

Курсовая Работа по ТПР

Терешков Алексей, Алимов Исмаил, Мамченков Дмитрий

1 Задача

Рассмотрим алгоритм экспоненциального взвешивания на основе градиента с переменным шагом обучения. Приведите исчерпывающие выкладки и получите оценку на кумулятивный регрет. Как следует выбирать параметр обучения на каждом шаге?

2 Решение

Алгоритм экспоненциального взвешивания на основе градиента с переменным шагом обучения (EWGD) применяется для минимизации функционала потерь в задачах оптимизации. Этот алгоритм является модификацией стандартного метода градиентного спуска, в котором используется переменный шаг обучения, зависящий от номера итерации.

Пусть дана функция потерь $f(w)$, которую необходимо минимизировать по параметру w . Тогда на каждом шаге алгоритма мы вычисляем градиент функции $f(w)$ и изменяем параметр w следующим образом:

$$w_{t+1} = w_t - \alpha_t \nabla f(w_t),$$

где α_t - шаг обучения на шаге t , который может меняться на каждой итерации.

В алгоритме EWGD шаг обучения задается следующим образом:

$$\alpha_t = \frac{\eta}{\sqrt{\sum_{i=1}^t g_i^2}},$$

где η - параметр скорости обучения, g_i - градиент функции на i -ой итерации.

Для оценки кумулятивного регрета используется следующая формула:

$$Regret = \sum_{t=1}^T (f(w_t) - f(w^*)),$$

где $f(w^*)$ - оптимальное значение параметра w .

Теперь рассмотрим, как выбирать параметр обучения на каждом шаге. В алгоритме EWGD шаг обучения зависит от градиентов функции на предыдущих итерациях, поэтому выбор этого параметра является важным вопросом. Обычно значение η выбирается заранее и остается постоянным на протяжении всего процесса обучения. Однако, в некоторых случаях может быть полезно использовать адаптивный подход и изменять значение параметра η на каждой итерации в зависимости от поведения функции потерь.

Существуют различные методы адаптивной настройки параметра η , такие как Adagrad, RMSprop, Adam, которые позволяют эффективно подбирать параметр скорости обучения в зависимости от свойств оптимизируемой функции. Один из распространенных методов адаптивной настройки параметра η - это метод экспоненциального затухания (Exponential Decay). Согласно этому методу, значение η уменьшается на каждой итерации пропорционально числу итераций:

$$\eta_t = \eta_0 \cdot \epsilon^{-\gamma t},$$

где η_0 - начальное значение параметра скорости обучения, γ - коэффициент затухания.

Другой подход - это использование алгоритма Adam, который позволяет эффективно настраивать параметры η и моменты градиента на каждой итерации. Алгоритм Adam сочетает в себе методы первого и второго порядка и является одним из наиболее эффективных методов оптимизации в задачах машинного обучения.

Независимо от выбранного метода настройки параметра η , важно следить за процессом обучения и вносить коррективы в выбор параметров, если необходимо. В целом, выбор параметра скорости обучения является важным вопросом в алгоритме EWGD, и его правильная настройка может значительно повысить эффективность процесса оптимизации.