

# 过拟合

维基百科，自由的百科全书

在统计学中，**过拟合**（英语：**overfitting**，或称过度拟合）现象是指过于紧密或精确地匹配特定数据集的结果，以致于无法拟合其他数据或预测未来的观察结果。<sup>[1]</sup>

对比于可获取的数据总量来说，一个荒谬的模型只要足够复杂，是可以完美地适应数据。过拟合一般可以视为违反奥坎剃刀原则。

当可选择的参数的自由度超过数据所包含信息内容时，这会导致最后（拟合后）模型使用任意的参数，这会减少或破坏模型一般化的能力更甚于适应数据。过拟合的可能性不只取决于参数个数和数据，也跟模型架构与数据的一致性有关。此外对比于数据中预期的噪声或错误数量，跟模型错误的数量也有关。

过拟合现象的观念对机器学习也是很重要的。通常一个学习算法是借由训练示例来训练的。亦即预期结果的示例是可知的。而学习者则被认为须达到可以预测出其它示例的正确的结果，因此，应适用于一般化的情况而非只是训练时所使用的现有数据（根据它的归纳偏向）。然而，学习者却会去适应训练数据中太特化但又随机的特征，特别是在当学习过程太久或示例太少时。在过拟合的过程中，当预测训练示例结果的表现增加时，应用在未知数据的表现则变更差。

在统计和机器学习中，为了避免过拟合现象，须要使用额外的技巧（如交叉验证、提早停止、贝斯信息量准则、赤池信息量准则或模型比较），以指出何时会有更多训练而没有导致更好的一般化。人工神经网络的过拟合过程亦被认知为过度训练（英语：overtraining）。在treatmeant learning中，使用最小最佳支持值（英语：minimum best support value）来避免过拟合。

相对于过拟合是指，使用过多参数，以致太适应数据而非一般情况，另一种常见的现象是使用太少参数，以致于不适应数据，这则称为乏适（英语：underfitting，或称：拟合不足）现象。

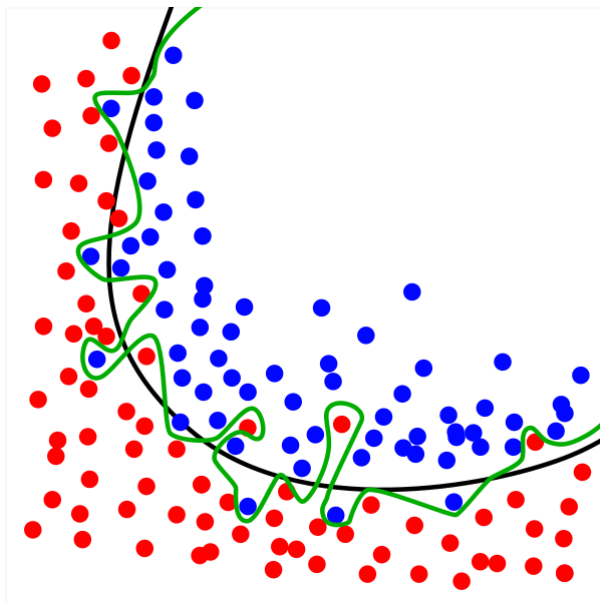
## 参考文献

1. OxfordDictionaries.com上overfitting (<https://en.oxforddictionaries.com/definition/overfitting>)在统计学中的定义。
- Tetko, I.V.; Livingstone, D.J.; Luik, A.I. Neural network studies. 1. Comparison of Overfitting and Overtraining, J. Chem. Inf. Comput. Sci., 1995, 35, 826–833 (<http://www.vcclab.org/articles/tetko.html#overtraining>)

## 外部链接

- <http://www.cs.sunysb.edu/~skiena/jaialai/excerpts/node16.html>
- 过度训练 (<http://www.vcclab.org/articles/tetko.html#overtraining>)（英文）

## 参见



绿线代表过拟合模型，黑线代表正则化模型。虽然绿线完美的符合训练数据，但太过依赖，并且与黑线相比，对于新的测试数据上具有更高的错误率。

取自“<https://zh.wikipedia.org/w/index.php?title=過適&oldid=56454626>”

---

本页面最后修订于2019年10月13日（星期日）01:26。

本站的全部文字在知识共享 署名-相同方式共享 3.0协议之条款下提供，附加条款亦可能应用。（请参阅使用条款）  
Wikipedia®和维基百科标志是维基媒体基金会的注册商标；维基™是维基媒体基金会的商标。  
维基媒体基金会是按美国国内税收法501(c)(3)登记的非营利慈善机构。