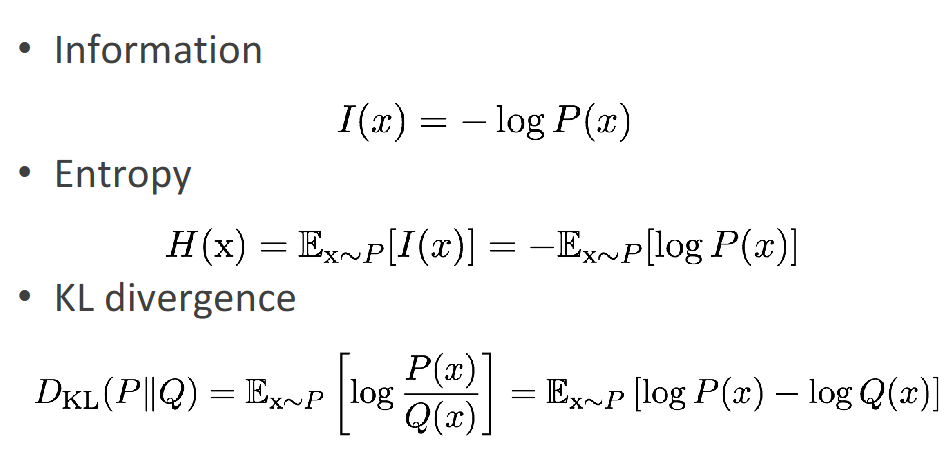
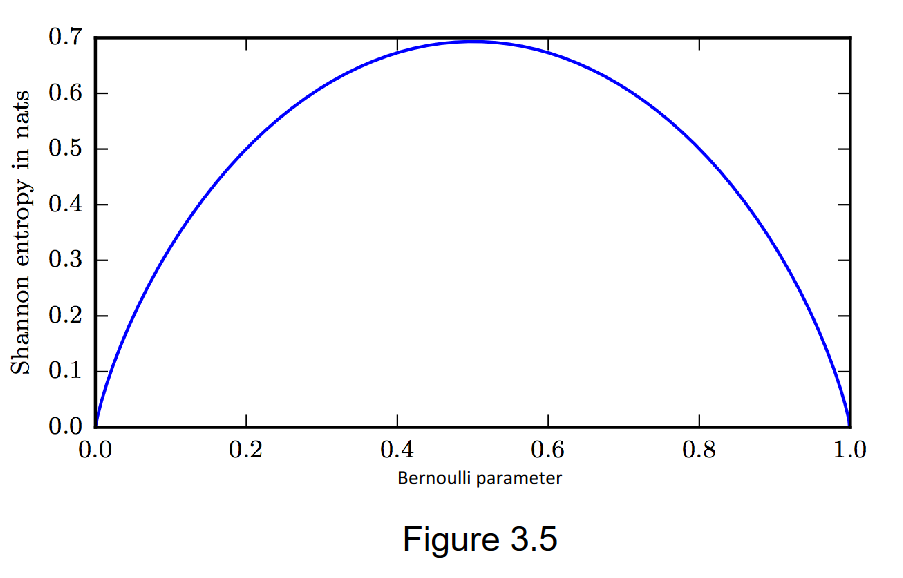
Mathematical background(数学背景)（下）

PS：之后改用md编辑了，幕布还是太难排版了

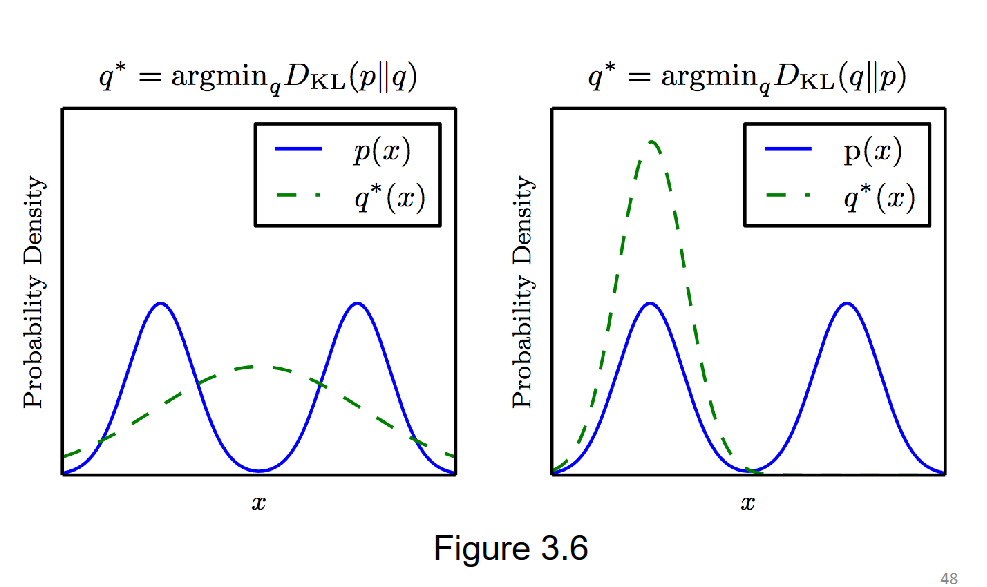
Information theory



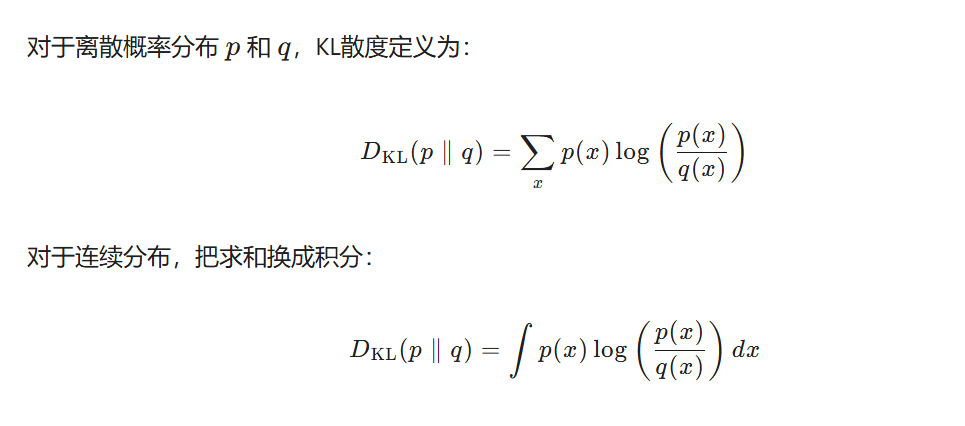
伯努利变量的熵

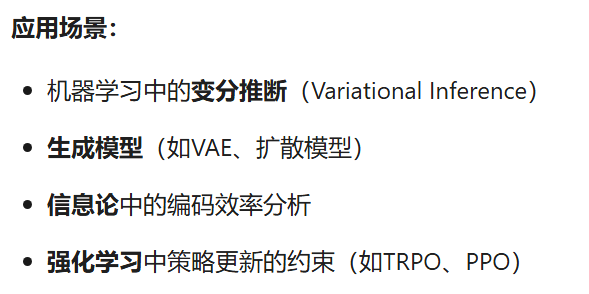


KL散度是不对称的



KL散度衡量的是：当你用一个分布 q 去近似另一个分布 p 时，损失了多少信息

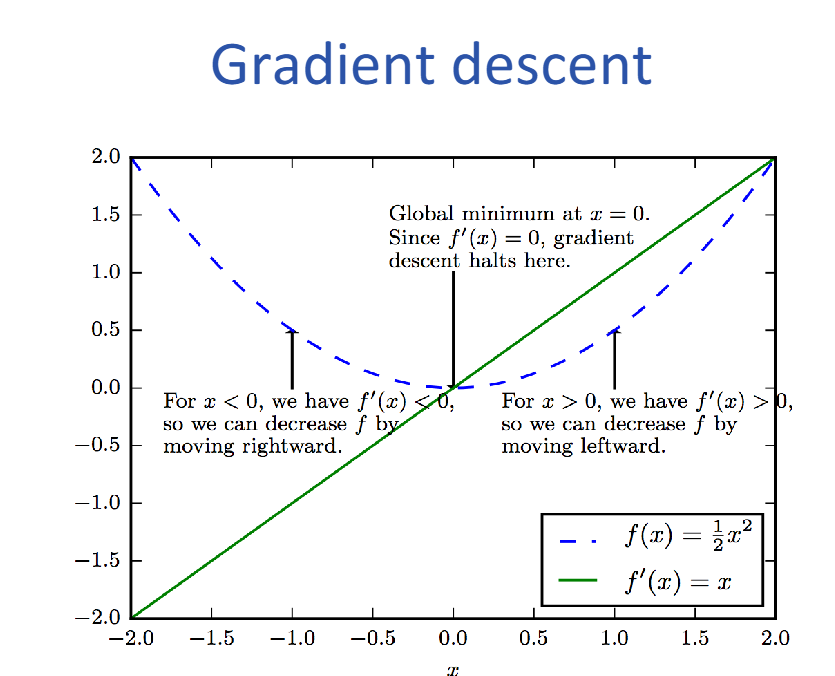




Optimisation andnumerical computation（优化和数值计算）

First order derivatives（一阶导数）

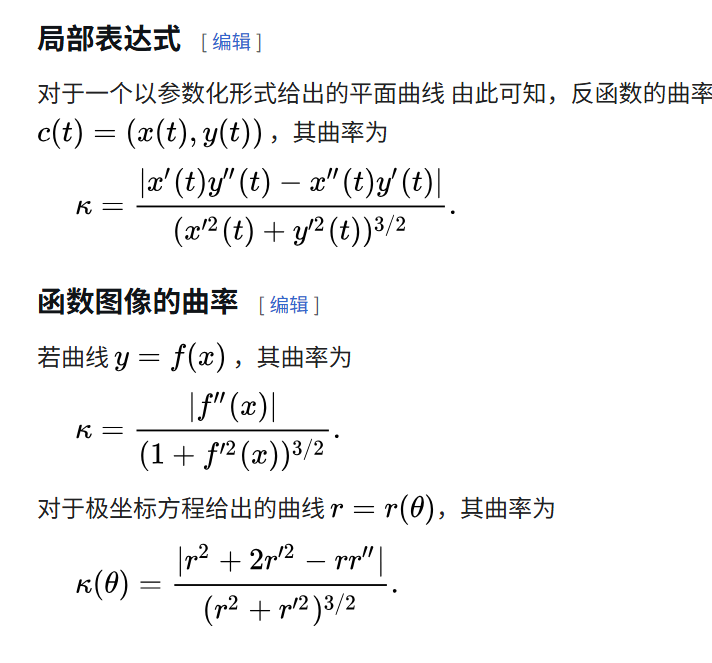
Gradient descent（梯度损失）（梯度是各个方向上的偏导数之和）



Critical points（极值点）

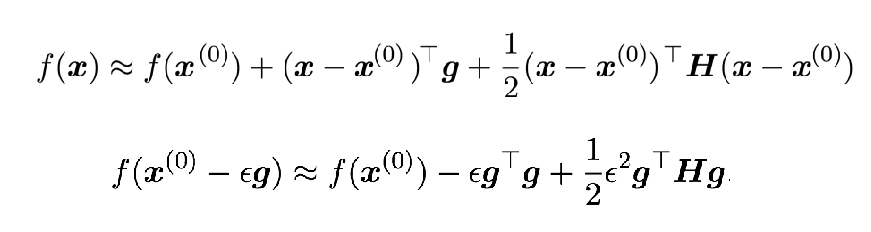
Curvature（曲率）

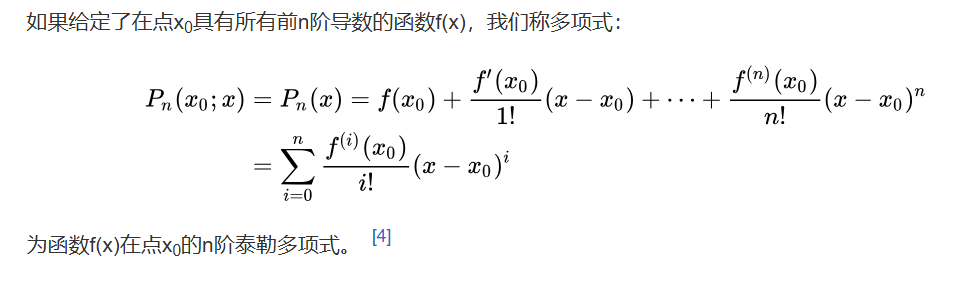
计算公式：（也是差不多不记得了，复习一下）



Second order derivatives

基于Taylor的最优步长系列（泰勒展开式，牢记）





后面是一些代码中使用的函数解释，具体内容见实验课。