大数定律：

[切比雪夫](https://baike.baidu.com/item/%E5%88%87%E6%AF%94%E9%9B%AA%E5%A4%AB)大数定理 ： 随着[样本容量](https://baike.baidu.com/item/%E6%A0%B7%E6%9C%AC%E5%AE%B9%E9%87%8F)n的增加，[样本](https://baike.baidu.com/item/%E6%A0%B7%E6%9C%AC)平均数将接近于[总体](https://baike.baidu.com/item/%E6%80%BB%E4%BD%93)平均数。从而为统计推断中依据样本平均数估计总体平均数提供了理论依据。

[伯努利](https://baike.baidu.com/item/%E4%BC%AF%E5%8A%AA%E5%88%A9)大数定律 ： 当n足够大时，事件A出现的频率将几乎接近于其发生的概率，即频率的稳定性。

辛钦大数定律 ：用算术平均值来近似实际真值是合理的，用**算术平均值来估计数学期望**就是根据此定律

数学基本知识点：

<https://www.zhihu.com/question/59946227/answer/771078151>

如何理解拉格朗日乘子法和KKT条件？

<https://blog.csdn.net/ccnt_2012/article/details/81811903>

卡特兰数

出栈次序

一个栈(无穷大)的进栈序列为1，2，3，…，n，有多少个不同的出栈序列

C（2n,n）-C(2n,n+1) = C(2n,n) / (n+1)

基尼系数

定义：基尼指数（基尼不纯度）：表示在样本集合中一个随机选中的样本被分错的概率。

 基尼指数（基尼不纯度）= 样本被选中的概率 \* 样本被分错的概率

注意： Gini指数越小表示集合中被选中的样本被分错的概率越小，也就是说集合的纯度越高，反之，集合越不纯。

<https://blog.csdn.net/qq_41430142/article/details/90905086>

ID3 信息增益

C4.5 信息增益率

CART 基尼指数

决策树与信息熵

用来描述随机事件的“混乱”程度，也即该随机事件所有结果所带来平均不确定性；显然，我们可以看出**信息熵的计算就是信息量的数学期望**

<https://www.jianshu.com/p/3ad5d3ff9a9e>

LASSO 回归 L1正则 可以使部分参数权重为0，达到降维的效果

岭回归 L2正则

朴素贝叶斯

贝叶斯分类法是统计学分类方法，它可以预测类隶属关系的概率，如一个给定元组属于一个特定类的概率。贝叶斯分类基于贝叶斯定理。朴素贝叶斯分类法假定一个属性值在给定类上的概率独立于其他属性的值，这一假定称为类条件独立性。

贝叶斯原理、贝叶斯分类和朴素贝叶斯这三者之间是有区别的

贝叶斯公式（定理）

贝叶斯原理是最大的概念，它解决了概率论中“逆向概率”的问题，在这个理论基础上，人们设计出了贝叶斯分类器，朴素贝叶斯分类是贝叶斯分类器中的一种，也是最简单，最常用的分类器。朴素贝叶斯之所以朴素是因为它假设属性是相互独立的，因此对实际情况有所约束，\*\*如果属性之间存在关联，分类准确率会降低。\*\*不过好在对于大部分情况下，朴素贝叶斯的分类效果都不错。

<https://blog.csdn.net/qiu_zhi_liao/article/details/90671932>

生成式模型与判别式模型

生成式模型(Generative Model)与判别式模型(Discrimitive Model)是分类器常遇到的概念，它们的区别在于：

　　对于输入x，类别标签y：

　　生成式模型估计它们的联合概率分布P(x,y)

　　判别式模型估计条件概率分布P(y|x)

　　生成式模型可以根据贝叶斯公式得到判别式模型，但反过来不行。

<https://www.cnblogs.com/Winnie-T/p/9316622.html>

隐马尔科夫模型

**回忆一下马尔可夫链的两个假设**

|  |  |
| --- | --- |
| (1)齐次马尔可夫假设： | 即t时刻的状态只受t-1时刻状态的影响 |
| (2)观测独立性假设： | 即任意时刻的观测只受该时刻所处状态的影响 |

<https://blog.csdn.net/gzj_1101/article/details/79955340>

条件随机场

条件随机场（Conditional random field，CRF）是给定一组输入随机变量条件下另一组输出随机变量的条件概率分布模型，其特点是假设输出随机变量构成马尔可夫随机场。条件随机场常用于序列标注问题，比如命名实体识别等。本文主要介绍线性链条件随机场。

EM算法 —— 聚类算法

EM算法是一种迭代优化策略，由于它的计算方法中每一次迭代都分两步，其中一个为期望步（E步），另一个为极大步（M步），所以算法被称为EM算法（Expectation-Maximization Algorithm）。EM算法受到缺失思想影响，最初是为了解决数据缺失情况下的参数估计问题，其算法基础和收敛有效性等问题在Dempster、Laird和Rubin三人于1977年所做的文章《Maximum likelihood from incomplete data via the EM algorithm》中给出了详细的阐述。其基本思想是：首先根据己经给出的观测数据，估计出模型参数的值；然后再依据上一步估计出的参数值估计缺失数据的值，再根据估计出的缺失数据加上之前己经观测到的数据重新再对参数值进行估计，然后反复迭代，直至最后收敛，迭代结束。

LDA 线性判别分析

　LDA是一种监督学习的降维技术，也就是说它的数据集的每个样本是有类别输出的。这点和PCA不同。PCA是不考虑样本类别输出的无监督降维技术。LDA的思想可以用一句话概括，就是“投影后类内方差最小，类间方差最大”。什么意思呢？ 我们要将数据在低维度上进行投影，投影后希望每一种类别数据的投影点尽可能的接近，而不同类别的数据的类别中心之间的距离尽可能的大。

Bagging 和 Boosting

Baggging 和Boosting都是模型融合的方法，可以将弱分类器融合之后形成一个强分类器，而且融合之后的效果会比最好的弱分类器更好。

https://www.cnblogs.com/earendil/p/8872001.html

SVM为什么要转化为对偶问题？

1.对偶问题将原始问题中的约束转为了对偶问题中的等式约束

2.方便核函数的引入

3.改变了问题的复杂度。由求特征向量w转化为求比例系数a，在原始问题下，求解的复杂度与样本的维度有关，即w的维度。在对偶问题下，只与样本数量有关。

梯度下降 步长是怎么迭代调整的？

对于full gradient方法，一般都是line search，基本思想就是每次试一个步长，如果用该步长走的话，看函数值会不会比当前点下降一定的程度，如果没有，就按比例减小步长，再试，直到满足条件（根据泰勒展开式我们知道步长足够小时总会满足下降条件）

sgd 一般用自适应的learning rate，比如 ADAGRAD,AdaDelta, ADAM

批量梯度下降法，是梯度下降法最常用的形式，具体做法也就是在更新参数时使用所有的样本来进行更新，这个方法对应于前面3.3.1的线性回归的梯度下降算法，也就是说3.3.1的梯度下降算法就是批量梯度下降法。　　‘

　随机梯度下降法，其实和批量梯度下降法原理类似，区别在与求梯度时没有用所有的m个样本的数据，而是仅仅选取一个样本j来求梯度。对应的更新公式是：

https://www.cnblogs.com/pinard/p/5970503.html’

最小二乘法：

最小二乘法是勒让德( A. M. Legendre)于1805年在其著作《计算慧星轨道的新方法》中提出的。它的主要思想就是求解未知参数，使得理论值与观测值之差（即误差，或者说残差）的平方和达到最小：

所谓最小二乘，其实也可以叫做最小平方和，其目的就是通过最小化误差的平方和，使得拟合对象无限接近目标对象。换句话说，最小二乘法可以用于对函数的拟合。

**局限性**

* 首先，最小二乘法需要计算(XTX)−1逆矩阵，有可能逆矩阵不存在，这样就没有办法直接用最小二乘法。
* 第二，当样本特征n非常的大的时候，计算逆矩阵是一个非常耗时的工作，甚至不可行。建议不超过10000个特征。
* 第三，如果拟合函数不是线性的，这时无法使用最小二乘法，需要通过一些技巧转化为线性才能使用。

怎么判断一个n是否为2次幂（2次幂的二进制特征）

n&(n-1)==0

**1. 一阶优化算法**

这种算法使用各参数的梯度值来最小化或最大化损失函数E(x)。**最常用的一阶优化算法是梯度下降。**

函数梯度：导数dy/dx的多变量表达式，用来表示y相对于x的瞬时变化率。往往为了计算多变量函数的导数时，会用梯度取代导数，并使用偏导数来计算梯度。梯度和导数之间的一个主要区别是函数的梯度形成了一个向量场。

因此，对单变量函数，使用导数来分析；而梯度是基于多变量函数而产生的。更多理论细节在这里不再进行详细解释。

**2. 二阶优化算法 海森矩阵**

二阶优化算法使用了二阶导数(也叫做**Hessian方法**)来最小化或最大化损失函数。由于二阶导数的计算成本很高，所以这种方法并没有广泛使用。

<https://blog.csdn.net/dddxxxx/article/details/77944000>

梯度，海森矩阵，雅可比矩阵

离散数学

合取成为积 ∧ 析取称为和∨

主析取范式：极小项之和 主合取范式：极大项之积

幂集：给定集合A，以集合A的所有子集为元素组成的集合P（A）

集合的划分与覆盖:

划分：每个元素均应出现且仅出现一次在子集里（即A中的元素互不相交）

覆盖：只要求每个元素都出现，没有要求只出现一次

等价关系：自反，对称，可传递 举例：模m同余关系

相容关系：自反，对称 举例：

偏序关系：自反，反对称，可传递 举例：大于等于，小于等于，整除

拟序关系：反自反，可传递（隐含反对称） 举例：小于，大于

全序关系：对一个偏序集合，如果对于每一个x,y 属于A，都有x<=y或y<=x，则称这种偏序关系为全序关系

良序关系：给定集合X , R是X上的二元关系。如果R是全序关系，且X的每一个非空子集都有一个最小元，则称R是良序关系

极小元：集合A中没有比它更小的元素（若存在可能不唯一）

极大元：集合A中没有比它更大的元素（若存在可能不唯一）

最小元：比集合A中任何其他元素都小（若存在就一定唯一）

最大元：比集合A中任何其他元素都大（若存在就一定唯一）

自然数集合N{0,1,2,3…}的基数是阿列夫零

实数集合R 的基数是阿列夫一

代数系统：设S是一个非空集合，且fi是S上的运算，则称S和fi组成的结果，称为代数系统

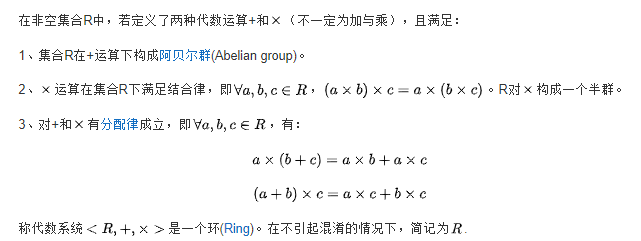
半群：给定代数系统，运算满足结合律，则称为半群

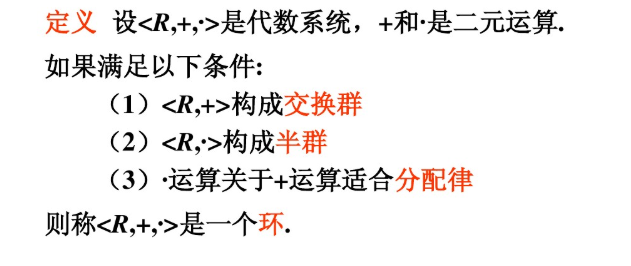
独异点/含幺半群/拟群：半群且含有幺元

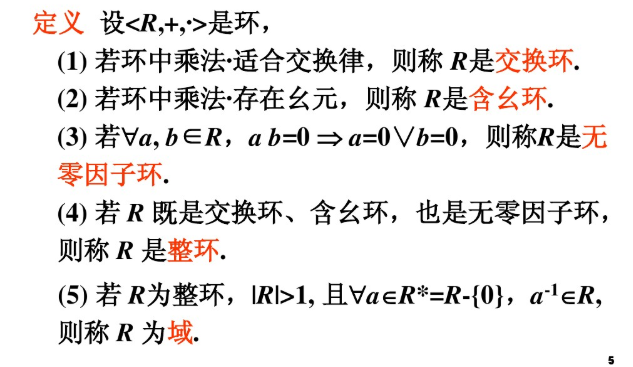
群：代数系统是独异点且每个元素均存在逆元

阿贝尔群（交换群）：满足交换律的群

群的性质：封闭性；结合律；单位元存在；逆元存在







* 乘法为交换半群，则为交换环
* 乘法为独异点，则为含幺环
* [环](https://baike.baidu.com/item/%E7%8E%AF/12795512)R中一个元a≠0，若有0≠b∈R使得ab=0或ba=0，称a是环R的零因子

若R既是交换环，含幺环，无零因子环，则称R为整环

简单路径：边不同

基本路径：结点不同

图G中包含所有边的简单开路径成为图G的欧拉路径

每个结点都是偶结点的连通无向图成为欧拉图（遍历边）

哈密尔顿回路：图G中包含其所有结点的基本闭路径成为图G的哈密尔顿回路

哈密尔顿图——具有哈密尔顿回路的图

数据库

1、第一范式：

当关系bai模式R的所有属性都不能在分解为更基本du的数据单位时，称R是满足第一范式的，简记为1NF。满足第一范式是关系模式规范化的最低要求，否则，将有很多基本操作在这样的关系模式中实现不了。

2、第二范式：

如果关系模式R满足第一范式，并且R得所有非主属性都完全依赖于R的每一个候选关键属性，称R满足第二范式，简记为2NF。

3、第三范式：

设R是一个满足第一范式条件的关系模式，X是R的任意属性集，如果X非传递依赖于R的任意一个候选关键字，称R满足第三范式，简记为3NF。