СОДЕРЖАНИЕ

1 Методы локальной оптимизации

Оптимизация – это задача нахождения экстремума (минимума или максимума) целевой функции в некоторой области конечномерного векторного пространства, ограниченной набором линейных и/или нелинейных равенств и/или неравенств.

Во многих практически важных случаях для целевой функции многих переменных $f(\boldsymbol{x})$ задача оптимизации может быть сформулирована в виде:

$$f(\boldsymbol{x}) \to \min,$$

где $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$ – вектор неизвестных (управляющих параметров); min – минимальное значение функции в ограниченной или неограниченной области изменения неизвестных.

1.1 Метод градиентного спуска

Градиентный спуск – метод нахождения локального экстремума (минимума или максимума) функции многих переменных $f(x_1, x_2, \dots, x_n)$ с помощью движения вдоль градиента этой функции. Это наиболее простой в реализации из всех методов локальной оптимизации, но имеет относительно малую (линейную) скорость сходимости.

Градиент ∇ это вектор, указывающий направление наибольшего возрастания некоторой функции f, значение которой меняется от одной точки пространства к другой (скалярного поля), а по величине (модулю) равный скорости роста этой величины в этом направлении. Компонентами вектора градиента являются частные производные f по всем её аргументам:

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n}\right) \tag{1}$$

Для случая трёхмерного пространства градиентом скалярной функции $f(x, y, z)$ называется векторная функция:

$$\operatorname{grad} f = \nabla f,$$

где ∇ – векторный дифференциальный оператор набла, компоненты которого являются частными производными по координатам:

$$\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)$$

Следует отметить, что оператор набла не принадлежит тому же пространству, что и обычные векторы, а говоря точнее, скалярное и векторное произведение для него определено с некоторыми отличиями. Оператор ∇ действует на те скалярные поля, что стоят от него справа, и не действует на стоящие от него слева. Поэтому скалярное и векторное произведение с участием ∇ *не коммутативны* и не антикоммутативны, как это свойственно для таких произведений обычных векторов.

Минимизация целевой функции $f(\boldsymbol{x})$ сводится к итерационному процессу последовательного выбора нового вектора неизвестных $\boldsymbol{x}_{k+1}$, такого чтобы значение функции в новой точки было меньше чем в предыдущих:

$$f(\boldsymbol{x}_0) > f(\boldsymbol{x}_1) > ... > f(\boldsymbol{x}_k) > f(\boldsymbol{x}_{k+1}) > ...$$

Предполагая, что новый вектор неизвестных мало отличается от предыдущего ($\boldsymbol{x}_{k+1} - \boldsymbol{x}_k \approx 0$), можно воспользоваться линейным приближением для разложения в ряд Тейлора целевой функции:

$$f(\boldsymbol{x}_{k+1}) = f(\boldsymbol{x}_k) + \left(\nabla f(\boldsymbol{x}_k), \boldsymbol{x}_{k+1} - \boldsymbol{x}_k\right) \tag{2}$$

Если в качестве нового вектора неизвестных выбрать:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \lambda \cdot \nabla f(\boldsymbol{x}_k), \tag{3}$$

то из (2) получим:

$$f(\boldsymbol{x}_{k+1}) = f(\boldsymbol{x}_k) - \lambda \cdot \|\nabla f(\boldsymbol{x}_k)\|^2 \tag{4}$$

где $\lambda > 0$ – малое положительное число; $\|\nabla f(\boldsymbol{x}_k)\| \geqslant 0$ – норма вектора градиента:

$$\|\nabla f\| = \sqrt{(\nabla f, \nabla f)}$$

Основная идея метода градиентного спуска заключается в том, чтобы последовательно идти в направлении наибольшего уменьшения целевой функции, которое задаётся антиградиентом $-\nabla f(\boldsymbol{x})$:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \boldsymbol{\lambda} \cdot \nabla f(\boldsymbol{x}_k) \tag{5}$$

где k – номер итерационного шага процесса; $\boldsymbol{x}_k$ – значение неизвестных на k-ой итерации; $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n)$ – скорость градиентного спуска в направлении осей координат.

Практически можно задать некоторое число $\varepsilon > 0$, связанное с выбранной точностью вычислений, и проводить итерации до тех пор, пока на k-ой итерации не будут выполнены одно или несколько неравенств вида:

$$\|\boldsymbol{x}_k - \boldsymbol{x}_{k-1}\| < \varepsilon_1, \quad \|f(\boldsymbol{x}_k) - f(\boldsymbol{x}_{k-1})\| < \varepsilon_2 \tag{6}$$

1.2 Алгоритм метода градиентного спуска

1) Задают начальное приближение (x^0, y^0), скорость градиентного спуска λ_x и λ_y, а также точность расчёта ε.
2) Рассчитывают градиент целевой функции в текущей точке $\nabla^{(0)} = \nabla S(x^0, y^0)$.
3) Определяют новую точку в соответствии с соотношением:

$$\begin{aligned} x^{(1)} &= x^{(0)} - \lambda_x^{(0)} \cdot \nabla_x^{(0)} \\ y^{(1)} &= y^{(0)} - \lambda_y^{(0)} \cdot \nabla_y^{(0)} \end{aligned}$$

4) Рассчитывают величину расстояния между двумя точками:

$$r = \sqrt{\left[x^{(0)} - x^{(1)}\right]^2 + \left[y^{(0)} - y^{(1)}\right]^2}$$

5) Проверяют условие остановки: если $r < \varepsilon$, то итерационный процесс останавливается; иначе текущую точку считают начальной $x^{(0)} = x^{(1)}$ и $y^{(0)} = y^{(1)}$ и переходят к шагу (2) итерационного процесса.

1.3 Метод тяжелого шара

Поиск минимума функции многих переменных $f(\boldsymbol{x})$ методом “тяжелого шара“ основан на аналогии движения материальной частицы массой m в консервативном силовом поле $\boldsymbol{F}(\boldsymbol{x})$ в вязкой среде.

В соответствии с принципом минимальной энергии тело смещается в положение, которое минимизирует общую потенциальную энергию системы

$f(\boldsymbol{x}) \to \min$. Поэтому если предположить, что функция $f(\boldsymbol{x})$ является потенциальной энергией частицы в консервативном силовом поле $\boldsymbol{F}(\boldsymbol{x}) = -\nabla f(\boldsymbol{x})$, и частица перемещается в пространстве $\boldsymbol{x}$ минимизируя свою энергию, то уравнение движения этой частицы можно записать в виде:

$$\begin{cases} \dfrac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t} &= \boldsymbol{v} \\ m\,\dfrac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} &= \boldsymbol{F} - \alpha \cdot \boldsymbol{v} \end{cases} \tag{7}$$

где $\boldsymbol{r}$ – положение частицы в выбранной системе координат; $\boldsymbol{v}$ и α – скорость и коэффициент вязкого трения частицы в среде, соответственно.

Этот метод используется в методе стохастического градиентного спуска и в качестве расширения алгоритмов обратного распространения ошибок для обучения искусственных нейронных сетей.

Поиск минимума данным методом начинается с задания начальных условий, которые, как правило, формулируются в виде:

$$\begin{cases} \boldsymbol{r}(0) = & \boldsymbol{r}_0 \\ \boldsymbol{v}(0) = & 0 \end{cases}, \tag{8}$$

где $\boldsymbol{r}_0$ – начальное приближения для поиска минимума функции.

Масса частицы m и коэффициент вязкого трения α являются эвристическими параметрами метода и выбираются произвольным образом, отражающим специфику решаемой задачи.