## 微积分

## 线性代数

矩阵导数

### 特征向量

方阵A, Av=λv,v为特征向量，λ为特征值。则A=QΣQT，Q列向量为特征向量，Σ主对角线为特征值。、

矩阵A乘以x表示，对向量x进行一次转换（旋转或拉伸）（是一种线性转换），而该转换的效果为常数c乘以向量x（即只进行拉伸）。

### SVD分解

一个m×n的实数矩阵A，分解为A=UΣVT，Σ仅在主对角线上有值，我们称它为奇异值。

AAT=UΣVTVΣTUT=UΣΣTUT

ATA=VΣTUTUΣVT=VΣTΣVT

适合一般矩阵。只取头部奇异值，可近似分解矩阵。

### PCA主成分分析

将矩阵的点缩放到超平面上，则样本点到超平面的距离尽量近，样本在超平面的投影尽量远。

组成这个超平面的向量称为主成分。

解法：

对所有点中心化

计算样本的协方差矩阵XXT

特征分解协方差矩阵XXT

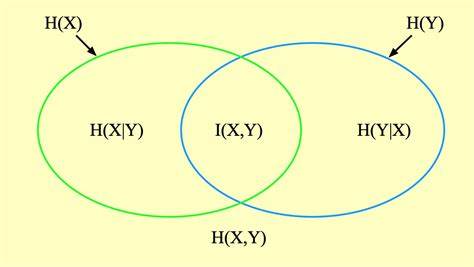
提取前几个成分作为主成分。

## 概率

最大似然 先验概率 后验概率

## 信息熵

条件熵 相对熵 互信息 交叉熵kl散度



<https://www.cnblogs.com/kyrieng/p/8694705.html>

H(X) 就被称为随机变量 x 的熵,它是表示随机变量不确定的度量，是对所有可能发生的事件产生的信息量的期望。

交叉熵用来度量两个离散分布的差异，常用来做分类损失函数。

## 梯度

梯度是各个维度的偏导方向，是当前点增长增快的方向，常常使用梯度下降求解凸函数最小值。

## 凸优化

通常建模设计任务为求目标函数最大值或者最小值，为保证局部极值是全局极值，设计任务函数为凸函数。

有约束条件时候，拉格朗日乘子法转换为无约束函数。

<https://blog.csdn.net/lijil168/article/details/69395023>