

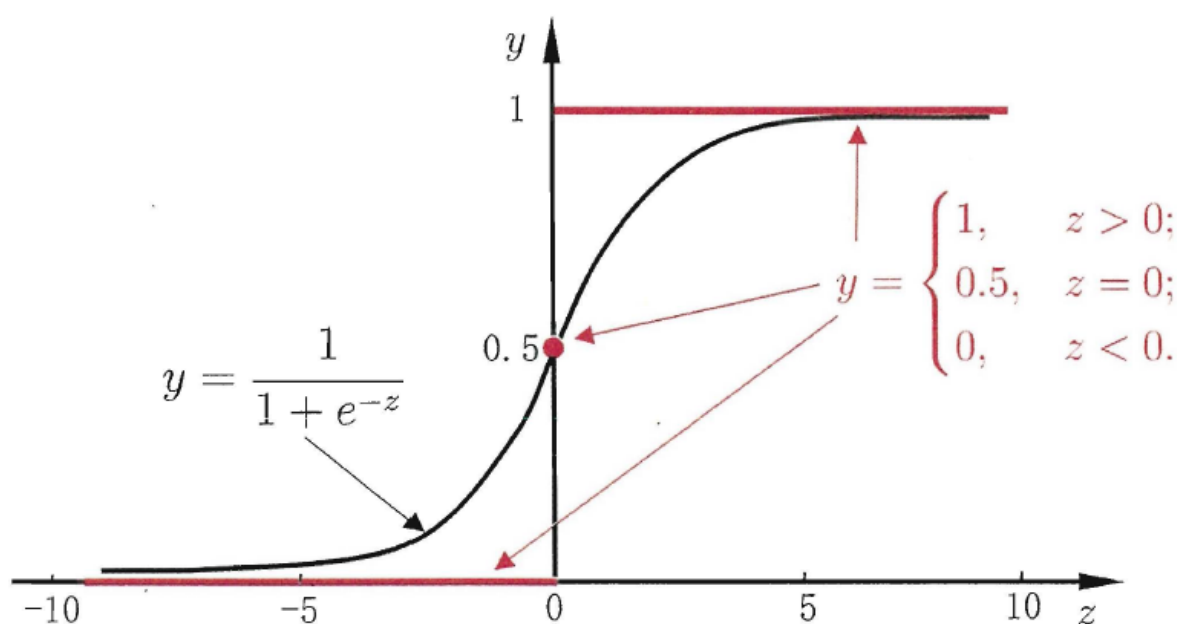
一、概述

上一章主要学习了回归问题，这一章学习分类问题。我们将在代码中实现判别式分类器和生成式分类器，分别为逻辑回归分类器和朴素贝叶斯分类器。并对分类器的可靠性进行交叉验证。

二、内容

朴素贝叶斯分类器：根据已知数据计算变量属于 y_1, y_2, \dots 中属于哪一类的概率。选择其中概率最大的 y 作为结果。朴素贝叶斯是贝叶斯分类器的一个分支。有兴趣可以了解下里面的算法（这也是大二概率论中所讲到的概念）。在代码2.4中举例了估算收入阶层使用了朴素贝叶斯分类器。

逻辑回归分类器：逻辑回归名字上是做了一个回归任务，但实际上是一个分类器，将输入值映射到 $[0,1]$ 之间的小数（目前将这个小数解读为概率值），映射结束后将大于阈值的设为1，小于阈值的设为0。



交叉验证：评估模型的表现，在整个集合中抽取

三、心得

我目前认为数据的预处理还是重中之重，在接下来的2.3、2.4的代码中，虽然只是将数据单纯的导入分类，但是里面还是有很多小细节，需要自己读代码多写写。

参考博客:

朴素贝叶斯: https://blog.csdn.net/huanfeng_AI/article/details/107675987