数学基础

向量(vector)

在数学中，向量（也称为欧几里得向量、几何向量、矢量），指具有大小（magnitude）和方向的量。它可以形象化地表示为带箭头的线段。箭头所指：代表向量的方向；线段长度：代表向量的大小

向量在不同的领域中有不同的概念。

- 物理上，向量是空间中的箭头
- 计算机里，向量是有序的数字列表
- 数学上，向量是有大小和方向的量，可加可乘。

一个n维向量x的表达式可写成:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ . \\ . \\ . \\ x_n \end{bmatrix}$$

其中$x_1, ..., x_n$是向量的元素，将各元素均为实数的n维向量x记作$x \in R_n$或$x \in R_{n*1}$

标量(scalar)

亦称“无向量”。有些物理量，只具有数值大小，而没有方向，部分有正负之分。

矩阵(Matrix)

在数学中，矩阵（Matrix）是一个按照长方阵列排列的复数或实数集合，最早来自于方程组的系数及常数所构成的方阵。
一个m行n列矩阵的表达式可写成

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}$$

其中x_{ij}是矩阵X中第i行第j列的元素$(1 \le i \le m, 1 \le j \le n)$，将各元素均为实数的$m$行$n$列矩阵$X$记作$X \in \mathbb{R}_{m*n}$.

> 向量是特殊的矩阵

张量

导数(Derivative)

导数（Derivative）是微积分中的重要基础概念。当函数$y = f(x)$的自变量x在一点x_0上产生一个增量Δx时，函数输出值的增量Δy与自变量增量Δx的比值在Δx趋于0时的极限a如果存在，a即为在x_0处的导数，记作$f_,(x0)$或$df(x0)/dx$。

导数是函数的局部性质。一个函数在某一点的导数描述了这个函数在这一点附近的变化率。如果函数的自变量和取值都是实数的话，函数在某一点的导数就是该函数所代表的曲线在这一点上的切线斜率。导数的本质是通过极限的概念对函数进行局部的线性逼近。

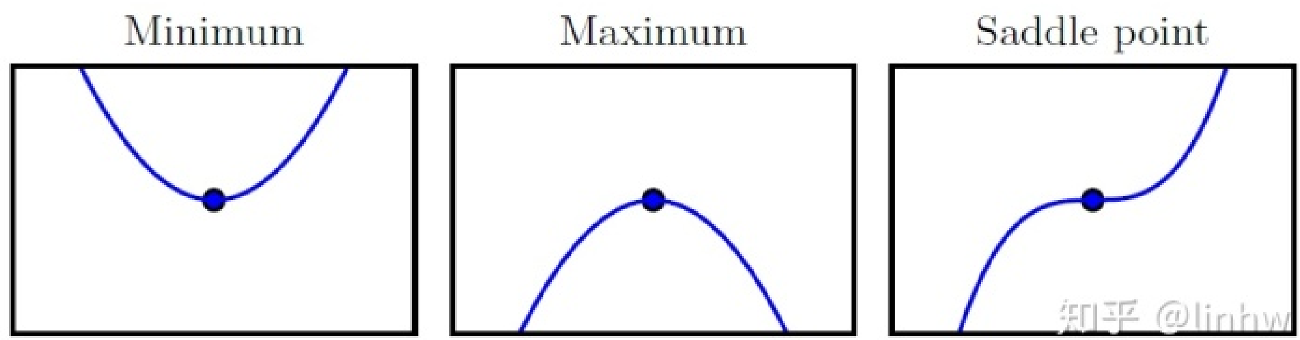
机器学习中，导数：函数的变化率（变化速度）

单变量导数的一阶导

对于单变量来说，其导数的计算方法如下：

$$f_,(x) = \lim_{h->0} \frac{f(x+h)-f(x)}{h}$$

导数代表的意义是函数在x处的变化率，当导数为0时x可能为函数的鞍点或者极值点。



单变量函数的二阶导数

二阶导数代表的意义是函数在x点处导数的变化率
二阶导数的计算方式如下：

$$f''(x) = \lim_{h->0} \frac{f'(x+h) - f'(x)}{h}$$

二阶导数为0，说明该点的导数变化率为0

偏导数

当一个函数有多个变量，而你只想计算函数与某个变量的变化关系时，就需要计算偏导数。
设u为一个有n个自变量的函数，$u = f(x1, x2, ..., xn)$，它有关第i个变量x_i的偏导数为:

$$\frac{\theta u}{\theta x_i} = \lim_{h->0} \frac{f(x_1, ...x_{i-1}, x_i + h, x_{i+1}..., x_n) - f(x_1, ..., x_i, ..., x_n)}{h}$$

为了计算$\partial u / \partial x_i$，只需将$x_1, ..., x_{i-1}, x_{i+1}, ..., x_n$视为常数并求$u$有关$x_i$的导数

梯度

假设$f : \mathbb{R}_n - > \mathbb{R}$的输入维度是一个$n$维向量$x = [x_1, x_2, ..., x_n]_T$，输出是标量。

函数$f(x)$有关x的梯度是一个由n个偏导数组成的向量

$$\nabla_x f(x) = \left[\frac{\theta f(x)}{\theta x_1}, \frac{\theta f(x)}{\theta x_2}, \ldots, \frac{\theta f(x)}{\theta x_n} \right]_T$$

在机器学习的学习中，多元函数的所有偏导数构成的向量即为梯度。
梯度的本意是一个向量，表示某一函数在该点处的方向导数沿着该方向取得最大值，即函数在该点处沿着该方向（此梯度的方向）变化最快，变化率最大，为该梯度的模。