

《统计学习方法》读书笔记

3月1日

丁自民

10.3 学习算法

1、本节介绍了两种算法，一种是监督学习方法，一种是无监督学习方法

(Baum-Welch 算法，即 EM 算法)。

2、监督学习方法需要标签化的数据集。(就像下面蓝色标注数据)

10.3.1 监督学习方法

假设已给训练数据包含 S 个长度相同的观测序列和对应的状态序列 $\{(O_1, I_1), (O_2, I_2), \dots, (O_S, I_S)\}$ ，那么可以利用极大似然估计法来估计隐马尔可夫模型的参数。具体方法如下。

然后用普通的贝叶斯公式估算参数。

3、Baum-Welch 算法只需要一系列观测值作为训练数据集就好，不需要标注。

(就像下面红色标注数据)

10.3.2 Baum-Welch 算法

假设给定训练数据只包含 S 个长度为 T 的观测序列 $\{O_1, O_2, \dots, O_S\}$ 而没有对应的状态序列，目标是学习隐马尔可夫模型 $\lambda = (A, B, \pi)$ 的参数。我们将观测序列数据看作观测数据 O ，状态序列数据看作不可观测的隐数据 I ，那么隐马尔可夫模型事实上是一个含有隐变量的概率模型

后面的思路与 EM 算法一样。

E: 写出联合概率密度

M: 将联合概率密度公式通过 Lagrange 方法最大化，求出参数。