中档题

偏导数与梯度

1. **什么是偏导数?**

   偏导数是多变量函数中的一种导数，用于描述函数在某一个特定方向上的变化率，数学定义如下：

$$f(x_1,x_2,\ldots,x_n)\frac{\partial f}{\partial x_i}\bigg|_{(a_1,a_2,\ldots,a_n)}=\lim_{h\to 0}\frac{f(a_1,a_2,\ldots,a_i+h,\ldots,a_n)-f(a_1,a_2,\ldots,a_i,\ldots,a_n)}{h}$$

2. **什么是梯度?**

   梯度是一个向量，其每个分量是函数对该点个坐标轴方向的偏导数，其数学定义如下：

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

3. **偏导数与梯度的关系**
   - 梯度的分量是偏导数，且梯度向量指向了函数在该点增加最快的方向
   - 梯度向量的模表示函数在该方向上的最大变化率

梯度下降与反向传播

1. **什么是梯度下降?**

   梯度下降是一种常用的**优化算法**，用于最小化一个可微函数（注意，求的只是局部最小值，但若是严格的凹函数，则可以求到整体最小值）

   梯度下降通过迭代的方法沿着负梯度的方向更新参数，逐步逼近函数的局部最小值

2. **梯度下降的基本步骤：**

$$(1)\text{初始化参数：选定一个点}x_0$$

$$(2)\text{计算梯度：计算当前点}x_k\text{处的梯度}\nabla f$$

$$x_{k+1}=x_k-\alpha\nabla f_{x_k}$$

$$(3)\text{更新参数：根据计算出的梯度，按照一定的步长（学习率）沿梯度的负方向移动}$$

$$(4)\text{重复步骤}(2)\text{和}(3)\text{：直到找到极小值}$$

3. **什么是反向传播?**

   反向传播是一种在神经网络中用于计算损失函数关于网络权重的梯度的算法。

   使用链式法则，将梯度从输出层传递回输入层，然后逐层计算每一层的权重和偏置的梯度

什么是泰勒公式

泰勒公式是一种用于近似函数的方法，我们可以使用泰勒公式在某一点附近使用多项式来逼近，这个多项式称为泰勒多项式，它可以用来估计该函数在该点附近的值

泰勒公式的定义

$$\text{假设}f(x)\text{在点}a\text{的某个领域内有}n+1\text{阶连续导数，那么}f(x)\text{可以在}a\text{附近展开为}f(x)=f(a)+f'(a)(x-a)+\frac{f''(a)}{2!}(x-a)^2+\cdots+\frac{f^{(n)}(a)}{n!}(x-a)^n$$

矩阵和向量的关系

向量是一维的，只有一个方向（行或列），而矩阵是多维的，有多个行和列。所以矩阵类似于向量的推广，一个矩阵能储存比一个向量更多的数据和信息。

因而向量主要用于表示单个数据点或特征，而矩阵可以用于解线性方程组、线性变换、机器学习中的数据处理等

在机器学习中的应用

1. **偏导数**和**梯度**主要使用于梯度下降算法中，在梯度下降算法中，我们需要求某个自变量的偏导数，并用梯度储存某点的所有偏导数
2. **矩阵可以用来储存数据**，在吴恩达的视频中讲到，面对大量的数据，倘若使用一般方法——循环来完成，则计算机一个一个计算，从 1 到 n ，但倘若使用矩阵，则计算机会直接通过硬件来计算，所有步同时进行，**大大提升了程序的效率和速度**
3. **泰勒公式**可以用于拟合复杂函数，简化计算

   比如用泰勒多项式近似某些复杂的激活函数，如sigmoid和Tanh，以简化计算

   在优化理论中，利用二阶泰勒展开分析函数的凹凸性，设计有效的优化算法