深度之眼
deepshare.net

法律声明

课程详情请咨询

- 微信公众号：深度之眼
- 客服微信号：deepshare0920

公众号　　微 信

deepshare.net
深度之眼

西瓜书公式推导

导师：Sm1les（Datawhale南瓜书项目负责人）

deepshare.net
深度之眼

隐马尔可夫模型公式推导

Derivation of Hidden Markov Model

本节大纲

Outline

先修内容：西瓜书14.1、统计学习方法第10章

1.隐马尔可夫模型简介

2.概率计算问题

3.学习问题

4.预测问题

隐马尔可夫模型简介

Introduction to Hidden Markov Model

1.定义

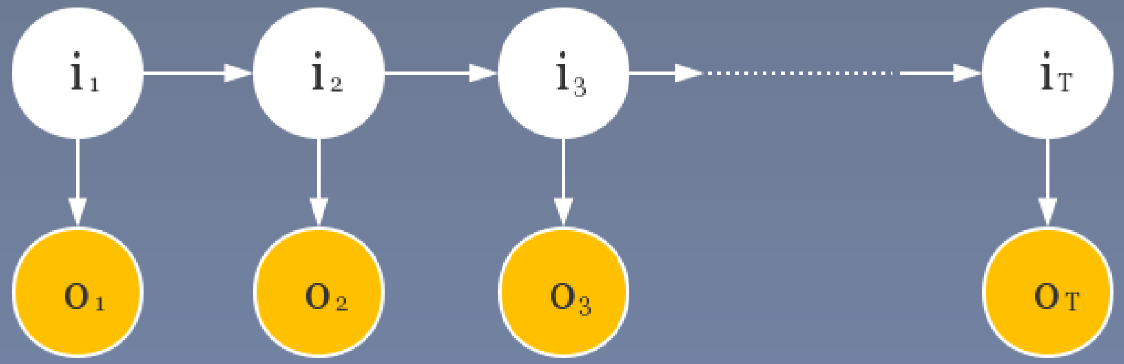
2.两个基本假设

3.三个基本问题

隐马尔可夫模型简介

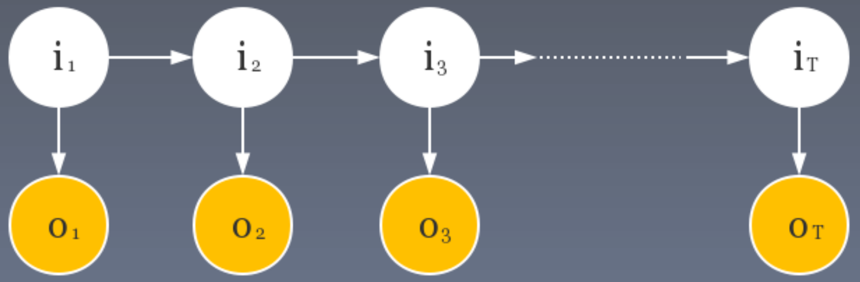
Introduction to Hidden Markov Model

定义：隐马尔可夫模型（Hidden Markov Model，HMM）是关于时序的概率模型，描述由一个隐藏的马尔可夫链随机生成不可观测的状态随机序列，再由各个状态生成一个观测而产生观测随机序列的过程。隐藏的马尔可夫链随机生成的状态的序列，称为状态序列；每个状态生成一个观测，而由此产生的观测的随机序列，称为观测序列，序列的每一个位置又可以看作是一个时刻。



隐马尔可夫模型简介

Introduction to Hidden Markov Model



Q是所有N种可能的状态的集合：$Q = \{q_1, q_2, ..., q_N\}$

V是所有M种可能的观测的集合：$V = \{v_1, v_2, ..., v_M\}$

I是长度为T的状态序列，O是对应的观测序列：$I = (i_1, i_2, ..., i_T), O = (o_1, o_2, ..., o_T)$

A是状态转移概率矩阵：$A = [a_{ij}]_{N\times N}$ 其中，$a_{ij} = P(i_{t+1} = q_j | i_t = q_i), \quad i = 1, 2, ..., N; j = 1, 2, ..., N$

B是观测概率矩阵：$B = [b_{jk}]_{N\times M}$ 其中，$b_{jk} = P(o_t = v_k | i_t = q_j), \quad j = 1, 2, ...N; k = 1, 2, ..., M$

π是初始状态概率向量：$\pi = (\pi_1, \pi_2, ..., \pi_N)$ 其中，$\pi_i = P(i_1 = q_i), \quad i = 1, 2, ..., N$

隐马尔可夫模型由初始状态概率向量π、状态转移概率矩阵A和观测概率矩阵B决定。π和A决定状态序列，B决定观测序列。因此，隐马尔可夫模型λ可以用三元符号表示，即：$\lambda = (A, B, \pi)$

隐马尔可夫模型简介

Introduction to Hidden Markov Model

两个基本假设:

1. 齐次马尔可夫性假设，即假设隐藏的马尔可夫链在任意时刻t的状态只依赖于其前一时刻的状态，与其他时刻的状态及观测无关，也与时刻t无关:

$$P(i_t|i_{t-1},o_{t-1},...,i_1,o_1)=P(i_t|i_{t-1})$$

2. 观测独立性假设，即假设任意时刻的观测只依赖于该时刻的马尔可夫链的状态，与其他观测及状态无关:

$$P(o_t|i_T,o_T,i_{T-1},o_{T-1},...,i_t,i_{t-1},o_{t-1},...,i_1,o_1)=P(o_t|i_t)$$

隐马尔可夫模型简介

Introduction to Hidden Markov Model

三个基本问题：

1. 概率计算问题：给定模型 $\lambda=(A,B,\pi)$ 和观测序列 $O=(o_1,o_2,...,o_T)$，计算在模型λ下观测序列O出现的概率 $P(O|\lambda)$；

2. 学习问题：已知观测序列 $O=(o_1,o_2,...,o_T)$ 估计模型 $\lambda=(A,B,\pi)$ 参数，使得在该模型下观测序列概率 $P(O|\lambda)$ 最大，即用极大似然估计的方法估计参数；

3. 预测问题：也称为解码问题，已知模型 $\lambda=(A,B,\pi)$ 和观测序列 $O=(o_1,o_2,...,o_T)$，求对给定观测序列条件概率 $P(I|O)$ 最大的状态序列 $I=(i_1,i_2,...,i_T)$，即给定观测序列，求最有可能的对应的状态序列。

概率计算问题

Probability calculation problem

1.直接计算法

2.前向算法

3.后向算法

概率计算问题

Probability calculation problem

直接计算法:

对于求 $P(O|\lambda)$ 最直接的方法就是按照概率公式直接计算，即:

$$P(O|\lambda) = \sum_I P(O, I|\lambda) = \sum_I P(O|I, \lambda)P(I|\lambda)$$

其中，$P(I|\lambda)$ 表示给定模型参数时，产生状态序列 $I = (i_1, i_2, ..., i_T)$ 的概率:

$$P(I|\lambda) = \pi_{i_1} a_{i_1 i_2} a_{i_2 i_3} \cdots a_{i_{T-1} i_T}$$

概率计算问题

Probability calculation problem

$P(O|I,\lambda)$ 表示给定模型参数且状态序列为 $I=(i_1,i_2,...,i_T)$ 时，产生观测序列 $O=(o_1,o_2,...,o_T)$ 的概率：

$$P(O|I,\lambda)=b_{i_1o_1}b_{i_2o_2}...b_{i_To_T}$$

所以

$$\begin{aligned} P(O|\lambda) &= \sum_I P(O|I,\lambda)P(I|\lambda) \\ &= \sum_{i_1,i_2,\dots,i_T} \pi_{i_1} b_{i_1o_1} a_{i_1i_2} b_{i_2o_2} \cdots a_{i_{T-1}i_T} b_{i_To_T} \end{aligned}$$

其中，$\sum_{i_1,i_2,\dots,i_T}$ 共有 N^T 种可能，计算 $\pi_{i_1} b_{i_1o_1} a_{i_1i_2} b_{i_2o_2} \cdots a_{i_{T-1}i_T} b_{i_To_T}$ 的时间复杂度为 $O(T)$，所以上式整体的时间复杂度为 $O(TN^T)$，显然这种算法是不可行的。

概率计算问题

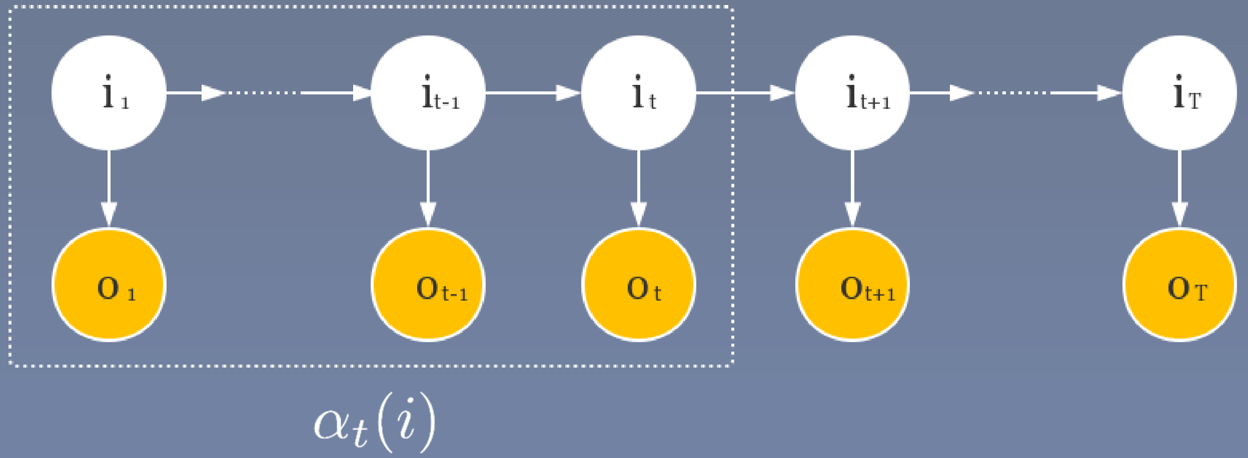
Probability calculation problem

前向算法:

首先定义前向概率：给定隐马尔可夫模型 λ，定义到时刻t部分观测序列为$o_1, o_2, ..., o_t$

且状态为q_i 的概率为前向概率，记作:

$$\alpha_t(i) = P(o_1, o_2, ..., o_t, i_t = q_i|\lambda)$$



概率计算问题

Probability calculation problem

根据前向概率的定义可推得:

$$P(O|\lambda) = P(o_1, o_2, ..., o_T|\lambda) = \sum_{i=1}^{N} P(o_1, o_2, ..., o_T, i_T = q_i|\lambda) = \sum_{i=1}^{N} \alpha_T(i)$$

于是求解 $P(O|\lambda)$ 的问题被转化为了求解前向概率 $\alpha_t(i)$ 的问题。由前向概率的定义可知:

$$\alpha_1(i) = P(o_1, i_1 = q_i|\lambda) = \pi_i b_{io_1}$$

$$\alpha_2(i) = P(o_1, o_2, i_2 = q_i|\lambda) = [\sum_{j=1}^{N} \alpha_1(j) a_{ji}] \times b_{io_2}$$

$$\alpha_3(i) = P(o_1, o_2, o_3, i_3 = q_i|\lambda) = [\sum_{j=1}^{N} \alpha_2(j) a_{ji}] \times b_{io_3}$$

依次此类推可得如下递推公式:

概率计算问题

Probability calculation problem

deepshare.net
深度之眼

$$\alpha_{t+1}(i)=[\sum_{j=1}^{N}\alpha_t(j)a_{ji}]\times b_{io_{t+1}}$$

因此可以递推求得：

$$\alpha_T(i)=[\sum_{j=1}^{N}\alpha_{T-1}(j)a_{ji}]\times b_{io_T}$$

将上式所求结果代回

$$P(O|\lambda)=P(o_1,o_2,...,o_T|\lambda)=\sum_{i=1}^{N}P(o_1,o_2,...,o_T,i_T=q_i|\lambda)=\sum_{i=1}^{N}\alpha_T(i)$$

即可求得 $P(O|\lambda)$

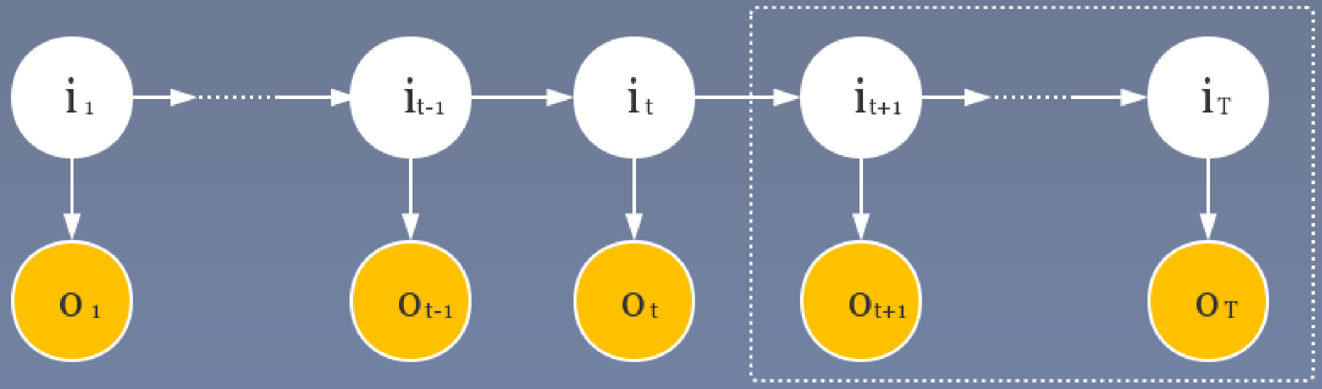
概率计算问题

Probability calculation problem

后向算法：

同前向算法一样，首先定义后向概率：给定隐马尔可夫模型λ，定义在时刻t状态为 q_i 的条件下，从t+1到T的部分观测序列为 $o_{t+1}, o_{t+2}, ..., o_T$ 的概率为后向概率，记作：

$$\beta_t(i) = P(o_{t+1}, o_{t+2}, ..., o_T | i_t = q_i, \lambda)$$



$\beta_t(i)$

概率计算问题

Probability calculation problem

由后向概率的定义可知:

$$\beta_T(i) = P(i_T = q_i, \lambda) = 1$$

$$\beta_{T-1}(i) = P(o_T | i_{T-1} = q_i, \lambda) = \sum_{j=1}^{N} a_{ij} b_{j o_T} \beta_T(j)$$

$$\beta_{T-2}(i) = P(o_{T-1}, o_T | i_{T-2} = q_i, \lambda) = \sum_{j=1}^{N} a_{ij} b_{j o_{T-1}} \beta_{T-1}(j)$$

依次类推可得递推公式:

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_{j o_{t+1}} \beta_{t+1}(j)$$

概率计算问题

Probability calculation problem

根据递推公式可求得 $\beta_1(i)$ 又

$$P(O|\lambda) = P(o_1, o_2, ..., o_T|\lambda) = \sum_{i=1}^{N} P(o_1, i_1 = q_i|\lambda)P(o_2, o_3, ..., o_T|i_1 = q_i, \lambda) = \sum_{i=1}^{N} \pi_i b_{io_1} \beta_1(i)$$

所以也可求得 $P(O|\lambda)$

综上可以看出前向算法和后向算法都是先计算局部概率，然后递推到全局，每一时刻的概率计算都会用上前一时刻计算出的结果，整体的时间复杂度大约为 $O(TN^2)$明显小于直接计算法的 $O(TN^T)$

概率计算问题

Probability calculation problem

利用前向概率和后向概率，可以得到关于单个状态和两个状态概率的一些计算公式：

1.给定模型参数λ和观测O，在时刻t处于状态 q_i 的概率，记为：$\gamma_t(i) = P(i_t = q_i|O, \lambda)$

可以通过前向概率和后向概率进行计算，推导如下：

$$\gamma_t(i) = P(i_t = q_i|O, \lambda) = \frac{P(i_t = q_i, O|\lambda)}{P(O|\lambda)} = \frac{P(i_t = q_i, O|\lambda)}{\sum_{j=1}^{N} P(i_t = q_j, O|\lambda)}$$

又由前向概率和后向概率的定义可知：

$$\alpha_t(i)\beta_t(i) = P(o_1, o_2, ..., o_t, i_t = q_i|\lambda)P(o_{t+1}, o_{t+2}, ..., o_T|i_t = q_i, \lambda) = P(i_t = q_i, O|\lambda)$$

所以

$$\gamma_t(i) = \frac{P(i_t = q_i, O|\lambda)}{\sum_{j=1}^{N} P(i_t = q_j, O|\lambda)} = \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^{N} \alpha_t(j)\beta_t(j)}$$

概率计算问题

Probability calculation problem

2. 给定模型参数λ和观测O，在时刻t处于状态 q_i 且在时刻t+1处于状态 q_j 的概率，记为：

$$\xi_t(i,j) = P(i_t = q_i, i_{t+1} = q_j | O, \lambda)$$

可以通过前向概率和后向概率进行计算，推导如下：

$$\xi_t(i,j) = \frac{P(i_t = q_i, i_{t+1} = q_j, O|\lambda)}{P(O|\lambda)} = \frac{P(i_t = q_i, i_{t+1} = q_j, O|\lambda)}{\sum_{i=1}^{N}\sum_{j=1}^{N} P(i_t = q_i, i_{t+1} = q_j, O|\lambda)}$$

又 $P(i_t = q_i, i_{t+1} = q_j, O|\lambda) = P(i_t = q_i, i_{t+1} = q_j, o_1, o_2, ..., o_T|\lambda)$

$$= P(o_1, o_2, ..., o_t, i_t = q_i|\lambda)P(o_{t+1}, i_{t+1} = q_j|i_t = q_i, \lambda)P(o_{t+2}, o_{t+3}, ..., o_T|i_{t+1} = q_j, \lambda)$$

$$= \alpha_t(i)a_{ij}b_{jo_{t+1}}\beta_{t+1}(j)$$

所以

$$\xi_t(i,j) = \frac{\alpha_t(i)a_{ij}b_{jo_{t+1}}\beta_{t+1}(j)}{\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_t(i)a_{ij}b_{jo_{t+1}}\beta_{t+1}(j)}$$

结 语

在这次课程中，我们学习了隐马尔可夫模型的简介和概率计算问题

那么在下次课程中，我们将会学习

隐马尔可夫模型的学习问题