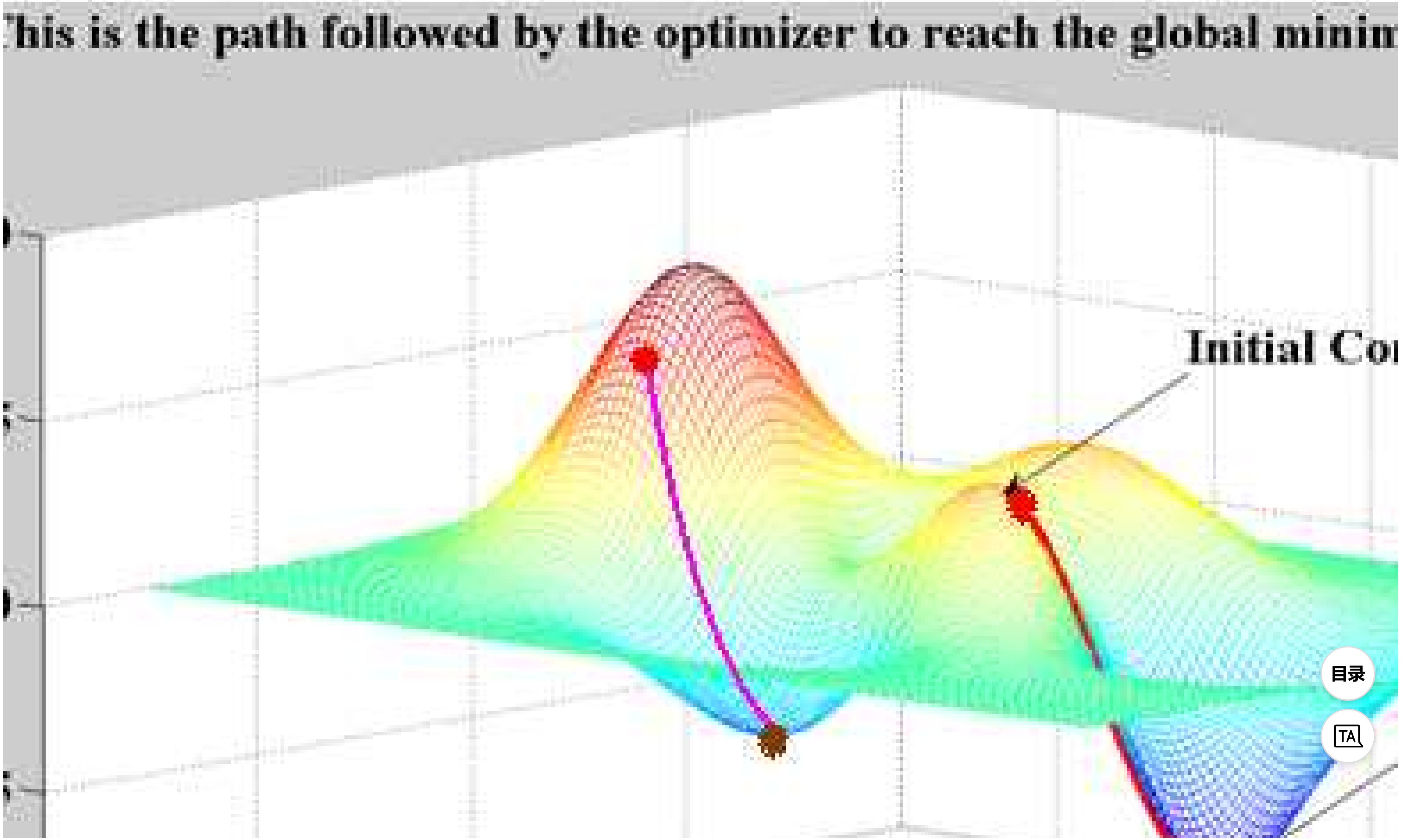
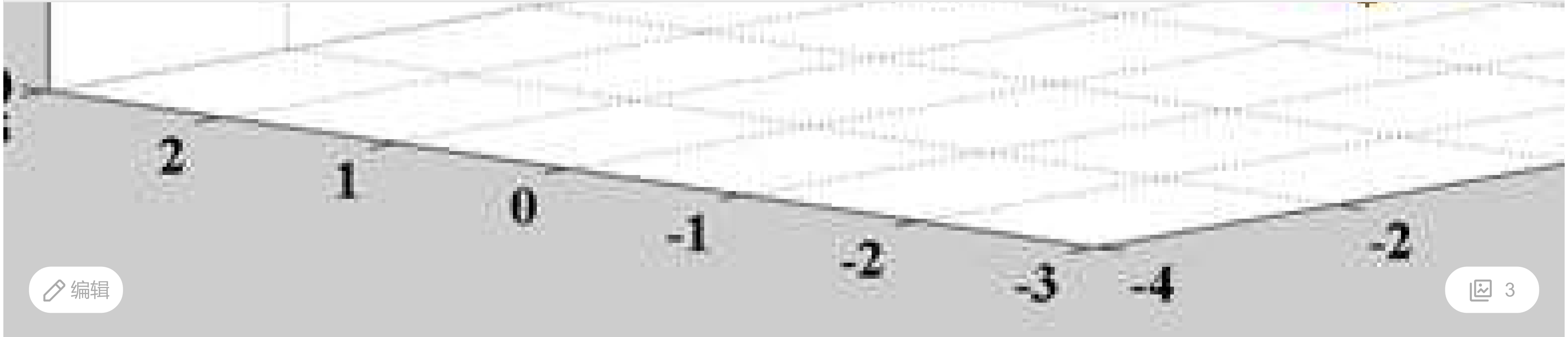

'his is the path followed by the optimizer to reach the global minim
Initial Co



编辑 3

梯度下降

科普中国
本词条由“科普中国”百科科学词条编写与应用工作项目审核
贡献者 孙和军

详情

梯度下降是迭代法的一种,可以用于求解最小二乘问题(线性和非线性都可以)。在求解机器学习算法的模型参数，即无约束优化问题时，梯度下降（Gradient Descent）是最常采用的方法之一，另一种常用的方法是最小二乘法。在求解损失函数的最小值时，可以通过梯度下降法来一步步的迭代求解，得到最小化的损失函数和模型参数值。反过来，如果我们需要求解损失函数的最大值，这时就需要用梯度上升法来迭代了。在机器学习中，基于基本的梯度下降法发展了两种梯度下降方法，分别为随机梯度下降法和批量梯度下降法。

目录

TA

中文名	梯度下降
外文名	steepest descent (gradient descent)
用于	求解非线性方程组
类型	最优化算法

简介

梯度：对于可微的数量场 $f(x,y,z)$ ，以 $\left(\frac{\partial f}{\partial x},\frac{\partial f}{\partial y},\frac{\partial f}{\partial z}\right)$ 为分量的向量场称为f的梯度或斜量。[1]

梯度下降法(gradient descent)是一个最优化算法，常用于机器学习和人工智能当中用来递归性地逼近最小偏差模型。

求解过程

顾名思义，梯度下降法的计算过程就是沿梯度下降的方向求解极小值（也可以沿梯度上升方向求解极大值）。

其迭代公式为 $a_{k+1}=a_k+\rho_k\bar{s}^{(k)}$,其中 $\bar{s}^{(k)}$ 代表梯度负方向， ρ_k 表示梯度方向上的搜索步长。梯度方向我们可以通过对函数求导得到，步长的确定比较麻烦，太大了的话可能会发散，太小收敛速度又太慢。一般确定步长的方法是由线性搜索算法来确定，即把下一个点的坐标看做是a_{k+1}的函数，然后求满足f(a_{k+1})的最小值的 即可。

因为一般情况下，梯度向量为0的话说明是到了一个极值点，此时梯度的幅值也为0.而采用梯度下降算法进行最优化求解时，算法迭代的终止条件是梯度向量的幅值接近0即可，可以设置个非常小的常数阈值。

应用

举一个非常简单的例子，如求函数 $f(x)=x^2$ 的最小值。

利用梯度下降的方法解题步骤如下：

1、求梯度， $\nabla=2x$

目录

TA

$x \leftarrow x - \gamma \cdot \nabla$ ，其中， γ 为步长。如果步长足够小，则可以保证每一次迭代都在减小，但可能导致收敛太慢，如果步长太大，则不能保证每一次迭代都减少，也不能保证收敛。

3、循环迭代步骤2，直到 x 的值变化到使得 $f(x)$ 在两次迭代之间的差值足够小，比如0.00000001，也就是说，直到两次迭代计算出来的 $f(x)$ 基本没有变化，则说明此时 $f(x)$ 已经达到局部最小值了。

4、此时，输出 x ，这个 x 就是使得函数 $f(x)$ 最小时的 x 的取值 。

MATLAB如下。

```
%%  最速下降法图示
%  设置步长为0.1，f_change为改变前后的y值变化，仅设置了一个退出条件。
syms  x;f=x^2;
step=0.1;x=2;k=0;                    %设置步长,初始值,迭代记录数
f_change=x^2;                            %初始化差值
f_current=x^2;                          %计算当前函数值
ezplot(@(x,f)f-x.^2)                %画出函数图像
axis([-2,2,-0.2,3])                 %固定坐标轴
hold  on
while  f_change>0.000000001                                %设置条件，两次计算的值之差小于某个数，跳出循环
        x=x-step*2*x;                                                  %-2*x为梯度反方向，step为步长，！最速下降法！
        f_change  =  f_current  -  x^2;                           %计算两次函数值之差
        f_current  =  x^2  ;                                           %重新计算当前的函数值
        plot(x,f_current,'ro','markersize',7)  %标记当前的位置
        drawnow;pause(0.2);
        k=k+1;
end
```

目录

梯度下降法处理一些复杂的非线性函数会出现问题，如Rosenbrock函数： $f(x,y)=(a-x)^2+b\left(y-x^2\right)^2$,其最小值在 $(x,y)=(1,1)$ 处，函数值为 $f(x,y)=0$ 。但是此函数具有狭窄弯曲的山谷，最小点 $(x,y)=(1,1)$ 就在这些山谷之中，并且谷底很平。优化过程是之字形的向极小值点靠近，速度非常缓慢。

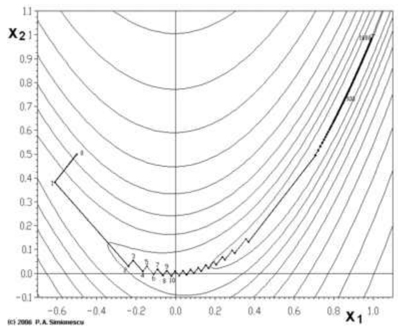


图1

缺点

靠近极小值时收敛速度减慢。

直线搜索时可能会产生一些问题。

可能会“之字形”地下降。

参考资料

猜你关注

excel入门视频教程

南昌到兰州的飞机票

求是村二手房

python