

实验 2. 隐马尔科夫模型实践

2017 年 11 月 22 日

综述

本次试验将带你一步一步实现隐马尔科夫模型 (Hidden Markov Model, HMM), 并将其应用在金融时序数据分析与预测方面。具体而言, 对于一个已经训练好的 HMM, 你将实现实现一个维特比算法¹, 通过动态规划的思想对模型进行推断, 其次, 如果 HMM 的参数未知, 则需要通过数据进行学习与训练, 这里我们将部分实现 Baum-Welch_algorithm. 你将负责其中两个关键函数: HMM 的前向与后向算法。最后, 我们将利用自己从零开始写好的 HMM, 进行股票的涨跌预测 (for fun), 我们将预测中国某支与 AI 相关的股票的走势。

HMM 的框架已经写好, 你只需要实现部分分类方法即可。框架文件参见 myHMM.py (在作业的附件中), 需要你实现的函数, 已经定义好了函数名称和输入输出, 并已明确注明了 “TODO”。未标记 TODO 的地方, 不用进行修改。所有数据处理和获得的函数已帮你实现好。

在进行实验之前, 请参阅 [1] 相关章节, 以及维基百科上对 Forward-backward algorithm²和 Baum-Welch algorithm³的算法描述。

问题背景

对于单只股票数据, 我们每日可以观测到的值可以是涨、跌, 不涨不跌 (相对于昨天的收盘价) 三种情况。这里将观测的涨跌平编码为 0,1,2 三种取值 (0: 跌, 1: 涨, 2: 平)。

我们假设股票的涨跌由内在的隐变量驱动 (这是一个十分简化的假设), 即牛市或熊市。换言之, 牛市 (编码为 1) 比较有可能驱动股票价格上涨, 熊市 (编码为 0) 比较有可能驱动股票下跌。换言之, 在本 HMM 模型中, 隐变量仅仅是 1 维的 0/1 离散状态。

实验一. 维特比算法 (40%)

本任务将实现维特比算法。请实现 myHMM.py 中 HMMViterbi 这个函数, 具体而言:
输入: a, b, o, pi

¹https://en.wikipedia.org/wiki/Viterbi_algorithm

²https://en.wikipedia.org/wiki/Forward-backward_algorithm

³https://en.wikipedia.org/wiki/Baum-Welch_algorithm

输出: path

这里, a 是状态转移矩阵 (transition matrix), b 是发射矩阵 (emission matrix), o 是观测到的序列 (例如 $[0, 1, 1, 1, 2, 1, 0, 1, 0]$, π 是初始分布 initial probabilities) 输出是最有可能的隐变量序列 (0,1 序列)

换言之, 如果别人告诉你 a , b 和 π 这三个参数, 你可以通过观测实际序列, 对隐状态进行推断了。

实验二. 实现 Forward Algorithm (30%)

Forward/backward algorithm 是 Baum-Welch algorithm 的关键步骤。请先完全掌握 Baum-Welch algorithm 算法的具体过程, 并详细阅读 myHMM.py 中对 BW 算法的实现步骤。

本次作业, 其实 BW 算法的流程和主体已经实现好, 除了两个子函数的调用需要你来实现, 本任务将实现 Forward Algorithm 这个子函数。

具体而言, 对于任务 2, 请实现 myHMM.py 中 HMMfwd(self, a , b , o , π) 这个函数, 其中:

输入: a, b, o, π

输出: α

这里, a 是状态转移矩阵 (transition matrix), b 是发射矩阵 (emission matrix), o 是观测到的序列, 输出的 α 是 $N * T$ 的 numpy 数组, 其中维度和数据格式已先行初始化。这里的命名均遵从通用称呼, 具体意义请自行参阅维基百科。

实验三. 实现 Backward Algorithm (30%)

本任务将实现 myHMM.py 中 HMMbwd(self, a , b , o) 这个函数, 其中:

输入: a, b, o, π

输出: β

这里, a 是状态转移矩阵 (transition matrix), b 是发射矩阵 (emission matrix), o 是观测到的序列, 输出的 β 也是 $N * T$ 的 numpy 数组, 其中维度和数据格式已先行初始化。这里的命名均遵从通用称呼, 具体意义请自行参阅维基百科。

实验检测

至此, 你已经学会如何从数据中训练一个 HMM, 并通过维特比算法进行推断。

为了检验你的程序, 请运行 test_myHMM.py 脚本, 这个脚本将验证代号为 002415 的股票在今年的预测准确度, 应该在 64.7% 左右。

附加任务. Do something EXTRA! (10%)

基于这个背景，做任何你觉得有意思的事情，以额外的文件和说明文档提交。例如增加隐变量状态，etc.

实验结果提交

提交的实验结果.zip 文件中需包含 myHMM.py，以及一份实验报告（pdf 格式），实验结果的提交方法和命名规则见课程主页。

注意!

本次实验使用的方法十分原始，请勿用于实际投资。

参考文献

[1] 周志华. 机器学习. 清华大学出版社, 2016.