梯度下降法

梯度下降法(gradient descent)或最速下降法(steepest descent)是求解无约束最优化问题的一种最常用的方法。

梯度下降是迭代算法，每一步需要求解目标函数的梯度向量

假设$f(x)$是R^n上具有一阶连续偏导数的函数。要求解的无约束最优化问题是 $\min_{x \in R^n} f(x)$，x^*表示目标函数$f(x)$的极小点。

迭代的基本思想：选取适当的初值$x^{(0)}$，不断迭代，更新x的值，进行目标函数的极小化，直到收敛。由于负梯度方向是使函数值下降最快的方向，在迭代中的每一步，以负梯度方向更新x的值，从而达到减少函数值的目的。

由于$f(x)$具有一阶连续偏导数，若第k次迭代值为$x^{(k)}$，则可将$f(x)$在$x^{(k)}$附近进行一阶泰勒展开：

$$f(x) = f(x^{(k)}) + g_k^T (x - x^{(k)})$$

补充：泰勒公式 $f(x) = \frac{f(a)}{0!} + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + R_n(x)$

这里，$g_k = g(x^{(k)}) = \nabla f(x^{(k)})$为$f(x)$在$x^{(k)}$的梯度。

求出第$k+1$次迭代值$x^{(k+1)}$：

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

其中，p_k是搜索方向，取负梯度方向$p_k = -\nabla f(x^{(k)})$，λ_k是步长，由一维搜索确定，即λ_k使得

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

算法A.1 (梯度下降法)

输入：目标函数$f(x)$，梯度函数$g(x) = \nabla f(x)$，计算精度ε；

输出：$f(x)$的极小点x^*

(1) 取初始值$x^{(0)} \in R^n$，置$k=0$

(2) 计算$f(x^{(k)})$

(3) 计算梯度$g_k = g(x^{(k)})$，当$\|g_k\| < \varepsilon$时，停止迭代，令$x^* = x^{(k)}$；否则，令$p_k = -g(x^{(k)})$，求λ_k，使 $f(x^{(k)} + \lambda_k p_k) = \min_{\lambda \geq 0} f(x^{(k)} + \lambda p_k)$

(4) 置$x^{(k+1)} = x^{(k)} + \lambda_k p_k$，计算$f(x^{(k+1)})$

当$\|f(x^{(k+1)}) - f(x^{(k)})\| < \varepsilon$或$\|x^{(k+1)} - x^{(k)}\| < \varepsilon$时，停止迭代，令$x^* = x^{(k+1)}$

(5) 否则，置$k = k+1$，转(3)