

中文译者:

白雪(某大型国企 10 年+IT 从业者, 初稿译者)

龙星镖局(互联网从业者, 终稿译者)

译者声明:

本文翻译自 KDnuggets, 略有删减, 原始链接如下。

第二部分:

<http://www.kdnuggets.com/2016/02/21-data-science-interview-questions-answers-part2.html/>

附加日: 什么是过拟合? 怎么防止过拟合? 【[Gregory Piatetsky](#) 解答】

[Google correlate](#)   [Spurious correlations](#)

伪发现率(False Discovery Rate)

可重用保持法(reusable holdout method)

延伸阅读

- [The Cardinal Sin of Data Mining and Data Science: Overfitting](#)
- [Big Idea To Avoid Overfitting: Reusable Holdout to Preserve Validity in Adaptive Data Analysis](#)
- [Overcoming Overfitting with the reusable holdout: Preserving validity in adaptive data analysis](#)
- [11 Clever Methods of Overfitting and how to avoid them](#)
- [Tag: Overfitting](#)

第十三天: “长”(高)、“宽”数据(tall/wide data)的差别是什么? 【[Gregory Piatetsky](#) 解答】

[Statistical Learning with Sparsity: The Lasso and Generalizations](#), 由 Hastie, Tibshirani, 和 Wainwright 编写

第十五天: 解释 Edward Tufte's 的“图表垃圾”(chart junk)的概念 【[Gregory Piatetsky](#) 解答】

[Edward Tufte](#) 1983 年的书 [《定量信息的视觉展示》](#)

第十八天: 什么是推荐引擎(recommendation engine)? 它是怎么工作的? 【[Gregory Piatetsky](#) 解答】

Dataconomy 对推荐引擎有一个很好的介绍 [Introduction to Recommendation Engines](#). Toptal 有一个关于协同过滤的概览 [building a Collaborative Filtering Recommendation Engine](#). 更新的推荐技术进展可参见 [ACM RecSys conference](#).

第二十天: 你使用什么工具来做可视化(visualization)? 你觉得 Tableau 怎么样? R 呢? SAS 呢(对图形来说)。怎样在一张图里(或者视频里)有效地展示 5 维数据? 【[Gregory Piatetsky](#) 解答】

下面是 KDnuggets 上一些有用的资源链接:

- [Visualization and Data Mining Software](#)
- [Overview of Python Visualization Tools](#)

- [21 Essential Data Visualization Tools](#)
- [Top 30 Social Network Analysis and Visualization Tools](#)
- [Tag: Data Visualization](#)

还可以参考

- [Quora: What's the best way to visualize high-dimensional data?](#)
- pioneering work of Georges Grinstein and his colleagues on [High-Dimensional Visualizations](#) .