

极大似然估计方法是求估计的另一种方法,1821年首先由德国数学家C. F. Gauss提出,但是这个方法通常被归功于英国的统计学家R. A. Fisher, 他在1922年的论文On the mathematical foundations of theoretical statistics, reprinted in Contributions to Mathematical Statistics (by R. A. Fisher), 1950, J. Wiley & Sons, New York 中再次提出了这个思想,并且首先探讨了这种方法的一些性质.极大似然估计这一名称也是费歇给的。这是一种目前仍然得到广泛应用的方法。

求极大似然函数估计值的一般步骤:

- (1) 写出似然函数;
- (2) 对似然函数取对数,并整理;
- (3) 求导数;
- (4) 解似然方程。

极大似然估计,只是一种概率论在统计学的应用,它是参数估计的方法之一。说的是已知某个随机样本满足某种概率分布,但是其中具体的参数不清楚,参数估计就是通过若干次试验,观察其结果,利用结果推出参数的大概值。极大似然估计是建立在这样的思想上:已知某个参数能使这个样本出现的概率最大,我们当然不会再去选择其他小概率的样本,所以干脆就把这个参数作为估计的真实值。

它是建立在极大似然原理的基础上的一个统计方法,极大似然原理的直观想法是:一个随机试验如有若干个可能的结果A, B, C, ...。若在一次试验中,结果A出现,则一般认为试验条件对A出现有利,也即A出现的概率很大。

牛顿法求解方程

并不是所有的方程都有求根公式,或者求根公式很复杂,导致求解困难。利用牛顿法,可以迭代求解。

原理是利用泰勒公式,在 x_0 处泰勒展开到一阶,即: $f(x) = f(x_0) + (x - x_0)f'(x_0)$

求解方程: $f(x)=0$, 即: $f(x_0) + (x-x_0)*f'(x_0)=0$, 求解: $x = x_1 = x_0 - f(x_0)/f'(x_0)$,

利用泰勒公式的一阶展开, $f(x) = f(x_0) + (x - x_0)f'(x_0)$ 处只是近似相等,这里求得的 x_1 并不能让 $f(x) = 0$,只能说 $f(x_1)$ 的值比 $f(x_0)$ 更接近 $f(x) = 0$,于是乎,迭代求解的想法就很自然了,可以进而推出 $x_{n+1} = x_n - f(x_n)/f'(x_n)$,通过迭代,这个式子必然在 $f(x^*) = 0$ 的时候收敛。整个过程如下图:

牛顿法用于最优化

在最优化的问题中,线性最优化至少可以使用单纯行法求解,但对于非线性优化问题,牛顿法提供了一种求解的办法。假设任务是优化一个目标函数 f ,求函数 f 的极大极小问题,可以转化为求解函数 f 的导数 $f'=0$ 的问题,这样求可以把优化问题看成方程求解问题($f'=0$)。剩下的问题就和第一部分提到的牛顿法求解很相似了。在极小值估计值附近,把 $f(x)$ 泰勒展开到2阶形式:

$$f(x + \Delta x) = f(x) + f'(x)\Delta x + \frac{1}{2}f''(x)\Delta x^2$$

当且仅当 Δx 无限趋近于0,上面的公式成立。令: $f'(x+\Delta x)=0$, 得到:

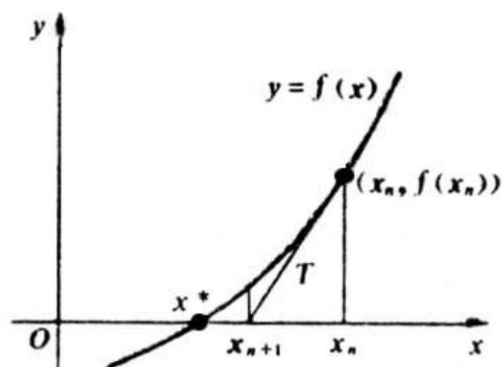
$$f'(x) + f''(x)\Delta x = 0.$$

求解:

$$\Delta x = -\frac{f'(x_n)}{f''(x_n)}$$

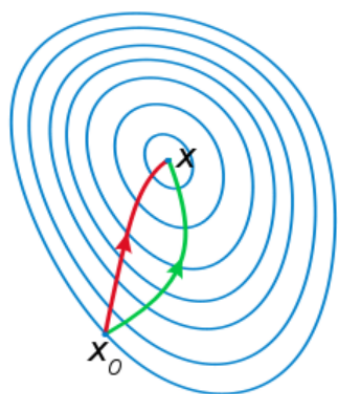
得出迭代公式:

$$x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)}, n = 0, 1, \dots$$



牛顿法求实根图示

从几何上说, 牛顿法就是用一个二次曲面去拟合你当前所处位置的局部曲面, 而梯度下降法是用一个平面去拟合当前的局部曲面, 通常情况下, 二次曲面的拟合会比平面更好, 所以牛顿法选择的下降路径会更符合真实的最优下降路径。如下图是一个最小化一个目标方程的例子, 红色曲线是利用牛顿法迭代求解, 绿色曲线是利用梯度下降法求解。



优化提要

- 各任务下 (回归、分类) 各个模型优化的目标
 - 最小二乘法: 最小化均方误差
 - 对数几率回归: 最大化样本分布似然
 - 线性判别分析: 投影空间内最小 (大) 化类内 (间) 散度
- 参数的优化方法
 - 最小二乘法: 线性代数
 - 对数几率回归: 凸优化梯度下降、牛顿法
 - 线性判别分析: 矩阵论、广义瑞利商

Ch2 总结

- 线性回归
 - 最小二乘法 (最小化均方误差)
- 二分类任务
 - 对数几率回归
 - 单位阶跃函数、对数几率函数、极大似然法
 - 线性判别分析
 - 最大化广义瑞利商
- 多分类学习

- 一对一
- 一对其余
- 多对多

□ 纠错输出码

□ 类别不平衡问题

- 基本策略：再缩放

正态分布的**概率密度函数**均值为 μ 方差为 σ^2 (或**标准差** σ) 是**高斯函数**的一个实例:

$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

([请看指数函数](#)以及 π .)

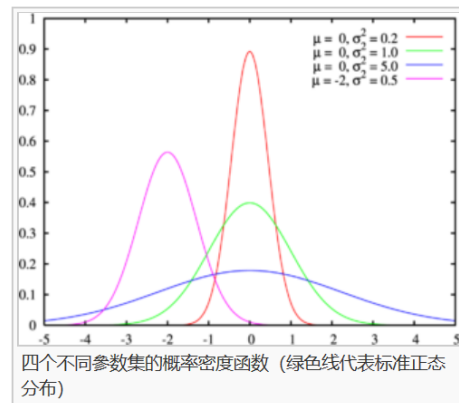
假设一个**随机变量**服从这个分布, 我们写作 $X \sim N(\mu, \sigma^2)$. 假设 $\mu = 0$ 而且 $\sigma = 1$, 这个分布被称为**标准正态分布**, 这个分布可以简化为

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

右边是给出了不同参数的正态分布的函数图。

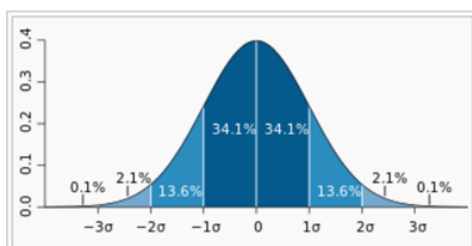
正态分布中一些值得注意的量:

- 密度函数关于平均值对称
- 平均值是它的**众数** (statistical mode) 以及**中位数** (median)
- 函数曲线下68.268949%的面积在平均值左右的一个**标准差**范围内
- 95.449974%的面积在平均值左右两个标准差 2σ 的范围内
- 99.730020%的面积在平均值左右三个标准差 3σ 的范围内
- 99.993666%的面积在平均值左右四个标准差 4σ 的范围内
- **反曲点** (inflection point) 在离平均值的距离为标准差之处



激活 Windows
转到“设置”以激活 Windows

在实际应用上, 常考虑一组数据具有近似于**正态分布**的概率分布。若其如果正确, 则约68%数值分布在距离平均值有1个标准差之内的范围, 约95%数值分布在距离平均值有2个标准差之内的范围, 以及约99.7%数值分布在距离平均值有3个标准差之内的范围。称为"68-95-99.7法则"或"经验法则".



深蓝色区域是距平均值小于一个标准差之内的数值范围。在**正态分布**中, 此范围所占比率为所有数值之68%。依据正态分布, 两个标准差之内 (蓝, 棕) 的比率合起来为95%。依据正态分布, 三个标准差之内 (深蓝, 橙, 黄) 的比率合起来为99%。

这种判别处理原理及方法仅局限于对正态或近似正态分布的样本数据处理, 它是以测量次数充分大为前提的, 当测量次数较少的情形用准则剔除粗大误差是不够可靠的。因此, 在测量次数较少的情况下, 最好不要选用准则, 而用其他准则。

在**正态分布**中 σ 代表**标准差**, μ 代表**均值**。 $x=\mu$ 即为图像的对称轴

3 σ 原则为

数值分布在 $(\mu-\sigma, \mu+\sigma)$ 中的概率为0.6827

数值分布在 $(\mu-2\sigma, \mu+2\sigma)$ 中的概率为0.9545

数值分布在 $(\mu-3\sigma, \mu+3\sigma)$ 中的概率为0.9973

可以认为, Y 的取值几乎全部集中在 $(\mu-3\sigma, \mu+3\sigma)$ 区间内, 超出这个范围的可能性仅占不到0.3%.

《饮料装填量不足与超量的概率》

某饮料公司装瓶流程严谨，每罐饮料装填量符合平均600毫升，标准差3毫升的常态分配法则。随机选取一罐，容量超过605毫升的概率？容量小于590毫升的概率

容量超过605毫升的概率 = $p(X > 605) = p\left(\frac{(X-\mu)}{\sigma} > \frac{(605 - 600)}{3}\right) = p(Z > 5/3) = p(Z > 1.67) = 0.9525$

容量小于590毫升的概率 = $p(X < 590) = p\left(\frac{(X-\mu)}{\sigma} < \frac{(590 - 600)}{3}\right) = p(Z < -10/3) = p(Z < -3.33) = 0.0004$

《6-标准差(6-sigma或6-σ)的品质管制标准》

6-标准差(6-sigma或6-σ)，是制造业流行的品质管制标准。在这个标准之下，一个标准常态分配的变量值出如今正负三个标准差之外，仅仅有 $2 \times 0.0013 = 0.0026$ ($p(Z < -3) = 0.0013$ 以及 $p(Z > 3) = 0.0013$)。也就是说，这样的品质管制标准的产品不良率仅仅有万分之二十六。如果例3-16的饮料公司装瓶流程采用这个标准，而每罐饮料装填量符合平均600毫升，标准差3毫升的常态分配法则。预期装填容量的范围应该多少？6-标准差的范围 = $p(-3 < Z < 3) = p\left(-3 < \frac{(X-\mu)}{\sigma} < 3\right) = p\left(-3 < \frac{(X-600)}{3} < 3\right) = p(-9 < X - 600 < 9) = p(591 < X < 609)$ 因此，预期装填容量应该介于591至609毫升之间。