无约束最优化问题求解方法

W.J.Z

0.1 梯度下降法

假设 $f(x)$ 有一阶偏导数，若第 K 次迭代值为 x^k，则可将 $f(x)$ 在 x^k 附近展开一阶泰勒展开式。

$$f(x) = f\left(x^k\right) + \bigtriangledown f\left(x^k\right)$$

$\bigtriangledown f\left(x^k\right)$ 为 $f(x)$ 在 x^k 的梯度。

第 $k+1$ 逸代值为：

$$x^{k+1} \leftarrow x^k + \lambda_k p_k$$

p_k 为搜索方向，$p_k = -\bigtriangledown f\left(x^k\right)$,$\lambda_k$ 为搜索步长，使得：

$$f\left(x^k + \lambda_k p_k\right) = \underset{\lambda \geq 0}{min} f\left(x^k + \lambda p_k\right)$$

Algorithm 1: 梯度下降法

Input: 目标函数 $f(x)$, 梯度函数 $\bigtriangledown f\left(x^k\right)$，计算精度 ε

Output: $f(x)$ 的极小值点

1 取初始值 x^0, 置 $k=0$

2 计算 $f(x^k)$

3 计算梯度 $\bigtriangledown f\left(x^k\right)$, 当 $\left\|\bigtriangledown f\left(x^k\right)\right\| < \varepsilon$ 停止迭代，令极小点为 x^k；否则令 $p_k = -\bigtriangledown f\left(x^k\right)$, 求 λ_k, 使

$$f\left(x^k + \lambda_k p_k\right) = \underset{\lambda \geq 0}{min} f\left(x^k + \lambda p_k\right)$$

4 $x^{k+1} = x^k + \lambda_k p_k$ 计算 $f(x^{k+1})$, 当 $\left\|f\left(x^{k+1}\right) - f\left(x^k\right)\right\| < \varepsilon$ 或 $\left\|x^{k+1} - x^k\right\| < \varepsilon$ 时，停止迭代令极小值点为 x^{k+1}

5 否则，$k=k+1$, 转到步骤 3

0.2 牛顿法和 BFGS

牛顿法时迭代算法，每一步需要求解目标函数的海塞矩阵的逆矩阵，计算比较复杂。拟牛顿法通过正定矩阵近似海塞矩阵，简化这一计算过程.

0.2.1 牛顿法

假设 $f(x)$ 有二阶连续偏导数，如第 k 次迭代值为 x^k，则可在 $f(x)$ 在 x^k 附近展开二阶泰勒公式：

$$f(x) = f\left(x^k\right) + \bigtriangledown\left(x - x^k\right) + \frac{1}{2}\left(x - x^k\right)^T H\left(x^k\right)\left(x - x^k\right)$$

$\bigtriangledown f\left(x^k\right)$ 为 $f(x)$ 在 x^k 的梯度,$H(x^k)$ 是 $f(x)$ 的海塞矩阵：

$$H(x) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}\right]$$

函数 $f(x)$ 有极值的必要条件是在极值点处一阶导数为 0，即梯度向量为 0，当海塞矩阵是正定矩阵时 $(X^{-1}H(x)X) > 0$, 函数的极值为极小值。

假设 x^{k+1} 满足 $\bigtriangledown f\left(x^{k+1}\right) = 0$, 泰勒公式求导并将 x^{k+1} 代入得：

$$\bigtriangledown f\left(x^k\right) + H_k\left(x^{(k+1)-}x^{(k)}\right) = 0$$

因此

$$x^{(k+1)} = x^{(k)} - H_k^{-1} \bigtriangledown f\left(x^k\right)$$

令 $P_k = -H_k^{-1} \bigtriangledown f\left(x^k\right)$.

Algorithm 2: 牛顿法

Input: 目标函数 $f(x)$，梯度函数 $\bigtriangledown f(x)$, 海塞矩阵 $H(x)$ 和精度 ε

Output: 函数极小值点

1 取初始值 x^0, 置 $k = 0$

2 计算 $\bigtriangledown f\left(x^k\right)$

3 若 $\left\|\bigtriangledown f\left(x^k\right)\right\| < \varepsilon$，则停止计算，得到近似解 x^k.

4 计算 $H(x^k)$，求

$$P_k = -H_k^{-1} \bigtriangledown f\left(x^k\right)$$

5 置 $x^{(k+1)} = x^k + p_k$

6 置 $k = k + 1$, 转到步骤 2

拟牛顿算法使用一个 n 阶矩阵来代替海塞矩阵，来降低因海塞矩阵导致计算度复杂的问题。

Algorithm 3: BFGS 算法

Input: 目标函数 $f(x)$，梯度函数 $\bigtriangledown f(x)$, 精度 ε

Output: 函数极小值点

1 初始点 $x^{(0)}$，取 B_0 为正定对称矩阵，置 K=0.

2 计算 $\left\|\bigtriangledown f\left(x^k\right)\right\| < \varepsilon$，则停止计算，得到近似解 x^k.

3 否则计算 $H(x^k)$，求

$$P_k = -H_k^{-1} \bigtriangledown f\left(x^k\right)$$

4 求 λ_k, 使

$$f\left(x^k + \lambda_k p_k\right) = \min_{\lambda \geq 0} f\left(x^k + \lambda p_k\right)$$

5 置 $x^{k+1} = x^k + \lambda_k p_k$

6 计算 $\bigtriangledown f\left(x^{k+1}\right)$，若 $\left\|\bigtriangledown f\left(x^{k+1}\right)\right\| < \varepsilon$ 则停止计算得近似解，否则

$$B_{k+1} = B_k + \frac{y_k y_k^T}{y_k^T \partial_k} - \frac{B_k \partial_k \partial_k^T B_k}{\partial_k^T B_k \partial_k}$$

$y_k = \bigtriangledown f\left(x^{k+1}\right) - \bigtriangledown f\left(x^k\right)$,$\partial_k = x^{k+1} - x^k$

7 置 k=k+1, 转到步骤 2