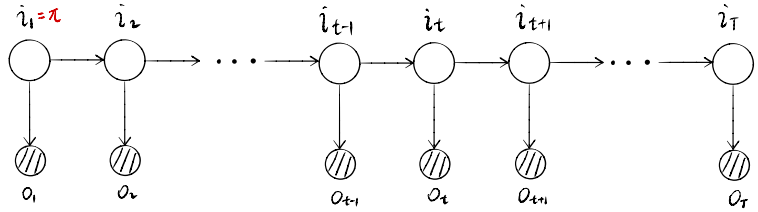
隐马尔可夫模型（Hidden Markov Model, HMM）

(1) 一张图



HMM模型的3个重要参数：

$\lambda = \{\pi, A, B\}$

A → 决定了观测序列

π → 决定了隐状态序列

① Q为隐状态集合$\{q_1, q_2, \cdots, q_N\}$

e.g. $Q=\{$晴，阴，雨，雪$\}$

② I为隐状态序列$\{i_1, i_2, \cdots, i_T\}$

e.g. $I=\{$晴，晴，阴，晴，雨，阴，雪，晴$\}$

③ π为i_1（即：隐状态序列在$t=1$时刻）的可能取值

④ V为可观测状态集合$\{v_1, v_2, \cdots, v_M\}$

e.g. $V=\{$出游，学习，发呆$\}$

⑤ O为可观测序列$\{o_1, o_2, \cdots, o_T\}$

e.g. $O=\{$出游，学习，发呆，学习，学习，发呆，学习，出游$\}$

⑥ A为隐状态转移概率矩阵

$A=[a_{ij}]_{N\times N}$

$p(i_{t+1}=q_j \mid i_t=q_i)$ 隐状态从（t时刻）q_i转变为（$t+1$时刻）q_j的概率

e.g.

	晴	阴	雨	雪	
晴	0.5	0.2	0.2	0.1	}1
阴	0.2	0.4	0.3	0.1	}1
雨	0.5	0.3	0.1	0.1	}1
雪	0.6	0.1	0.1	0.2	}1

$a_{34}=a_{雨\to雪}=0.1$ 即：t时刻是雨天，在$t+1$时刻变为下雪天的概率为0.1

⑦ B为发射概率矩阵 → 隐状态q_j的条件下，可观测到v_k的概率

$B=[b_j(k)]_{N\times M}$

k → 可观测状态v_k

M → 可观测状态集合元素个数（V）

j → 隐状态为q_j

N → 隐状态集合元素个数（Q）

e.g. 小张辞职后，行为有三种：$V=\{v_1, v_2, v_3\}=\{$出游，学习，发呆$\}$

不同天气条件下，v_i发生的概率是不一样的。

	出游	学习	发呆	
晴	0.4	0.5	0.1	}1
阴	0.2	0.7	0.1	}1
雨	0.1	0.3	0.1	}1
雪	0.3	0.5	0.2	}1

$b_{雪}(发呆)=0.2$ → 小张在下雪天发呆的概率是0.2

(2) 两个假设

HMM的成立，有两个前提假设：

① 齐次马尔可夫假设

"t时刻的隐状态，仅依赖于t-1时刻的隐状态"，即：

$$P(i_t \mid i_{t-1}, O_{t-1}, \cdots, i_1, O_1) = P(i_t \mid i_{t-1}) \qquad t=1,2,\cdots,T$$

② 观测独立性假设

"t时刻的观测值，仅依赖于t时刻的隐状态"，即：

$$P(O_t \mid i_T, O_T, i_{T-1}, O_{T-1}, \cdots, i_{t+1}, O_{t+1}, i_t, i_{t-1}, O_{t-1}, \cdots, i_1, O_1) = P(O_t \mid i_t) \qquad t=1,2,\cdots,T$$

(3) 三个问题

① 观测序列的概率计算问题

给定 $\begin{cases} \text{观测序列 } O=(O_1, O_2, \cdots, O_T) \\ \text{模型参数 } \lambda=(\pi, A, B) \end{cases}$

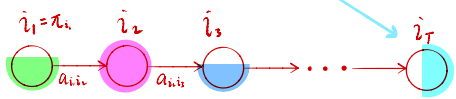
求：

(1) 整条观测序列O出现的概率 $P_\lambda(O)$

方法1：直接计算

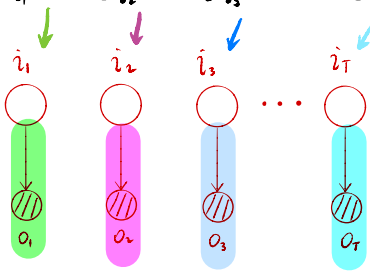
→ 隐状态序列I的概率（T个时刻隐状态的联合概率）

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



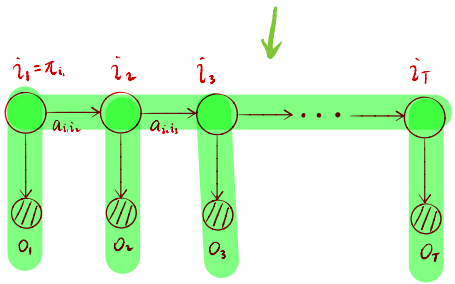
→ 可观测序列O的概率（T个时刻观测状态的联合概率）

$$P_\lambda(O \mid I) = b_{i_1}(O_1)\, b_{i_2}(O_2)\, b_{i_3}(O_3) \cdots b_{i_T}(O_T)$$



→ $P_\lambda(O, I) = P_\lambda(O \mid I)\, P_\lambda(I)$

$$= \pi_{i_1} b_{i_1}(O_1)\, a_{i_1 i_2} b_{i_2}(O_2) \cdots a_{i_{T-1} i_T} b_{i_T}(O_T)$$



$$\rightarrow P_\pi(O) = \sum_I P_\pi(O, I)$$

$$= \sum_{i_1, i_2, \cdots, i_T}^{N} P_\pi(O, I)$$

一共有T个变量，每个变量有N个可能的取值

⇓

共有N^T种组合 → 隐状态序列有N^T条
每条隐状态序列有T个时刻，每个时刻对应有一个发射概率

⇓

计算复杂度为$O(T \cdot N^T)$

⇓

当T、N较大时，计算不可行！

⇓

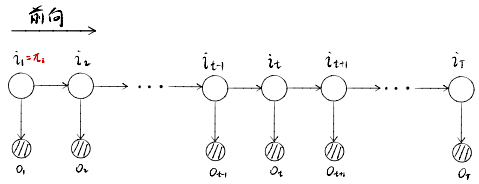
直接计算时，有很多步骤会被重复计算，所以我们可以
使用"动态规划"(Dynamic Programming)思想对计算过程进行改进

⇓

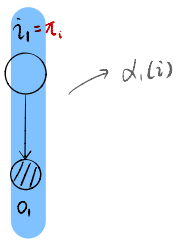
前向/后向算法

方法2：前向-后向(Forward-Backward)算法

(1) 前向算法



简记为i，$i=1,2,3,\cdots,N$

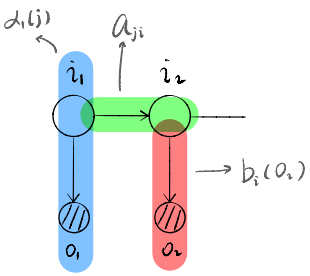


$t=1$时刻，在某一隐状态q_i下观测到O_1的概率为：

$\pi_i b_i(O_1)$，记为$\alpha_1(i)$

则在$t=1$时刻观测到O_1的概率为：

$$\sum_{i=1}^{N} \alpha_1(i)$$



$t=2$时刻，在某一隐状态q_i下观测到$O_{1:2}$的概率为：

$$\alpha_1(1)a_{1i}b_i(O_2) + \alpha_1(2)a_{2i}b_i(O_2) + \cdots + \alpha_1(N)a_{Ni}b_i(O_2)$$

$$= \sum_{j=1}^{N} \alpha_1(j) a_{ji} b_i(O_2)$$

$$= \left[\sum_{j=1}^{N} \alpha_1(j) a_{ji}\right] b_i(O_2)$$

因为j无关

$$= \alpha_2(i)$$

则在$t=2$时刻观测到$O_{1:2}$的概率为：

$$\sum_{i=1}^{N} \alpha_2(i)$$

$t=T$ 时刻，在某一隐状态 q_i 下观测到 $O_{1:T}$ 的概率为：

$$\left[\sum_{j=1}^{N} \alpha_{T-1}(j)\, a_{ji}\right] b_i(O_T) = \alpha_T(i)$$

则在 $t=T$ 时刻观测到 $O_{1:T}$ 的概率为：

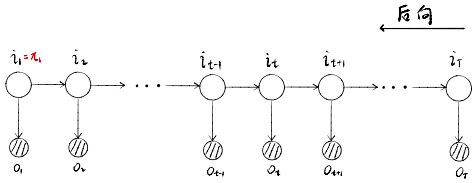
$$\sum_{i=1}^{N} \alpha_T(i)$$

$$\Downarrow$$

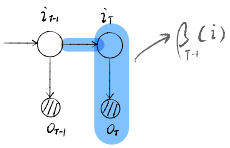
因为 $t=T$ 是最后一个时刻，至此我们计算出了完整的序列 O 的发生概率，即

$$P_\lambda(O) = \sum_{i=1}^{N} \alpha_T(i)$$

(1) 后向算法

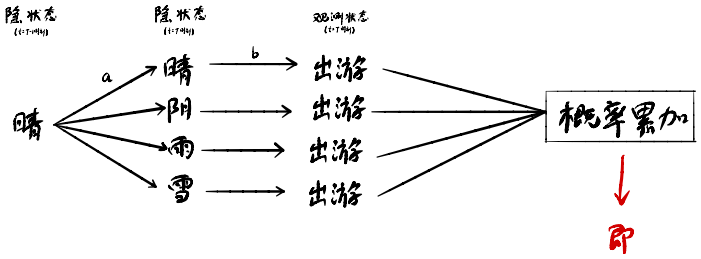


$t=T-1$ 时刻，在某一隐状态 q_i 下观测到 O_T 的概率为：



$$\sum_{j=1}^{N} a_{ij}\, b_j(O_T), \text{ 令其为 } \beta_{T-1}(i) \qquad i=1,2,3,\cdots,N$$

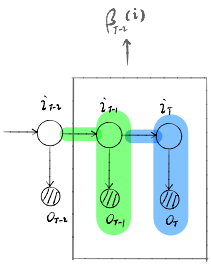
e.g.



在 $t=T-1$ 时刻隐状态为"晴"时，在 T 时刻能观测到"出游"的概率

则在 $t=T-1$ 时刻观测到 O_T 的概率为：

$$\sum_{i=1}^{N} \beta_{T-1}(i)$$



$t=T-2$ 时刻，在某一隐状态 q_i 下观测到 $O_{T-1:T}$ 的概率为：

$$\beta_{T-2}(i) = \sum_{j=1}^{N} a_{ij}\, b_j(O_{T-1})\, \beta_{T-1}(i) \qquad i=1,2,3,\cdots,N$$

则在 $t=T-2$ 时刻观测到 O_T 的概率为：

$$\sum_{i=1}^{N} \beta_{T-2}(i)$$

t-1时刻，在某一隐状态 q_i 下观测到 $O_{t:T}$ 的概率为：

$$\beta_{t-1}(i) = \sum_{j=1}^{N} a_{ij}\, b_j(O_t)\, \beta_t(i) \qquad i=1,2,3,\cdots,N$$

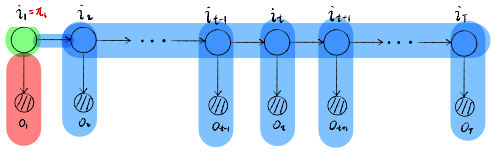
则在 t-1 时刻观测到 O_T 的概率为：

$$\sum_{i=1}^{N} \beta_{t-1}(i)$$

为了使 β_{T-1} 也满足这个通项公式，我们令 $\beta_T(i)=1$

t=1时刻，在某一隐状态 q_i 下观测到 $O_{2:T}$ 的概率为：

$$\beta_1(i) = \sum_{j=1}^{N} a_{ij}\, b_j(O_2)\, \beta_2(i) \qquad i=1,2,3,\cdots,N$$



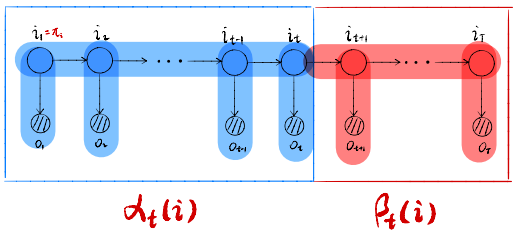
$$\Downarrow$$

那么，整条观测序列 $O_{1:T}$ 出现的概率值为

$$\sum_{j=1}^{N} \pi_i\, b_i(O_1)\, \beta_1(i)$$

(2) 求：t 时刻时，隐状态为 q_i 的概率

根据前向—后向算法，我们可以在t时刻将整个序列拆分成前向、后向两部分



$$P_\lambda(i_t=q_i, O) = \alpha_t(i)\cdot\beta_t(i)$$

$$\Rightarrow P_\lambda(i_t=q_i \mid O) = \frac{P_\lambda(i_t=q_i, O)}{P_\lambda(O)}$$

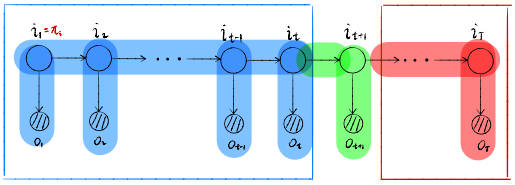
$$= \frac{P_\lambda(i_t=q_i, O)}{\sum_{j=1}^{N} P_\lambda(i_t=q_i, O)}$$

$$= \frac{\alpha_t(i)\,\beta_t(i)}{\sum_{j=1}^{N} \alpha_t(j)\,\beta_t(j)}$$

$$= \gamma_t(i) \rightarrow \text{记为 } \gamma_t(i)$$

<3> 求：t时刻时隐状态为q_i，在t+1时刻跳转为q_j的概率

根据前向—后向算法，我们可以在t时刻、t+1时刻将整个序列拆分成三部分



$\alpha_t(i)$　　$\beta_t(i)$

$$P_\lambda(i_t=q_i, i_{t+1}=q_j, O) = \alpha_t(i)\cdot a_{ij}b_j(O_{t+1})\cdot\beta_{t+1}(j)$$

$$\Rightarrow P_\lambda(i_t=q_i, i_{t+1}=q_j \mid O) = \frac{P_\lambda(i_t=q_i, i_{t+1}=q_j, O)}{P_\lambda(O)}$$

$$= \frac{P_\lambda(i_t=q_i, i_{t+1}=q_j, O)}{\sum_{m=1}^{N}\sum_{n=1}^{N} P_\lambda(i_t=q_m, i_{t+1}=q_n, O)}$$

$$= \frac{\alpha_t(i)\cdot a_{ij}b_j(O_{t+1})\cdot\beta_{t+1}(j)}{\sum_{m=1}^{N}\sum_{n=1}^{N} \alpha_t(m)\cdot a_{mn}b_n(O_{t+1})\cdot\beta_{t+1}(n)}$$

对隐状态序列进行预测

② 隐状态序列的概率计算问题　[又称为"解码(decoding)问题]

给定 $\begin{cases} \text{观测序列 } O=(O_1, O_2, \cdots, O_T) \\ \text{模型参数 } \lambda=(\pi, A, B) \end{cases}$

求：对应的最有可能的隐状态序列 $I=(i_1, i_2, \cdots, i_T)$ 是什么？

方法1：近似算法

我们之前计算过"在t时刻，隐状态为q_i的概率"，记为$\gamma_t(i)$
那么，我们直接取在每个时刻，概率最大的那个隐状态不就可以了吗？

即：$I^* = (i_1^*, i_2^*, \cdots, i_t^*, \cdots, i_T^*)$

$$= \max_{1\le i\le N} \gamma_t(i) \qquad t=1,2,\cdots,T$$

优点：计算简单

缺点：计算出来的最优隐状态序列，可能在实际的隐状态序列中并不存在。

比如：做分词时，有隐状态 B、M、E、S

B → Begin 词首　M → Middle 词中　E → End 词尾　S → Single 单字词

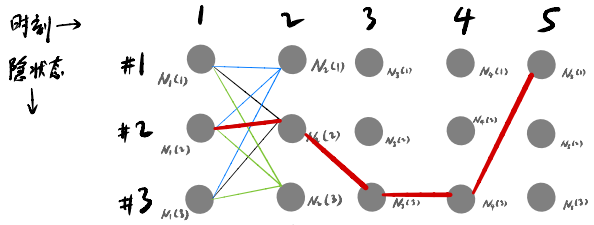
我们用"近似法"计算出的最大概率组合"可能是：

"B B E M S E"

↓

这是事实上不可能存在的隐状态序列

方法2：维特比（Viterbi）算法 → 本质上是"动态规划"求解



$$N_1(1) = \pi_1 \cdot b_1(O_1)$$
$$N_1(2) = \pi_2 \cdot b_1(O_1)$$
$$N_1(3) = \pi_3 \cdot b_1(O_1)$$

↓

$$N_2(3) = \max \begin{bmatrix} N_1(1) \cdot a_{13} \\ N_1(2) \cdot a_{23} \\ N_1(3) \cdot a_{33} \end{bmatrix} b_3(O_2)$$

→ 若 $N_1(2) \cdot a_{23}$ 的值最大，则在二维表中记录如下信息：

① $N_2(3)$ 的值

② 上一个节点为 $N_1(2)$ ← 路径回溯可得

⇓

写成通项公式的形式：

$$N_t(i) = \max_{1 \le j \le n} \left[N_{t-1}(j) a_{ji} \right] b_i(O_t)$$

然后，在每个时刻找最大值节点进行回溯，得到一条最佳路径

↓

即为最大概率出现的隐状态序列 I^*

② 模型参数 $\lambda = (\pi, A, B)$ 的学习问题

(1) 有监督学习

给定观测序列 $O = (O_1, O_2, \cdots, O_T)$

对应的隐状态为 $I = (i_1, i_2, \cdots, i_T)$

求：模型参数 λ

方法：因为已经知道 O 和 I，我们通过简单的"数数"和"占坑"计算即可得到 π, A, B

e.g. 分别观测小孙3周，记录如下：

(第一周)	晴	晴	阴	晴	雨	雨	阴	(I)
	↓	↓	↓	↓	↓	↓	↓	
	出游	学习	学习	出游	学习	发呆	学习	(O)
(第二周)	阴	晴	阴	雨	雪	雨	雪	(I)
	↓	↓	↓	↓	↓	↓	↓	
	学习	出游	学习	发呆	出游	学习	发呆	(O)
(第三周)	晴	晴	雨	晴	阴	雨	晴	(I)
	↓	↓	↓	↓	↓	↓	↓	
	出游	学习	学习	学习	学习	发呆	出游	(O)

我们可以统计出：

① 隐状态集合 $Q=$(晴, 阴, 雨, 雪)
观测状态集合 $V=$(出游, 学习, 发呆)

② 初始概率 π

$$\pi=\begin{pmatrix} \text{晴} & 2/3 \\ \text{阴} & 1/3 \\ \text{雨} & 0 \\ \text{雪} & 0 \end{pmatrix}$$

③ 状态转移矩阵 A

$$a_{\text{晴}\to\text{雨}}=\frac{\text{Count}(\overset{t\text{时刻}}{\text{晴}}\to\overset{t-1\text{时刻}}{\text{雨}})}{\text{Count}(\text{晴})}$$ (其他项，同理)

④ 发射概率矩阵 B

$$b_{\text{晴}}(\text{出游})=\frac{\text{Count}(\text{晴},\text{出游})}{\text{Count}(\text{晴})}$$ (其他项，同理)

(2) 有监督学习

给定观测序列 $O=(O_1, O_2, \cdots, O_T)$

求：模型参数 λ

方法：使用 Baum-Welch 算法（即 EM 算法）进行迭代估计

我自己还需要点时间梳理这部分的细节，过后更新！