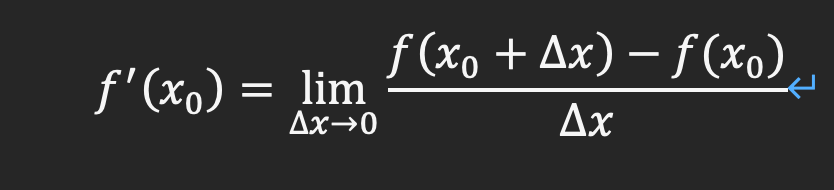
机器学习核心

## 高等数学

### 导数

#### 导数的基本概念

1. 微积分中的概念，是函数在这一点附近的变化斜率（切线斜率）；
2. 本质是通过极限的概念对函数进行局部的线性逼近；
3. 公式：



#### 基本函数的导数

|  |  |  |
| --- | --- | --- |
| **说明** | **公式** | **例子** |
| **常数的导数** |  |  |
| **幂函数的导数** |  |  |
| **指数函数的导数** |  |  |
|  | — |
| **对数函数的导数** |  |  |
|  | — |
| **三角函数的导数** |  | — |
|  | — |
|  | — |
|  | — |

注解：ln2 : 以e为底数的2的log对数

#### 1.13 导数的求导法则

|  |  |
| --- | --- |
| **说明** | **公式** |
| **两函数之和求导** |  |
| **两函数之积求导** |  |
| **两函数之商求导** |  |
| **复合函数的导数** |  |

#### 利用导数求极值

驻点：函数导数为零的极值

鞍点：函数导数为零的非极值

#### 二阶导数

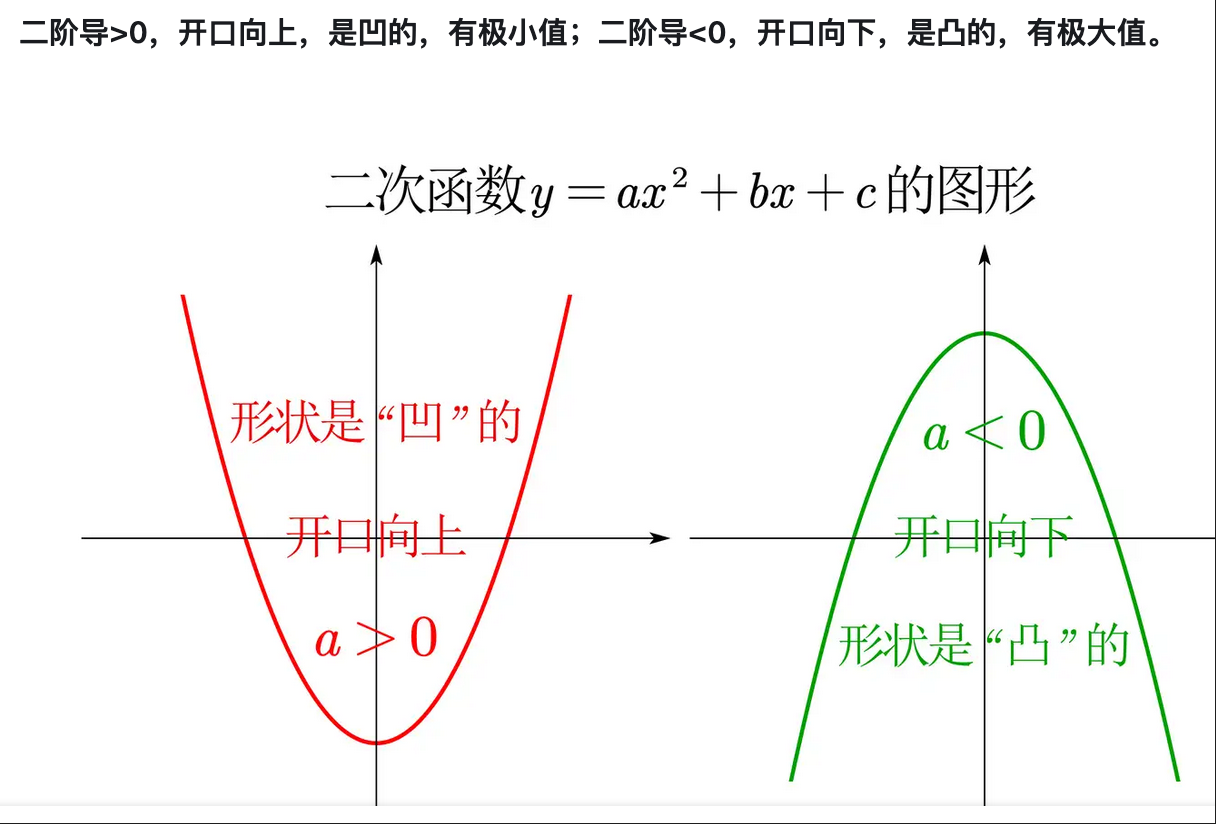
表达方式：、或

理解：位移时间函数、速度时间函数（一阶导）、加速度时间函数（二阶导）

二阶导数与函数的凹凸的关系：

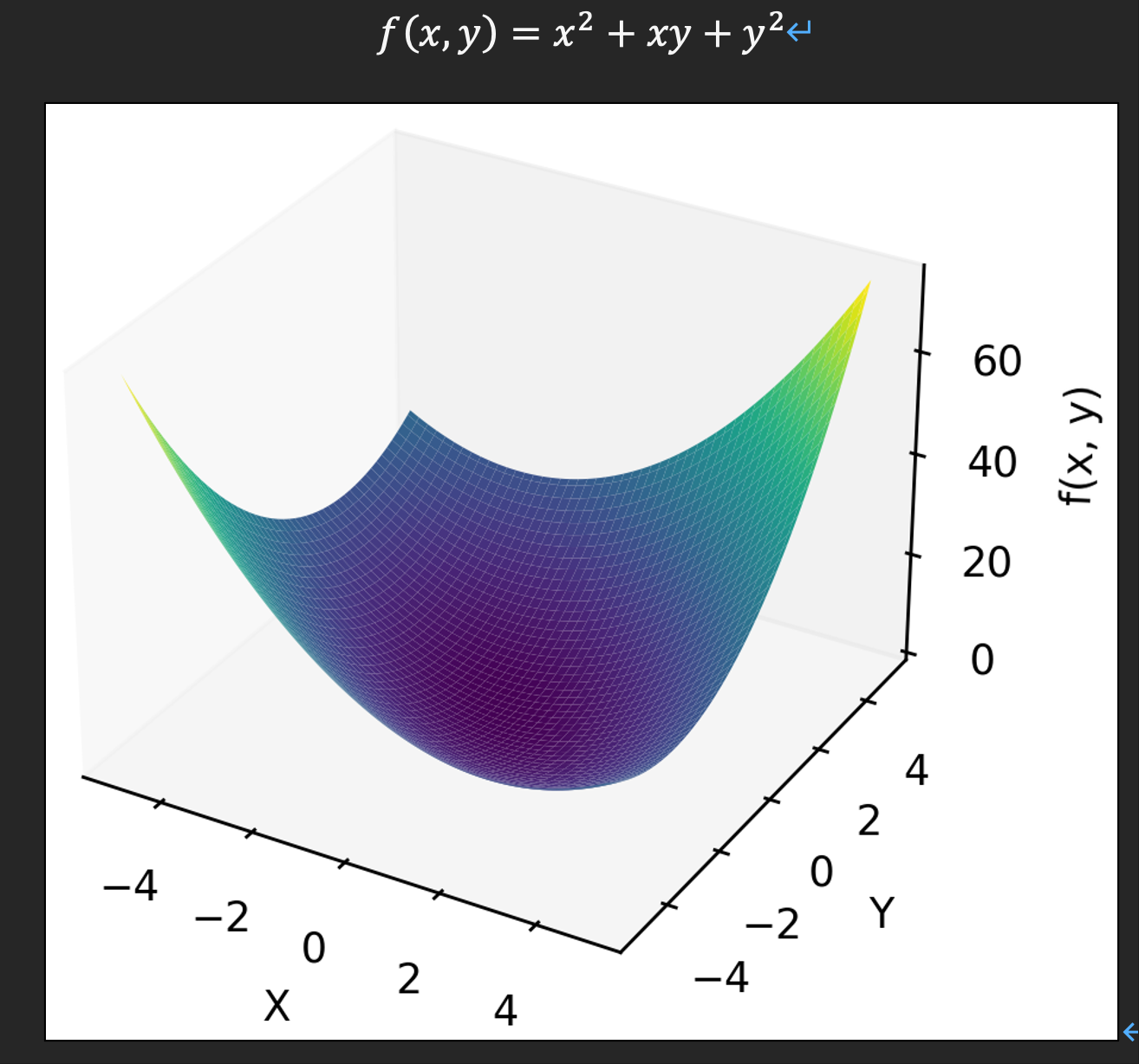
* 若，，则在取得极大值。
* 若，，则在取得极小值。
* 若，，则该点可能是拐点，也可能是极大值点或极小值点。

注释：标准的二元一次方程式



### 偏导与梯度

#### 1.2.1偏导



可将其中一个元素看作常数，此时可看作关于另一元素的函数。

在固定的情况下，可计算关于的导数：

这种导数称为偏导数，一般记作：

更一般地来说，一个多元函数在点处对的偏导数定义为：

注：

对x的偏导，就是对y定死了（即在垂直y的截面上）求x的变化率

对y的偏导，就是对x定死了（即在垂直x的截面上）求y的变化率

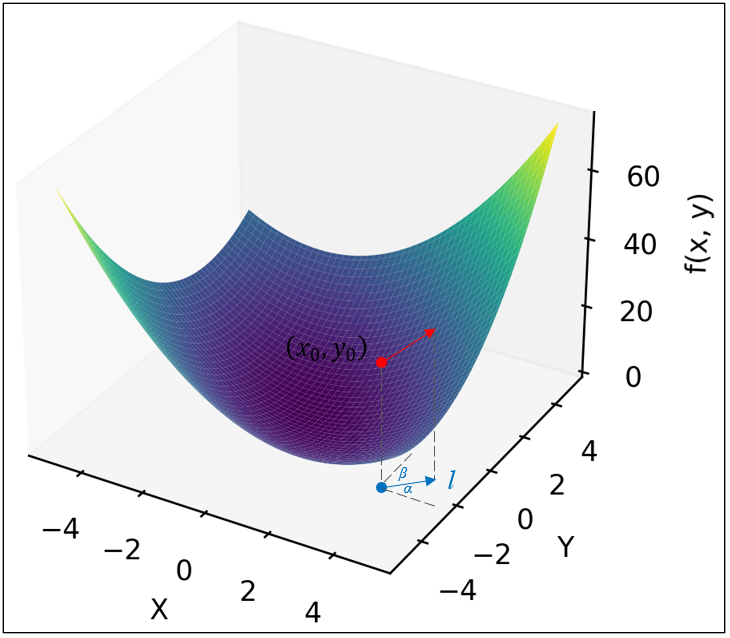
总结也就是偏导数可以看做多远函数f沿着自变量轴方向的变化率

#### 1.2.2 方向导数

基本概念：选取任意方向l,在某个点（x0,y0）处，二元函数f（x,y）沿着这个方向的变化率

这里， 就是沿方向*l*的微小改变量， 和 与 的关系为：

，



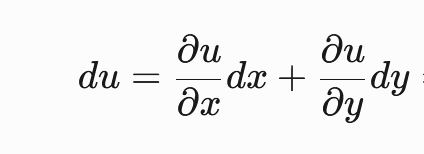
根据全微分公式，上式可以表示为：

其中 表示点处*f*对*x*、*y*的偏导数；是方向*l*的方向余弦，即*l*方向的单位方向向量可以表示为。

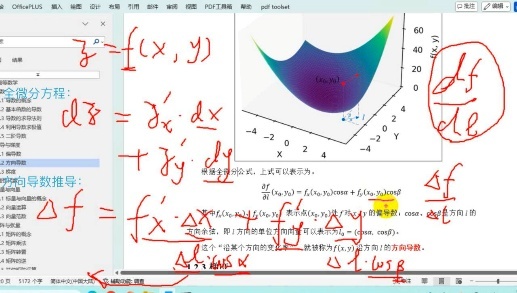
这个“沿某个方向的变化率”，就被称为沿方向*l*的**方向导数**。

注：

1. 全微分方程



1. 具体推导过程



#### .2.3 梯度

多元函数关于每个变量都有偏导数，在点处，这些偏导数定义了一个向量。

这个向量称为在点的梯度。

例如：在处的梯度为。

(1,1)梯度就是：

将x看作常数求的导数为：x+2y；此时值为3；

将y看作常数求的导数为：2x+y；此时的值为3；

注：梯度本身代表就是多维空间中，走势最快的坡度；

# 线性代数

## 标量与向量

标量（scalar）：一个单独的数，只有大小；

向量（verctor）：有标量组成，有大小和方向；常见的有行向量和列向量，他们是转置的关系；

注释：

具体可以查看代码操作：

F:\MyStudy\LLM\my\_codes\my\_study\_llm\_training\fundamentals\_of\_mathematics\Derivative.ipynb 中搜索“=向量的计算=

**”**

## 2.2矩阵与张量

矩阵的乘法：

矩阵乘法满足结合律、左分配律和右分配律。不满足交换律即。

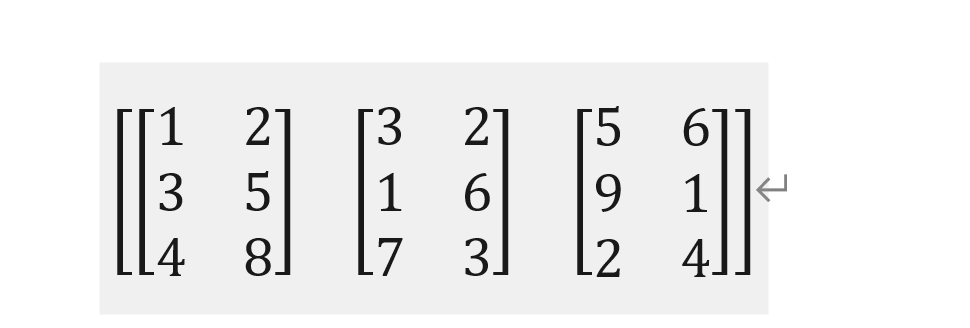
结合律：若，则

左分配律：若，则

右分配律：若，则

**张量：**

张量（tensor）可视为多维数组，是标量，1维向量和2维矩阵的n维推广。



注释：

具体可以查看代码操作：

F:\MyStudy\LLM\my\_codes\my\_study\_llm\_training\fundamentals\_of\_mathematics\Derivative.ipynb 中搜索“==矩阵与张量=**”**

## 矩阵求导

### 2.2.1基本概念

矩阵求导本质就是函数对变元的每个元素逐个求导，只是写成了向量或者矩阵的形式；

符号的统一：

为实向量变元。

为实矩阵变元。

为实标量函数，其变元为实向量。

为实标量函数，其变元为实矩阵。

为实向量函数，其变元为实向量。

为实向量函数，其变元为实矩阵。

当然，函数也可以是矩阵形式***F***，可以看作是向量函数的扩展。

注：

符号的逻辑：

x(小写未加粗的x)：实标量变元；

**x**(小写加粗的**x**)：实向量变元；

**X**（大写加粗的**X**）：实矩阵变元；

f(小写未加粗的f)：实标量函数；

**f**(小写加粗的**f**)：实向量函数；

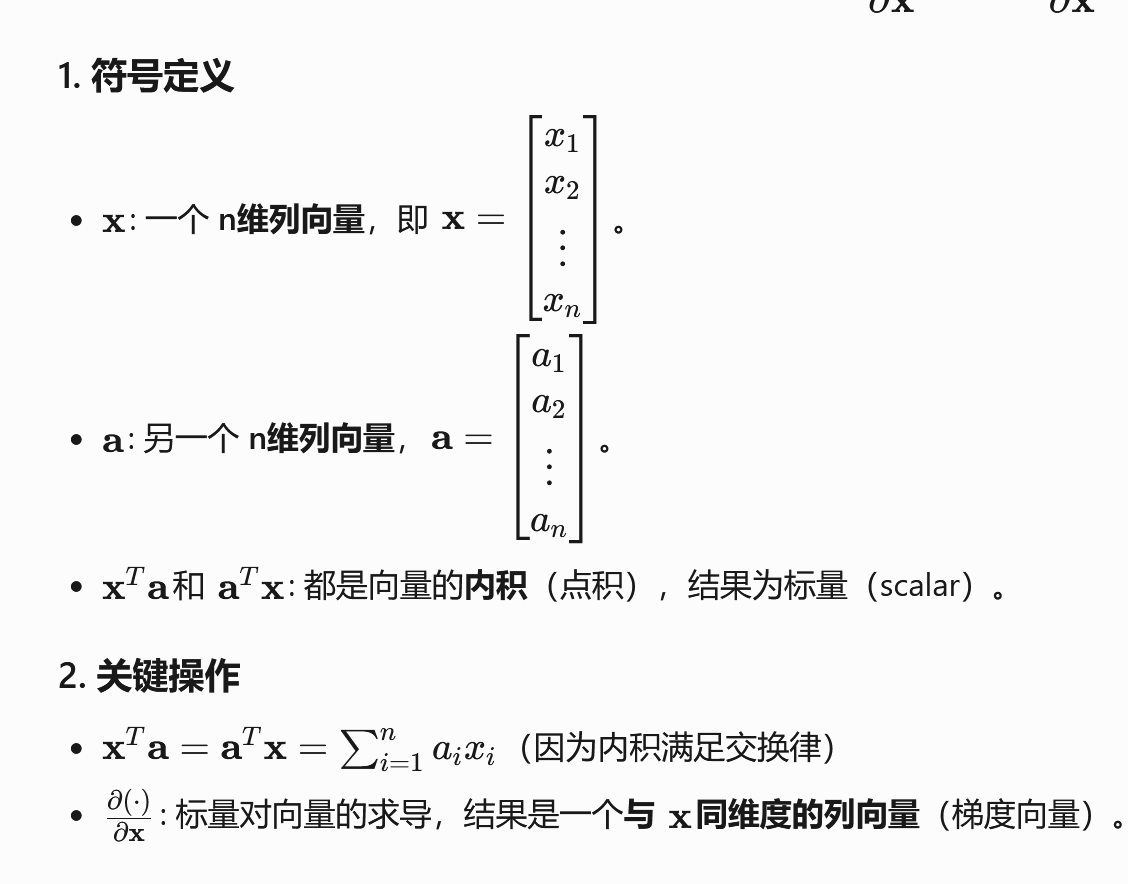
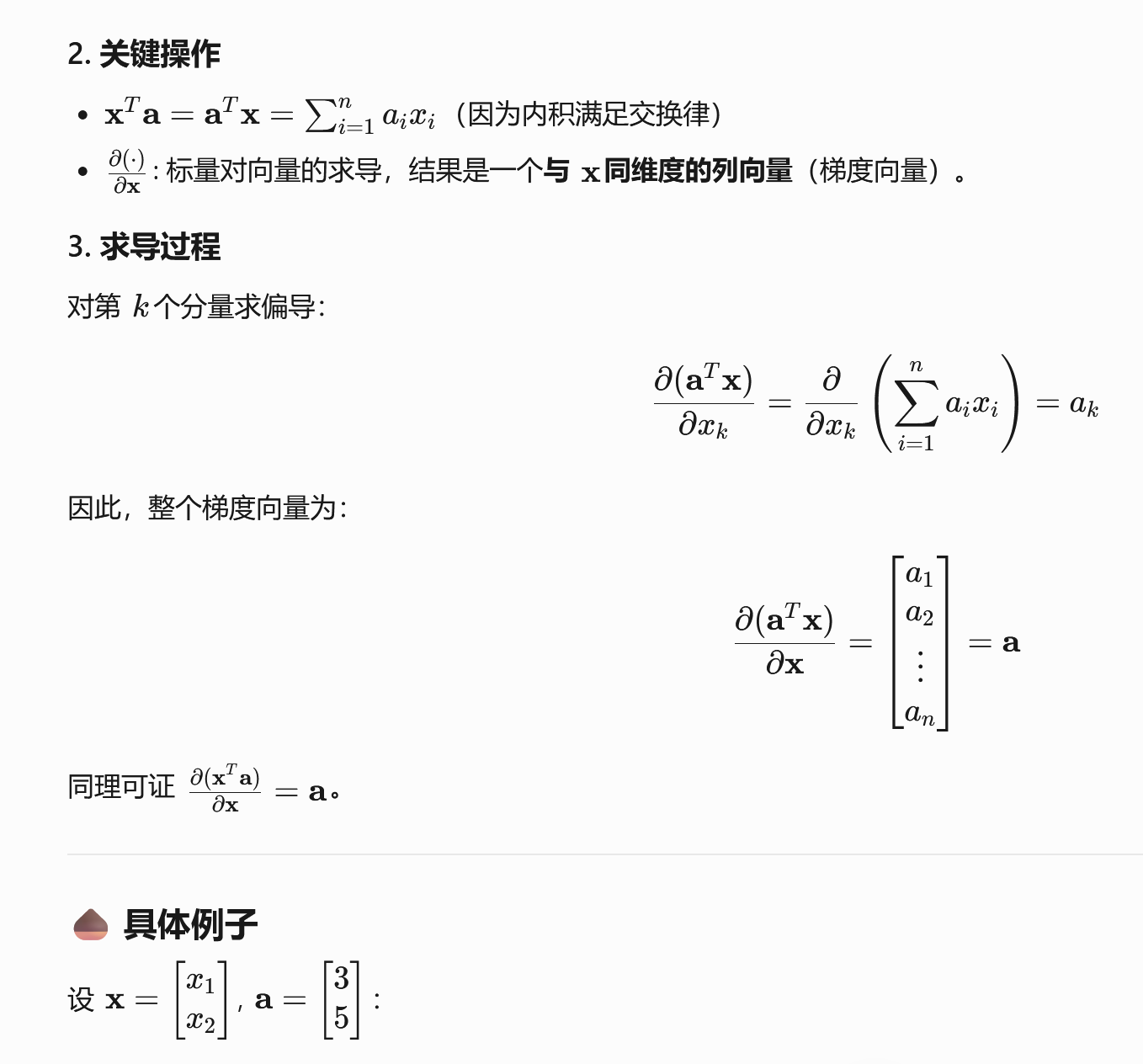
F（大写加粗的**F**）：实矩阵函数；

### 2.2.2 重点内容

1. 一般常见的f(x)=2x+4 其实就是标量变元的标量函数；

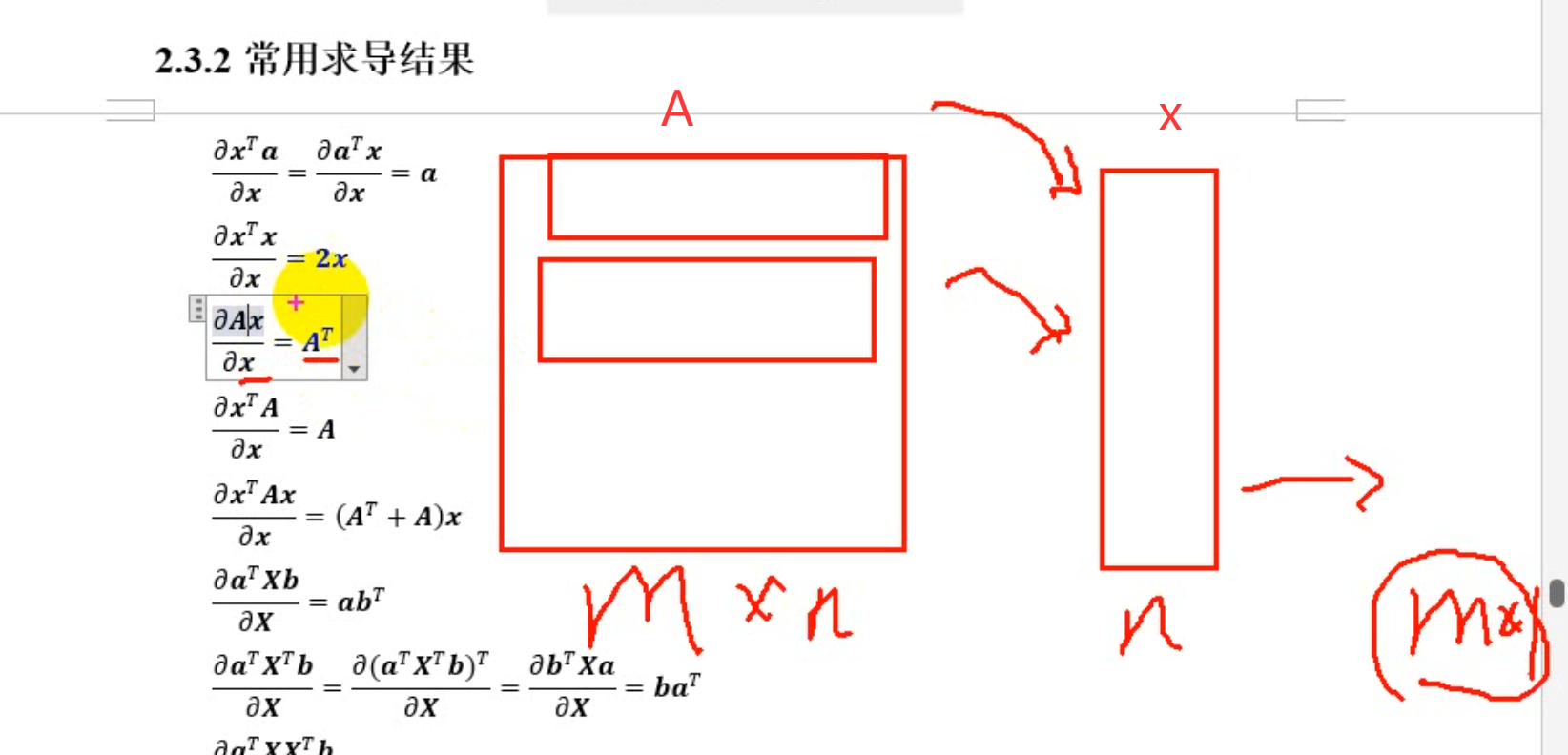
2. 常见的矩阵偏导公式

**解释：**

**解释：可类比 的导数是2x**

解释：常量矩阵**A**对变元向量**x**求导



解释：

**解释：可类比 ,但是矩阵乘法不符合乘法交换律，即 因此导数结果还不能完全实2Ax,而是**

1. **梯度矩阵**

对于实向量变元，实标量函数的梯度向量，为的列向量（与形状相同）：

对于矩阵变元，可以类似地得到的梯度矩阵：

类似地， 的二阶偏导构成的矩阵被称为“黑塞矩阵”（Hessian Matrix）：

本质就是：矩阵求导，得到的结果就是**梯度向量**或者**梯度矩阵**，其实就是对每一个变元位置的求导；

具体可以查看代码操作：

F:\MyStudy\LLM\my\_codes\my\_study\_llm\_training\fundamentals\_of\_mathematics\Derivative.ipynb 中搜索“=矩阵求导之梯度矩阵=**”**

### 2.2.3 重点总结

1. 后期我们更多研究的是： 实向量变元**x，**或者实矩阵变元**X(**m x n**)**的 标量函数f**(x)**或者**f(X)**;

因为后期最多的问题是在众多因素x中求出一个最大或者最小值，也就是只要一个结果；

2. 重点关注矩阵求导引出的梯度矩阵的概念

# 3. 概率论

## 3.1 概率

|  |  |
| --- | --- |
| **事件** | **概率** |
| **A** |  |
| **非A** |  |
| **A和B**  **（联合概率）** |  |
| **A或B** |  |
| **B的情况下A的概率**  **（条件概率）** |  |

重点：

1. 联合概率 、条件概率、或概率；
2. 联合概率和条件概率的关系；
3. 具体理解可以看文档中的例子；

## 概率分布

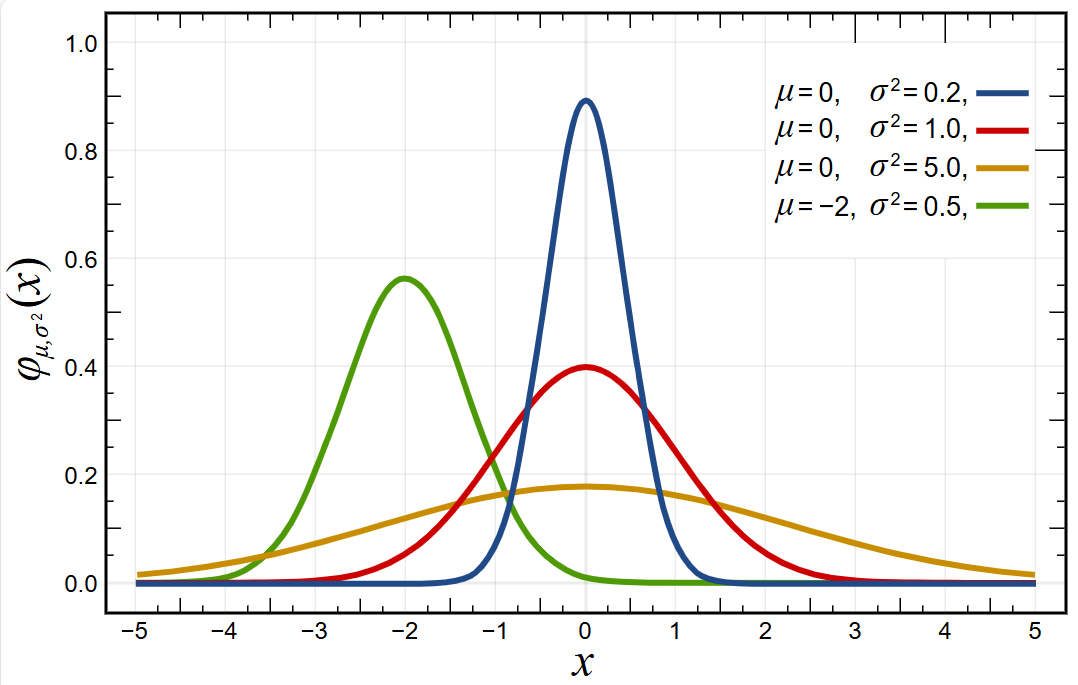
### 均匀分布

1. 均匀分布的概率密度函数可写为：

### 正态分布

若随机变量服从一个平均数为、标准差为（）的正态分布，则记为，其概率密度函数

正态分布的期望可解释为位置参数，决定了分布的位置；其方差可解释为尺度参数，决定了分布的幅度



## 贝叶斯定理

### 全概率公式

两两互斥的简单事件：A1、A2、…、An

事件A1发生的概率：

在事件A1的条件下B发生的概率：

事件A1和B的联合概率：

注：

联合概率就是同时发生的概率；

### 贝叶斯定律

* ：后验概率，发生的情况下发生的概率。
* ：似然概率，发生的情况下发生的概率。
* ：先验概率，发生的概率。
* ：发生的概率，通常通过全概率公式计算。

注：

1. 贝叶斯定律核心就是已知结果推论原因，即在结果概率B的基础上，计算原因事件A发生的概率；

## 似然函数

### 似然函数概念说明

概率：用于在已知一些参数的情况下，预测接下来的观测上所得到的结果；概率思维：从因推果（参数已知 → 预测结果）；

似然：用于在已知某些观测所得到的结果是，对有关事物之性质的参数进行估值；似然思维：从果推因（结果已知 → 反推参数）​​

似然函数：实一个关于参数的函数，而不是关于数据的函数；

似然函数的的定义：

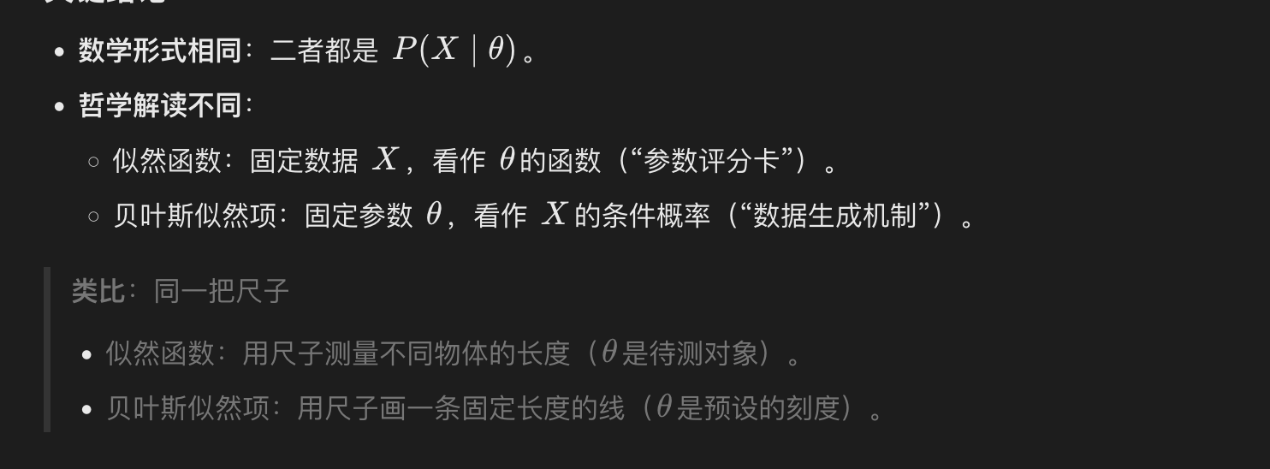
如服从参数为的某个分布，那么似然函数可以写作：

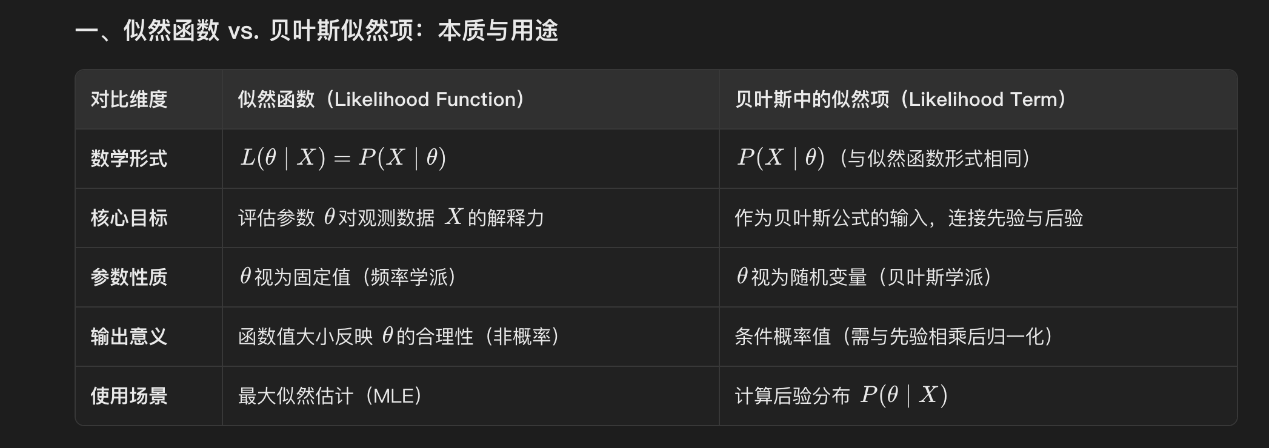
注：都是独立事件

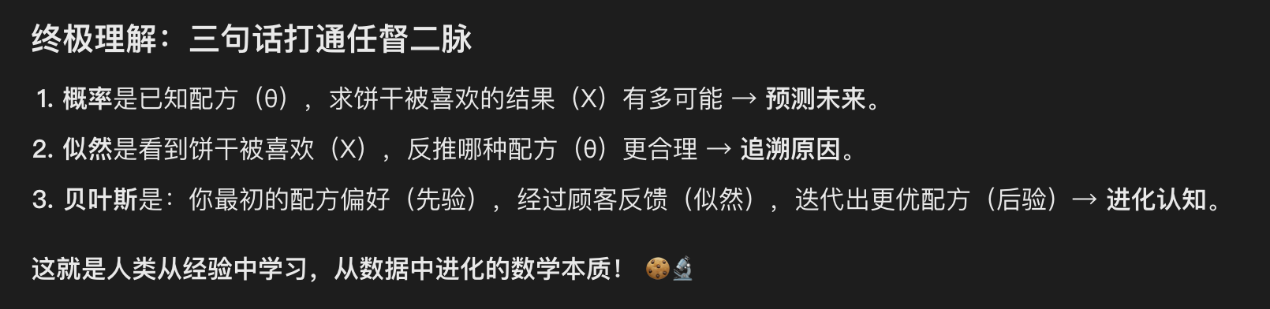
针对其中存在的乘法，可以使对数函数将其转化为加法：

注：

似然函数和贝叶斯定律中的似然项目的区别：







### 极大似然估计

似然函数常用于极大似然估计。我们希望找到使似然函数最大化的参数。这意味着在给定观测数据的情况下，选择最可能生成这些数据的参数值

注：

乘法对数化后，转化成加法，然后求导；

导数的值为零时概率是多少？

如下案例：

例如，掷硬币3次，2次正面1次背面，能否依据此结果逆推出正面的概率；正面概率为0.5的概率为多少、正面概率为0.6的概率为多少；最有可能的正面概率是多少？

我们用代表硬币正面朝上的概率，用代表2次正面1次背面的结果

注：代表“2次正面1次背面“会有3种可能的组合： HHT HTH THH

当正面概率为0.5时：

当正面概率为0.6时：

为了找出极大似然估计，对似然函数取对数并求导，使其等于0

解得，意味着当掷硬币3次，出现2次正面1次背面的结果时，硬币正面朝上的概率最有可能为。