

Note about the video about Christopher Bishop's Model-Based machine learning

Zhenxuan Zhang

September 24, 2019

1 Problem 1

在视频里，Christopher 提到每一种分类算法在对之前从未观测过的数据进行分析时都有着无明显差异的错判率。机器学习并不是去找去一个单一的适用于所有类型数据的算法，而是去找出针对每一种特定类型数据分析效果最好的算法。

我很认同他的观点，现实生活中信息无处不在，类型数也数不尽，一种算法也许只适用于数值型数据，而对分类型数据分析效果不好。数值型数据内部还能再根据数据特点分成无数个分支，比如高维的数据和低维的，数据量极大的和数据量较小的等等。没有普适的算法就好比教书时要因材施教，发展农业时要因地制宜。

而且每个算法都有使用的假设条件，而这些条件限制了它的应用范围，我们必须根据现实数据的特点来进行假设，而不是为了使用该算法而强行对数据进行不合理的假设。

2 Problem 2

PCA 作为一个算法，能把高维数据投射为低维数据。PCA 作为一个模型，将二维数据投射到三维数据，然后增加一些高斯噪声，重复多次，这样可以假设生成的样本数据不相关，然后反方向在三维空间中的二维平面上根据 Bayes 定理找出后验分布。通过 PCA 得到明确的假设条件。

3 Problem 3

贝叶斯框架易于实现，高效。在文本分类处理中，样本数据是独立同分布这一假设是成立的，此时运用贝叶斯方法分类效果较好。