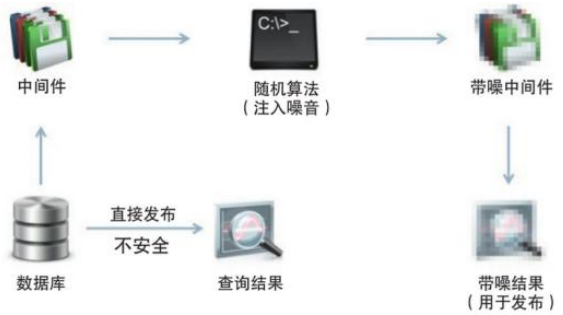
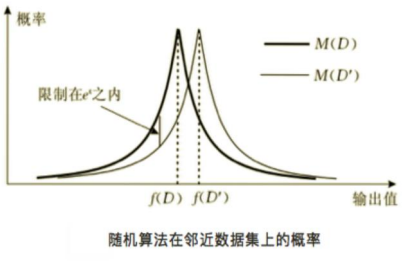
差分隐私
Differential Privacy

差分隐私是 Dwork 在 2006 年提出的一种隐私计算技术，目的是提供一种技术保证在数据库查询和统计时，能够最大化数据查询的准确性，同时最大限度来减少识别单个记录的机会，也就是让查询结果对于数据集中单个记录的变化不敏感，从而攻击者就无法通过加入或减少一个记录，观察查询结果的变化来推测个体的具体信息。

差分隐私中一个关键概念是相邻数据集，假设给定两个数据集 D 和 D'，如果它们有且仅有一条数据不一样，那么这两个数据集可称为相邻数据集。那么如果对于一个随机算法 A 如果其分别作用于两个相邻数据集得到的两个输出分布式难以区分的，那么这个算法就被认为达到差分隐私的效果。





随机算法在邻近数据集上的概率

差分隐私主要通过在输入输出中增加适量的随机噪音以确保修改数据集中一条个体记录不会对统计结果造成显著影响，差分隐私的主要噪声机制包括：拉普拉斯噪音(Laplace Noise)，高斯噪音(Gaussian Noise)，指数机制等。基于拉普拉斯机制和高斯机制的差分隐私是通过向查询结果中加入服从拉普拉斯分布和高斯分布的噪声来实现差分隐私保护，适用于数字型结果的差分隐私。基于指数机制的差分隐私，则采用指数机制来引入一个打分函数来实现差分隐私保护，也就是当接收到一个查询后，不是确定性的输出一个结果，而是以一定概率值返回结果，这个概率值是由打分函数确定，指数机制适用于离散型结果的差分隐私。

差分隐私根据数据收集分析中的保护对象的差异可以划分为中心化差分隐私和本地化差分隐私两种类型。中心化差分隐私通常具有可信第三方，第三方利用收集到的数据进行统计分析，因此差分隐私主要是保护数据收集分析后的结果发放过程，因此差分隐私保护机制是运行在可信第三方上。本地化差分隐私则认为第三方不可信，因此本地差分隐私主要保护数据上传到第三方的过程，因此差分隐私机制是运行在各个数据提供方本地。

差分隐私常用于多方参与的统计分析中，也用于联邦机器学习算法，在中间参数交互过程通过采用差分隐私，避免恶意参与方通过中间结果反推出训练集中的个体特征信息，增强联邦算法的安全性。