

Linear Classification 01 Introduction

Chen Gong

29 October 2019

本节的主要目的是，有关于机器学习的导图。对频率派的有关统计学习方法做一个大致的梳理。而在贝叶斯派的学习中，是使用有关于概率图的模型。在频率派的有关统计学习方法中，我们可以大致的分为，线性回归和线性分类。

1 线性回归

在前文中已经提到了，我们的线性回归模型可以写为 $f(w, b) = w^T x + b$ 。线性回归主要有三条性质：线性，全局性和数据未加工。而我们从每一条入手，打破其中的一条规则就是一个新的算法。

1.1 线性

线性可以分为，属性非线性，全局非线性和系数非线性。

1.1.1 属性非线性

所谓的属性非线性也就是从未知数入手，比如特征变换的方法还有将变量从一维，变换到高维。有点类似于引入二次型的思想，使用 $x_1^2 + x_2^2 + x_1 x_2 + \dots$ ，的方法打破属性的线性。

1.1.2 全局非线性

全局非线性的方法，是通过对函数的运算结果增加一个函数，来将线性函数改造成非线性函数。比如，神经网络中的激活函数，还有阈值函数来将软分类函数变成硬分类函数。

1.1.3 系数非线性

所谓系数非线性，感觉就是系数的生成结果并不是单一的，固定的。就像神经网络算法一样。算法的收敛结果是一个分布，也就是位于一个区间之中，这样的算法的结果一定不是线性的，这样通过不确定的方法来引入非线性。

1.2 全局性

所谓全局性，也就是将所有数据看成一个整体来进行拟合。而打破的方法很简单，也就是将数据之间分隔开，分段进行拟合。典型的方法有线性样条回归，决策树等方法。

1.3 数据未加工

从字面的意义上理解非常的简单，那就是输入数据不经过加工直接的输入模型中。有一系列类似的方法来打破，比如主成分分析法 (PCA)，流形等方法来对输入数据进行预处理。

2 线性分类

线性回归和线性分类之间有着很大的联系。从某种意义上说，线性分类就是线性回归函数使用激活函数的结果，同时也可以看成是线性回归降维的结果。对于一个线性回归函数，我们可以通过添加全局函数的形式来将其转换为线性分类函数。也就是

$$y = w^T x + b \longrightarrow y = f(w^T x + b) \quad (1)$$

这样就可以将值域从 $[0, 1]$ 转换为 $\{0, 1\}$ 。其中 f 被定义为 activation function, f^{-1} 定义为 link function。那么这个 f 实现了这样一个功能，也就是将 $w^T x + b \mapsto \{0, 1\}$ 。而 f^{-1} 恰好是反过来的，也就是将 $\{0, 1\} \mapsto w^T x + b$ 。

而线性分类，大致上可以划分成硬分类和软分类两个部分。

2.1 硬分类

所谓硬分类，也就是 $y \in [0, 1]$ ，大致上可以分成线性判别分析，也就是 Fisher 判别分析和感知机这两类。

2.2 软分类

所谓硬分类，也就是 $y \in \{0, 1\}$ ，大致上可以分成生成式模型，Gaussian Discriminate Analysis 和著名的判别式模型，Logistic Regression。

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)} \propto p(x|y)p(y) \quad (2)$$

也就是在求解 $p(y = 0|x)$ 或 $p(y = 1|x)$ 的时候，我们不直接求谁大谁小，而是转向求 $p(x|y = 0)p(y = 0)$ 和 $p(x|y = 1)p(y = 1)$ ，即求联合概率。

3 总结

通过这节的学习，我们已经大体上建立了有关于统计学习方法的知识的框架，包括线性分类和线性回归的内容，并作出了一定的梳理。