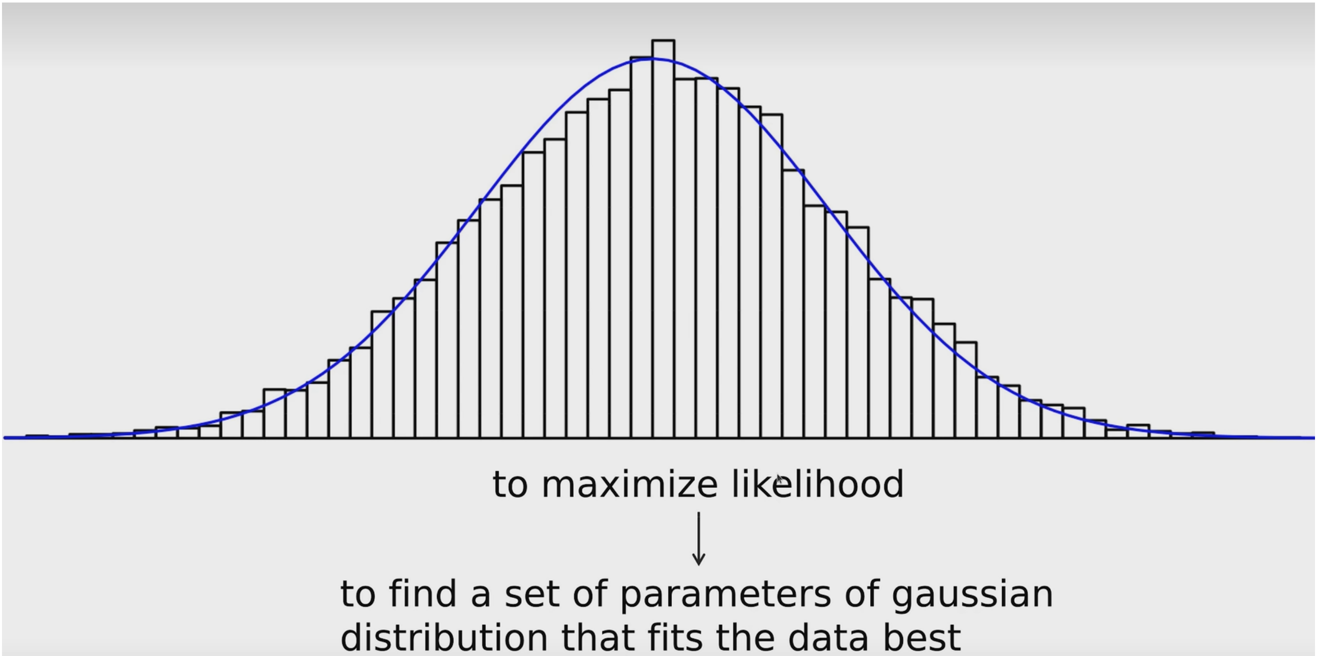
# Maximum Likelihood Estimation

author: doomx
time: 2023.9.22

假设我们有一批数据，这批数据是符合Gaussian Distribution的，现在我们要估计一个gaussian Distribution去拟合这个数据的一些规律，maximum likelihood estimation就是确定这个分布的一些参数，使得这组数据出现的概率最大，如图1所示。

example(fig1)




Bayes' Theorem(fig2)

$$p(\overset{\text{parameters}}{\theta}|\overset{\text{data}}{X}) = \frac{p(X|\theta)p(\theta)}{p(X)} \propto p(X|\theta)\overset{\text{uniformly distributed}}{p(\theta)} \propto p(X|\theta)$$

$$X = \{x_0, x_1, \cdots, x_n\}$$

bayes theorem可以表现为 后验概率 = （似然度 * 先验概率）/ 数据的概率，由于数据已经给定了，可以理解这个p(X)是一个常数，说明**后验概率正比于似然度乘以先验概率**。假设*我们对于这个*

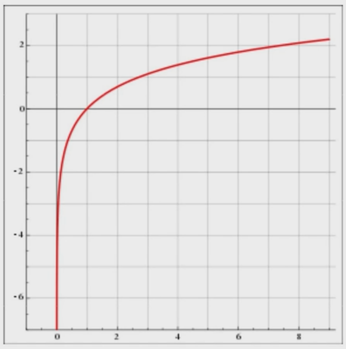
*参数的分布是没有一个先验知识的，可以假设为是一个均匀分布，参数的概率在全值域都是一个常数*，所以说最终可以认为**后验概率正比于似然度。**

如果是**独立同分布**的这个数据，如图2所示，那么p(X | θ)可以表示为一系列$x_0, x_1, \dots, x_n$相乘，即$p(x_0 \mid \theta)p(x_1 \mid \theta)\dots p(x_n \mid \theta)$。

Log Likelihood(fig3)

$$\hat{\theta} = \text{argmax } p(X|\theta) = \text{argmax } \prod_i^n p(x_i|\theta) = \text{argmax } log \prod_i^n p(x_i|\theta)$$

$$= \text{argmax } \sum_i^n log(p(x_i|\theta))$$




taking log turns multiplication to sum

因为对数是单调增长的，所以对连乘块取对数不会影响求最大值，因为对于连乘的操作要对它去做一些操作和处理比较麻烦，所以我们用取对数的形式可以很方便的把它变为一个连加求和的式子。这种方法是很常见的，值得学习。

Maximum Likelihood Estimation(fig4)

## Maximum Likelihood Estimation (Univariate Gaussian)

$$\theta := \{\mu, \sigma\} \qquad p(x_i|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{(-\frac{(x_i-\mu)^2}{\sigma^2})}$$

$$\frac{\partial \sum_i^n log(p(x_i|\mu,\sigma))}{\partial \mu} = 0 \qquad \mu = \frac{1}{n}\sum_i^n x_i$$

$$\frac{\partial \sum_i^n log(p(x_i|\mu,\sigma))}{\partial \sigma} = 0 \qquad \sigma^2 = \frac{1}{n}\sum_i^n (x_i-\mu)^2$$

给定一组参数之后，求这个对数似然度的极值，方法就是微积分求导求极值，用这个似然度函数对参数求偏导，令导数等于0。这个单变量的高斯分布有两个参数，一个是 μ ，一个是 σ ，对于这个高斯分布，很容易求出最大似然度的极值是 μ 取所有样本的均值， σ 取所有样本的方差适合可以得到。

1. Likelihood is the probability of data given distribution parameters. （似然度是在给定一系列参数的条件下数据出现的概率。）
2. MLE is trying to estimate a set of distribution parameters that maximizes the probability of the data set given the estimated parameters. （最大似然估计试图估计一组分布参数，该分布参数在给定估计参数的情况下使数据出现的概率最大。）
3. MLE is a special case of MAP with the assumption of uniform prior, but it's more feneral. （最大似然估计是最大后验估计的一个特例，假设*先验是均匀的*，这让它更普遍。）
4. Taking the log likelihood helps to convert multiplication to sum.（对似然度求对数可以帮助我们把连乘的操作转换为求和的操作，**方便后面求导。**）
5. Solve the equation of log likelihood partial derivative equal to zero gives the estimate of the respective parameter.（求解对数似然偏导数等于零的方程，得出相应参数的估计值。）