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:

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

где k – номер итерационного шага процесса; $\boldsymbol{x}_k$ – значение неизвестных на k-ой итерации.

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:

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

где $\lambda > 0$ – малое положительное число (параметр метода), имеющий смысл скорости градиентного спуска; $\|\nabla f(\boldsymbol{x}_k)\| \geqslant 0$ – норма вектора градиента (неотрицательное число):

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

Таким образом, выбор нового вектора неизвестных $\boldsymbol{x}_{k+1}$ в соответствии с выражением (3), гарантирует монотонное убывание целевой функции $f(\boldsymbol{x})$ в каждой итерации. Поэтому основная идея метода градиентного спуска заключается в том, чтобы последовательно идти в направлении наибольшего уменьшения целевой функции, которое задаётся антиградиентом $-\nabla f(\boldsymbol{x})$.

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:

$$\|\boldsymbol{x}_k - \boldsymbol{x}_{k-1}\| < \varepsilon_1, \quad \|f(\boldsymbol{x}_k) - f(\boldsymbol{x}_{k-1})\| < \varepsilon_2 \tag{5}$$

Алгоритм метода градиентного спуска

1) Задают начальное приближение (x_0, y_0), скорость градиентного спуска λ, а также точность расчёта ε.
2) Рассчитывают градиент целевой функции в текущей точке $\nabla_0 = \nabla f(x_0, y_0)$.
3) Определяют новый вектор неизвестных в соответствии с соотношением (3):

$$\begin{cases} x_1 &= x_0 - \lambda \cdot \nabla_{0_x} \\ y_1 &= y_0 - \lambda \cdot \nabla_{0_y} \end{cases},$$

   где ∇_{0_x} и ∇_{0_y} – компоненты вектора градиента в выбранной системе координат.
4) Рассчитывают величину расстояния между двумя точками:

$$r = \sqrt{(x_0 - x_1)^2 + (y_0 - y_1)^2}$$

5) Проверяют условие остановки итерационного процесса: 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.

1.2 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:

$$\begin{cases} \dfrac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} &= \boldsymbol{v} \\ m\,\dfrac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} &= \boldsymbol{F} - \alpha \cdot \boldsymbol{v} \end{cases} \tag{6}$$

где $\boldsymbol{x}$ – 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:

$$\begin{cases} \boldsymbol{x}(0) = & \boldsymbol{x}_0 \\ \boldsymbol{v}(0) = & \boldsymbol{v}_0 \end{cases}, \tag{7}$$

где $\boldsymbol{x}_0$ – начальное приближения для поиска минимума функции; $\boldsymbol{v}_0$ – “начальная скорость“ в пространстве неизвестных.

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