

Online Convex Optimization (OCO)

Михаил Лепехин и Роман Логинов, группа 694

24 декабря 2018 г.

Применение Online Convex Optimisation к задаче фильтрации спама

Предположим, что признаки email-сообщений принадлежат множеству \mathcal{X} . В качестве признаков будем рассматривать частоты вхождений слов (или групп слов, чтобы размерность мн-ва признаков не получилась слишком большой) в сообщение.

На каждом шаге t функция $a_t : \mathcal{X} \rightarrow [0, 1]$, сопоставляет вектору $x \in \mathcal{X}$ значений признаков некоторое число из отрезка $[0, 1]$. По смыслу это значение является оценкой вероятности (уверенности) того, что сообщение с данными значениями признаков является спамом.

На каждом шаге t соперник выбирает вектор значений x_t и индикатор y_t того, что данное сообщение является спамом.

Для оценки точности метода принятия решений a_t нужно взять некоторую функцию потерь f_t . Например, квадратичную функцию потерь:

$$f_t(a_t) := (y_t - a_t(x_t))^2.$$

Методы первого порядка

В данном разделе мы рассмотрим базовые алгоритмы для Online Convex Optimization, которые достаточно неплохо применимы на практике.

В целом данные методы похожи на соответствующие методы первого порядка для задач обычной выпуклой оптимизации. Но они принципиально отличаются целью применения. Ведь при помощи методов ОСО мы стремимся минимизировать не ошибку оптимизации, а *regret*:

$$\text{regret} = \sum_{t=1}^T f_t(x_t) - \min_{x \in \mathcal{K}} \sum_{t=1}^T f_t(x)$$

Для сравнения *regret* с ошибкой оптимизации полезно рассмотреть среднее значение *regret*, т.е. $\frac{\text{regret}}{T}$.

Введём обозначение:

$$\bar{x}_T := \frac{1}{T} \sum_{t=1}^T x_t$$

Пусть все функции f_t равны некоторой функции $f : \mathcal{K} \rightarrow \mathbb{R}$, то из неравенства Йенсена получим:

$$f(\bar{x}_T) - f(x^*) = f(\bar{x}_T) - \frac{1}{T} \sum_{t=1}^T f(x^*) \leq \frac{1}{T} \sum_{t=1}^T (f(x_t) - f(x^*))$$

Таким образом мы показали следующий факт:

функция $f(x_T)$ сходится к $f(x^*)$ не менее быстро, чем среднее значение *regret*.