс концептов $\mathcal{C} \subset 2^X$ называют PAC-обучаемым, если есть алгоритм L, принимающий на вход числа ε, δ и имеющий доступ к оракулу $EX(c, \mathcal{D})$, который для всех чисел $0 < \varepsilon, \delta < 1/2$, всех распределений $\mathcal{D}$ на X и всех $c \in \mathcal{C}$ с вероятностью по крайней мере $1 - \delta$ возвращает концепт $h \in \mathcal{C}$ с $error(h) < \varepsilon$.

Если L работает за время, полиномиальное от размерности X, $1/\varepsilon$ и $1/\delta$, то $\mathcal{C}$ называют эффективно PAC-обучаемым.

Пример

Разрешающий список L длины k над n булевыми переменными $x_1, \ldots, x_n$ — список из k пар $(\ell_1, b_1), \ldots, (\ell_k, b_k)$ и бит b_{k+1}, где ℓ_i — литералы, а $b_i \in \{0, 1\}$.

Значение L на входе x:

$$\textbf{if } \ell_1 \textbf{ then } b_1 \textbf{ else if } \ell_2 \textbf{ then } b_2 \ldots \textbf{else if } \ell_k \textbf{ then } b_k \textbf{ else } b_{k+1}$$

Пример

1. Создать пустой разрешающий список L.
2. Запросить m примеров и запомнить их в S.
3. Если все примеры в S имеют одинаковый ответ, то останавливаемся.
4. Иначе, находим литерал ℓ такой, что все примеры в S, для которых $\ell = 1$, имеют одинаковый ответ b. Добавляем пару (ℓ, b) к L, удаляем из S все примеры с $\ell = 1$, переходим к шагу 3.

Надо подобрать m и проверить, что алгоритм работает корректно.

Бритва Оккама

Лемма
Дан концепт $c \in \mathcal{C}$ и выборка S для c размера m, сгенерированная из распределения $\mathcal{D}$. Тогда вероятность, что существует концепт h с $error(h) \geqslant \varepsilon$, не больше $|C|(1-\varepsilon)^m$.

Если $m \geqslant 1/\varepsilon(\ln|C| + \ln 1/\delta)$, то:

$$\delta \geqslant |C|(1-\varepsilon)^m \geqslant \mathbf{Pr}[error(h) \geqslant \varepsilon]$$

Размерность Вапника-Червоненкиса

Для класса концептов $\mathcal{C}$ над X и $S \subseteq X$ определим:

$$\Pi_{\mathcal{C}}(S) := \{c \cap S : c \in \mathcal{C}\}$$

Если $|\Pi_{\mathcal{C}}(S)| = 2^{|S|}$, говорят, что $\mathcal{C}$ разбивает S.

$VCD(\mathcal{C})$ — мощность наибольшего конечного множества S, которое разбивается классом $\mathcal{C}$.

Примеры

Отрезки на прямой.

Полуплоскости.

Прямоугольники со сторонами, параллельными осям, на плоскости.

Замечательное свойство $\Pi_{\mathcal{C}}(m)$

Для натурального m определим:

$$\Pi_{\mathcal{C}}(m) := \max\{|\Pi_{\mathcal{C}}(S)| : |S| = m\}$$

Теорема

Либо всегда $\Pi_{\mathcal{C}}(m) = 2^m$, либо $\Pi_{\mathcal{C}}(m) = \mathcal{O}(m^d)$, где $d = VCD(\mathcal{C})$.

VCD и PAC

Теорема

Пусть $\mathcal{C}$ — класс концептов, $VCD(\mathcal{C}) = d$. L — алгоритм, который принимает выборку S из m объектов, соответствующую какому-то концепту в $\mathcal{C}$, и возвращает концепт $h \in \mathcal{C}$, который согласован с выборкой S. Тогда L — PAC алгоритм для $\mathcal{C}$, если ему дана выборка из m объектов из $EX(c, \mathcal{D})$, и m удовлетворяет неравенству:

$$m \geqslant \frac{c_0}{\varepsilon} \left(\log \frac{1}{\delta} + d \log \frac{1}{\varepsilon} \right)$$

для какой-то константы $c_0 > 0$.

В бритве Оккама:

$$m \geqslant \frac{1}{\varepsilon} \left(\ln \frac{1}{\delta} + \ln |\mathcal{C}| \right).$$

Модель с ограниченным числом ошибок

Пусть алгоритм $\mathcal{A}$ на каждом шаге принимает вектор $x \in X$, выдаёт α^* — догадку, а потом получает $\alpha = c(x)$. Тогда, если число ошибок, которое допускает $\mathcal{A}$, ограничено полиномом от n и $size(c)$ для всех возможных последовательностей векторов x и всех $c \in \mathcal{C}$, говорят, что $\mathcal{A}$ изучает $\mathcal{C}$ в модели с ограниченным числом ошибок.

Разрешающие списки

Алгоритм поддерживает список h из n непересекающихся множеств, в каждом лежат пары (ℓ, b), и, возможно, терминальные биты.

1. Инициализировать h, положив в первое множество все возможные пары (ℓ, b) и два возможных терминальных бита — всего $4n + 2$ элемента.
2. Получив пример x, найти первое множество с правилом, которое выполняется, предсказать бит, указанный в правиле.
3. Если ошиблись, перемещаем все правила, которые ошиблись, в следующее множество в списке.
4. Возвращаемся к шагу 2.

Алгоритм допускает не более $\mathcal{O}(nk)$ ошибок.

Ограниченное число ошибок и PAC

Теорема
Если алгоритм $\mathcal{A}$ изучает класс $\mathcal{C}$ в модели с ограниченным числом ошибок, то $\mathcal{C}$ PAC-обучаем.

Если $\mathcal{A}$ допускает не более M ошибок, то для $\mathcal{C}$ есть алгоритм, требующий всего $\frac{M}{\varepsilon}\ln(\frac{M}{\delta})$ примеров.

Можно улучшить до $\mathcal{O}(\frac{1}{\varepsilon}(\ln(\frac{1}{\delta}) + M))$ примеров.