Фундаментальные теоремы машинного обучения.

Николай Савельев

Московский физико-технический институт

Постановка задачи

Задана неотрицательная матрица $Q \in \mathbb{R}_{+}^{n \times k}$.

Рассмотрим функционал $F : \mathbb{R}^k \to \mathbb{R}$,

$$F(p) = \sum_{j=1}^{k} p_j \sum_{i=1}^{n} I(Q, p, i, j), \tag{1}$$

$$I(Q, p, i, j) = \begin{cases} 1, if\ S := \{s : q_{is} \geq p_s\} \neq \emptyset\ and\ j = arg \max\limits_{s \in S} q_{is} \\ 0, otherwise \end{cases}.$$

Требуется решить оптимизационную задачу

$$F(p) \to \max_{p}. \tag{2}$$

Замечание. Для любой матрицы $Q \in \mathbb{R}_+^{n\times k}$ существует решение оптимизационной задачи (2). Причем найдется такой набор индексов $i_1, \dots i_k$, что вектор $(q_{i_1 1}, \dots, q_{i_k k})^T$ является решением.

Доказательство. Введем множества $Q_j = \{q_{ij}\}_{i=1}^n$, $j = 1, \dots, k$ и множество
$P = \{p \in \mathbb{R}^k : \min Q_j \le p_j \le \max Q_j, j = 1, \dots, k\}$. Тогда
$\sup F(p) = \sup\limits_{p\in P} F(p)$.
Рассмотрим произвольный вектор $x \in P$. Постором для него вектор $y : y_j = \min\limits_{q\in Q_j, q\ge x_j} q$. Тогда $F(x) \le F(y)$.
Следовательно, можно искать максимум $F(p)$ на конечном множестве точек вида $(q_{i_1 1}, \dots, q_{i_k k})^T$.

Процедура clean

Пусть p^* - искомое решение оптимизационной задачи (2). Рассмотрим процедуру, позволяющую значительно уменьшить множество точек, внутри которого находится p^*.

Процедура clean:

1. Упорядочим элементы матрицы Q по убыванию.

2. Рассмотрим первый элемент получивщегося упорядоченного набора - $q_{i_1 j_1}$.
$p^*_{j_1} \leq q_{i_1 j_1} \Rightarrow I(Q, p, i_1, j) = 0, \; j \neq j_1 \; (q_{i_1 j} \leq q_{i_1 j_1}) \Rightarrow$ можно убрать из рассмотрения все элементы вида $q_{i_1 j} : q_{i_1 j} \leq q_{i_1 j_1}$.

3. Рассматриваем следующий элемент обновленного упорядоченного набора, имеющий другой второй индекс - $q_{i_2 j_2}$, первый индекс этого элемента тоже будет другой, так как иначе он был бы удален на предыдущем шаге.
$p^*_{j_2} \leq q_{i_2 j_2} \Rightarrow I(Q, p, i_2, j) = 0, \; q_{i_2 j} \leq q_{i_2 j_2} \Rightarrow$ можно убрать из рассмотрения все элементы вида $q_{i_2 j} : q_{i_2 j} \leq q_{i_2 j_2}$.

4. Аналогично повторяем шаг 3 для остальных индексов.

Пример

Пусть $Q = \begin{pmatrix} 3 & 5 & 1 \\ 2 & 4 & 6 \\ 7 & 0 & 8 \end{pmatrix}$.

Изначальный упорядоченный набор - 8, 7, 6, 5, 4, 3, 2, 1, 0.

Набор после первого шага - 8, 6, 5, 4, 3, 2, 1.

Набор после второго шага - 8, 6, 5, 4, 2.

Оставшиеся элементы матрицы Q : $\begin{pmatrix} & 5 & \\ 2 & 4 & 6 \\ & & 8 \end{pmatrix}$.

Теорема

Тем самым была доказана следующая теорема.

Теорема. Пусть $Q \in \mathbb{R}_+^{n \times k}$, $n \geq k$. Тогда функционал (1) достигает максимума на векторе вида $(q_1, \ldots, q_k)^T$, где q_j - это элемент j-го столбца матрицы Q, который не был удален процедурой clean.

Следствие

Следствие (Необходимое условие единственности максимума). Пусть $Q \in \mathbb{R}_+^{n \times k}$. Для того, чтобы точка максимума функционала (1) была единственна, необходимо выполнение неравенства $n \geq k$.

Доказательство. Пусть $n < k$. Применим к матрице Q процедуру clean. На каждом шаге процедуры clean рассматривается элемент матрицы Q с индексами, отличными от индексов элементов, рассмотренных на предыдущих шагах. Следовательно, будет сделано n шагов. Но тогда в матрице Q будет столбец, ни один элемент которого не рассматривался в ходе процедуры clean. А значит, все элементы этого столбца меньше рассмотренных и все они будут удалены. Обозначим индекс этого столбца, как $\hat{j}$.

Пусть p^* - точка максимума функционала (1). Тогда изменение компоненты $p^*_{\hat{j}}$ не изменяет значения функционала. Следовательно, точка максимума не единственна.