**AIC**

很多参数估计问题均采用似然函数作为目标函数，当训练数据足够多时，可以不断提高模型精度，但是以提高模型复杂度为代价的，同时带来一个机器学习中非常普遍的问题——过拟合。所以，模型选择问题在模型复杂度与模型对数据集描述能力（即似然函数）之间寻求最佳平衡。

人们提出许多信息准则，通过加入模型复杂度的惩罚项来避免过拟合问题，此处我们介绍一下常用的两个模型选择方法——赤池信息准则（Akaike Information Criterion，AIC）和贝叶斯信息准则（Bayesian Information Criterion，BIC）。

AIC是衡量统计模型拟合优良性的一种标准，由日本统计学家赤池弘次在1974年提出，它建立在熵的概念上，提供了权衡估计模型复杂度和拟合数据优良性的标准。

通常情况下，AIC定义为：

http://img.blog.csdn.net/20141105155641871?watermark/2/text/aHR0cDovL2Jsb2cuY3Nkbi5uZXQvanRlbmc=/font/5a6L5L2T/fontsize/400/fill/I0JBQkFCMA==/dissolve/70/gravity/Center

其中k是模型参数个数，L是似然函数。从一组可供选择的模型中选择最佳模型时，通常选择AIC最小的模型。

当两个模型之间存在较大差异时，差异主要体现在似然函数项，当似然函数差异不显著时，上式第一项，即模型复杂度则起作用，从而参数个数少的模型是较好的选择。

一般而言，当模型复杂度提高（k增大）时，似然函数L也会增大(即RSS變小)，从而使AIC变小，但是k过大时，似然函数增速减缓，导致AIC增大，模型过于复杂容易造成过拟合现象。目标是选取AIC最小的模型，AIC不仅要提高模型拟合度（极大似然），而且引入了惩罚项，使模型参数尽可能少，有助于降低过拟合的可能性。

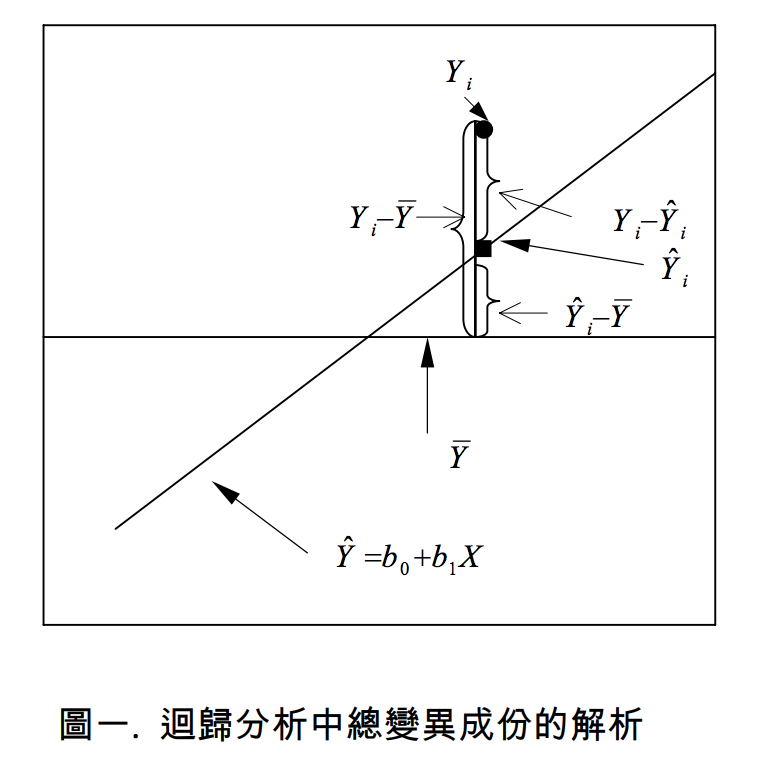
参考资料：<http://blog.csdn.net/lynnucas/article/details/47947943>

**AdjustedR2**

當因变数Y 与自变数 X 之间的关系可以用一个回归方程式来解释时，X

的解释能力有必要进一步地了解。该解释能力的程度大小，即回归分析的配合度(goodness of fit)，是以决定系数(R2)来描述。如图一所示，总变异()是由两个成分组成：配合值离平均值之变异()，以及观测值离配合值之变异()。前者()是由所建立回归式之X 所引起，而后者() 则为逢机机差所引起。由建立回归式之 X 所引起之平方和占总变异平方和的比例，称为决定系数

即 ，式中 为总和之代表符号；有些教科书则将 R2 翻译为判定系数。R2 的所在范围在 0 与 1 之间， 其结果的大小表示 Y 的变异中X 所能解释的程度。当 R2 值越接近 1 时，表示估计式中大部份 Y 之变异是由X 影响而来，也代表利用 X 来解释Y 的能力越强，因此所建立的回归模式为合适可接受。



当自变数不只一项时，例如同时探讨氮、磷、钾多种肥料对作物产量的影响关系，或气温、日照、雨量等各种气象因素对产量的综合影响，一组自变数 X1,X2,X3……与因变数 Y 之间的直线关系，可以利用复回归方程式(multiple regression function)来表示：Yˆ =b0+b1X1+b2X2+b3X3……。但并非所有生物现象都是呈直线关系，当肥料用量增加时，作物产量可能会以一缓慢的速度增加，以致于该曲线会逐渐平稳而接近水平；当过度施肥时，甚至对作物造成毒害而使曲线下降。因此有时候，非直线回归方程式的探讨，是有其必要性的。对同一套试验资料而言，到底应该适合于何种特定形式的回归方程式，常常也是我们探讨的重点。而此时，R2 是用来作为判断回归方程式是否有效的一个重要指标。也就是说，回归系数是看个别自变数与因变数间的净关系，而决定系数则是看全部自变数与因变数间的综合关系。

修正的公式是

其中n是样本数量，p是模型中变量的个数。

我们知道在其他变量不变的情况下，引入新的变量，总能提高模型的R2。修正R2就是相当于给变量的个数加惩罚项。

换句话说，如果两个模型，样本数一样，R2一样，那么从修正R2的角度看，使用变量个数少的那个模型更优。

参考资料：<http://ilc.hk.edu.tw/c/document_library/get_file?p_l_id=260741&folderId=261080&name=DLFE-3350.pdf>

<http://sofasofa.io/forum_main_post.php?postid=1000702>