第一章“数据科学-鸟瞰”

解释了数据科学或机器学习是一种让机器能够在不被告知或编程的情况下从数据集中学习的过程。例如，要编写一个以手写数字作为输入图像，并根据在图像中写入的数字从0-9输出值的程序将非常困难。这也同样适用于将收到的电子邮件分类为垃圾邮件或非垃圾邮件的任务。为了解决这些问题，数据科学家利用数据科学或机器学习领域的学习方法和工具，通过给计算机一些解释性特征来教计算机如何来自动识别数字，从而达到能够区分不同数字之间的特性。对于垃圾邮件/非垃圾邮件问题也是如此，我们可以通过特定的学习算法来教计算机如何区分垃圾邮件和无垃圾邮件，而不是使用正则表达式和编写上百条规则来对收到的电子邮件进行分类。

1. 数据科学入门-泰坦尼克号

线性模型是数据科学领域的基本学习算法。在你学习数据科学的过程中，了解线性模型是如何工作的是至关重要的。因为它是大多数复杂的学习算法的基本组成部分，包括神经网络在内。