**深度学习**

V. 1.0.1

备注：本文档大量常考《Deep Learning》（[美] Ian Goodfellow ,[加] Yoshua Bengio,[加] Aaron Courvile 著），

《深度学习入门-基于Python的理论与实现》（[日] 斋藤康毅 著），请勿用做商业。

注：如涉及侵权，务必联系 [leep233@foxmail.com](mailto:leep233@foxmail.com) . 本人将立即整改；

目的：本文档只用作记录，分析以及学习，并无任何商业目的

目录

[线性代数 2](#_Toc106382294)

[概念 2](#_Toc106382295)

[导数与偏导数 2](#_Toc106382296)

[积分 5](#_Toc106382297)

[向量与矩阵 5](#_Toc106382298)

[范数 5](#_Toc106382299)

[运算 7](#_Toc106382300)

[特殊向量与矩阵 7](#_Toc106382301)

[概率 8](#_Toc106382302)

[随机变量 8](#_Toc106382303)

[随机变量与概率分布 8](#_Toc106382304)

[边缘概率 8](#_Toc106382305)

[条件概率 8](#_Toc106382306)

[独立性与条件独立 8](#_Toc106382307)

[期望 9](#_Toc106382308)

[机器学习 9](#_Toc106382309)

# 线性代数

## 概念

**实数（real number）**

**有理数和无理数的总称。**

**用表示实数集。**

**标量（scalar）**

**表示一个具体实数。**

**使用小写斜体来表示（如：标量 *a* ）**

**向量（vector）**

**表示一列数。**

**使用小写加粗来表示（如：向量 x）**

注意: 假设有一个向量 x , 这个向量里面有 *n* 个元素，并且每个元素属于实数，那么我们可以记作 ; **其中向量里面的元素使用标量+下角标的形式表示**

**矩阵（matrix）**

**矩阵是一个二维数组。**

**使用加粗大写字母表示（如：假设矩阵 A ，是一个m 行 n 列 的矩阵 我们可以记作 ；如果我们需要表示A 中 第 i 行的 j 列 的元素 我们可以记作 ；）**

注意: 假设矩阵 A ，是一个m 行 n 列 的矩阵 我们可以记作 ；

如果我们需要表示A 中 第 i 行的 j 列 的元素 我们可以记作 ；

表示第i 行的所有列;

表示第 j 列的所有行;

## 导数与偏导数

导数（derivative）

**如下图所示，小明开车去旅游，图中是某段事件内，车速s 和 时间 t 对应的函数关系图。求t1到t2 时间里，小明开车的平均速度？**

**S**

**t**



t1

t2

s1

s2

**我们很简单可以得出小明t1~t2之间的平均速度为 ;**

**如果我们将上图中，s 和 t ，使用函数关系坐标系来表示 s = f(x) , t = x;**

**那么函数关系图如下：**

**x**



**通过上图我们用来代替轴，那么，之间的**平均变化率, 记作 ;

**通过上面求平均变化率的例子，当 无线趋近 时，即 ) , 那么求出来的就是**导数**，记作**， 或 ;

**导数的几何意义:在几何数学中，导数即时该点的斜率，同时我们可以通过点斜式来求得关于过目标点的切线方程。**

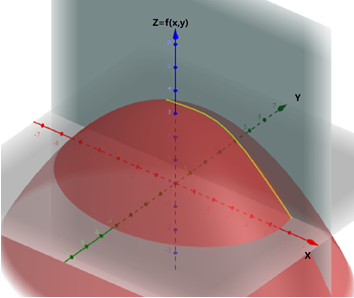
**点斜式：已知目标点（ , ）,斜率 *k*, 通过点斜式可以求出目标点的切线方程，即 ;**

**斜率与函数单调性：当斜率k>0时，表示函数单调递增，k<0时，函数单调递减，当k ≈ 0，时表示到达了函数的极值或鞍点**

偏导数（partial derivative**）**

**当函数的具有多个自变量（输入参数，）求其中一个的导数时，我们称为偏导数**

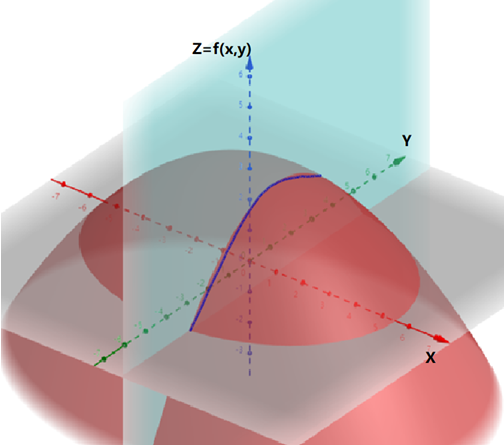
设函数 ,我们求关于函数 的偏导数时，需要将的值固定，的值来求偏导，记作 ,通过坐标系表示如下图



,

由图可以了解到求 关于 的偏导,其实就是求图中 在黄色线中的导数；

同样的如果我们需要求 ，我们只需要固定的值，就可以求出点的导数。如下图



,

## 积分

下图时关于函数f(x) 和 x 的函数关系图，求a,b区间内（阴影区域）的面积；

a

b



如图可知，由于我们的线性关系时曲线所以我们没有办法通过寻常求面积的方式求得阴影的面积，那么我们可以使往一个方向移动极小的距离，同时我们可以也可以得出移动 距离后的对应的函数的值，记作，那么我们可以近似这小块的面积等于 ,只要一直将这个区间内所有的小块面积相加就可以得到整个大块阴影的面积，我们记作;

同时如果我们的函数是关于多个自变量求积分则；如,需要求在 区域内的积分记作

## 向量与矩阵

### 范数

在机器学习中，我们使用范数来衡量向量的大小；

**范数:**

**, 其中 ,其实范数是衡量向量x和原点（所有元素都是零的点）之间的距离。我们使用函数来代表范数即，则范数需要满足以下条件**

* **正定性：**
* **三角不等式：**
* **正值齐次性：**

**范数**

**机器学习中，我们需要区分零与非零但是值很小的元素，这种情况下，我们使用各个位置的斜率相同，同时保持简单的数学形式的函数：范数**

**机器学习中，零和非零元素之间的差异非常重要时，通常使用范数。**

**范数（欧几里得范数）**

**当时，称为范数，也叫做欧几里得范数，他表示从原点出发到向量x点的欧几里得距离。由于此范数在机器学习中使用的非常频繁所有我们简写成|| x || , 同时经常用来衡量向量的大小，**

**范数（最大范数）**

**在机器学习中，也经常出现最大范数，它表示向量中具有最大幅值得元素得绝对值；**

**Frobenius范数**

**在机器学习中，当我们需要对矩阵大小进行衡量得时候，我们通常使用Frobenius范数**

### 运算

**转置**

**矩阵的转置即以主对角线（主对角线：从左上角到右下角的对角线）为轴的镜像，即矩阵的行变列，列变行。**

**矩阵的转置为，定义如下：;**

**标量可以看作一个1x1的矩阵，所以标量的转置等于自身 ；**

**向量可以看作是一个列的矩阵，它转置之后变成一个一行的矩阵 ();**

**加减法**

**矩阵与矩阵相加相当于每个对应的元素相加。如;**

**标量与矩阵相加或相乘时，相当于矩阵的每个元素乘以或加上这个标量。**

**如 ;**

**深度都学习中，我们允许矩阵与向量相加，产生一个新的矩阵。如,其中 ; 换言之，将向量b 与 矩阵 A 的每一行相加，这种隐式将向量b 复制到很多位置的方式叫做广播**

**乘积**

**Hadamard乘积**

**轨运算**

### 特殊向量与矩阵

**单位向量**

**方阵**

**矩阵的行数和列数相等时，我们称为方阵**

**单位矩阵**

**所有沿主对角线得元素都是1，而且其他位置得元素都是0的方阵叫做单位矩阵。任意向量和单位矩阵相乘都不会进行任何改变。**

**通常我们使用 来表示一个n行n列的单位矩阵。下图为 矩阵：**

**逆矩阵**

**矩阵，是否存在逆矩阵必须满足公式：**

**对角矩阵**

# 概率

### 随机变量

**离散型**

**连续型**

### **随机变量与概率分布**

**离散型变量和概率质量函数（probability mass function , PMF）**

**连续型变量和概率密度函数（probability density function , PDF）**

### **边缘概率**

### 条件概率

### 独立性与条件独立

### 期望

# 机器学习