

Математика: Градиентный перебор

Предыстория

В качестве финального задания Иван попросил Элис помочь ему с реализацией задачи машинного обучения «Зеленая линия». В частности, он попросил изучить математические основы алгоритма градиентного спуска. Помогите Элис выполнить задание!

Задача

Мы пытаемся обучить исключительно сложную модель (одномерная линейная регрессия, она же точка) на исключительно сложном наборе данных (1 одномерная точка, расположенная в точке $x = 5$), и нам нужна ваша помощь...

По сути, для такой маломерности задача сводится к поиску точки с помощью модели линейной регрессии. Чтобы определить для компьютера, насколько мы близки к цели, мы используем так называемую функцию потерь. Для этой задачи она определена как квадрат расстояния между текущей точкой и точкой $x = 5$:

$$L(x) = (x - 5)^2$$

Убедитесь, что значение функции потерь минимально и равно нулю при $x = 5$.

Модель начинает с произвольной точки x_0 и пытается достичь точки $x = 5$. На каждом шаге она проверяет, увеличивает ли движение вправо значение функции потерь. Для этого она вычисляет *наклон* или *производную* функции потерь L по отношению к ее аргументу x (или, для шага t , x_t), что математически записывается как $\frac{dL}{dx_t}$. Очевидно, что нам следует перемещаться в положение с меньшим значением L . Когда движение вправо увеличивает функцию потерь (что означает, что производная положительна), имеет смысл двигаться влево, и наоборот. Чтобы рассчитать, насколько большим должен быть шаг, мы берем производную и умножаем ее на константу α . Выбор α обычно остается за вами (и, спойлер, вы узнаете, как выбирать эту константу). Пока что предположим, что α положительна.

Более формально, переход от шага t к $t + 1$ задается формулой:

$$x_{t+1} = x_t - \alpha \cdot \frac{dL}{dx_t}$$

Производная задается как

$$\frac{dL}{dx_t} = 2 \cdot (x_t - 5)$$

Убедитесь, что если производная положительна, мы делаем шаг влево, и наоборот.

Если вы знакомы с машинным обучением, то наверняка поняли, что описанная процедура — это градиентный спуск, а α — скорость обучения.

Ваша задача следующая:

Задание {1} (5 баллов) Проведите градиентный спуск, начиная с $x_0 = 0$, с $\alpha = 0.3$. Вы можете остановиться, если значение функции потерь L на шаге x_t не превышает 1. Для каждого шага покажите следующее: $x_t, L(x_t), \frac{dL}{dx_t}, x_{t+1}$.

Обратите внимание! В соответствии с форматом $\LaTeX$ (и англоязычным форматом в принципе), мы используем точку, чтобы разделить целую часть числа от дробной.

Задание {2} (15 баллов) Изучите, как происходит градиентный спуск для разных значений α . Вы заметите, что для некоторых значений α модель будет приближаться к 5, а для других — нет. Опишите, для каких значений α модель сходится (или не сходится) к 5. Попытайтесь определить диапазон этих значений, для которых модель сходится. Ваше объяснение должно быть как можно более полным и рассматривать все возможные случаи.

Подсказка: вам могут пригодиться свойства геометрической прогрессии.

Формат

Для части {1} составьте таблицу со столбцами: t , x_t , $L(x_t)$, $\frac{dL}{dx_t}$, x_{t+1} .

Для части {2} дайте математическое объяснение. Мы не требуем формального математического доказательства, но для получения полного балла вы должны предоставить валидное математическое объяснение, а не просто наблюдения.

Включите важные детали, но избегайте ненужных или длинных разъяснений. Это поможет нам быстрее и точнее оценить ваше решение. Вы можете предоставить свое решение на латышском, русском или английском языках.

Оценивание

- 25% за правильную часть за задание {1}
- 75% за правильную часть за задание {2}