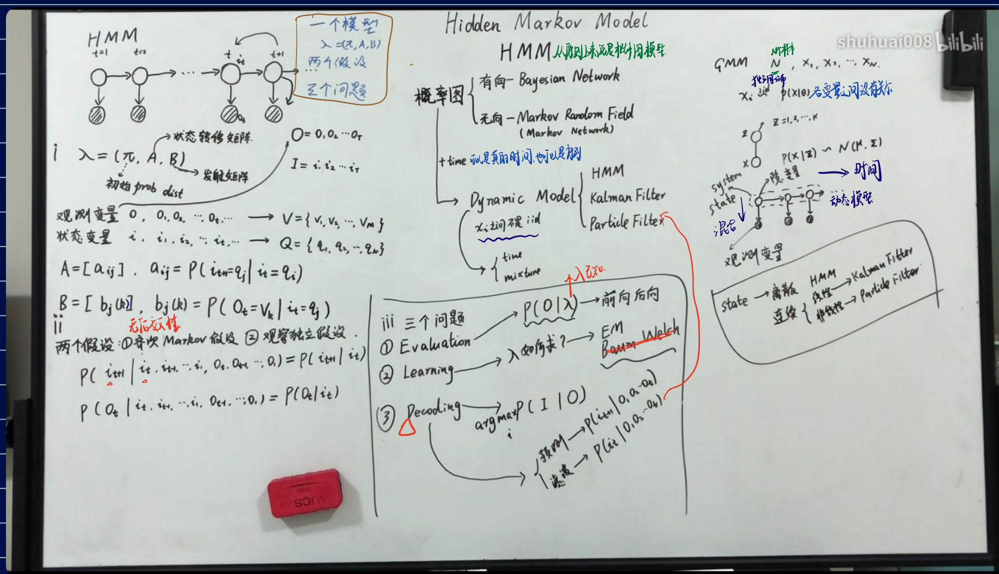
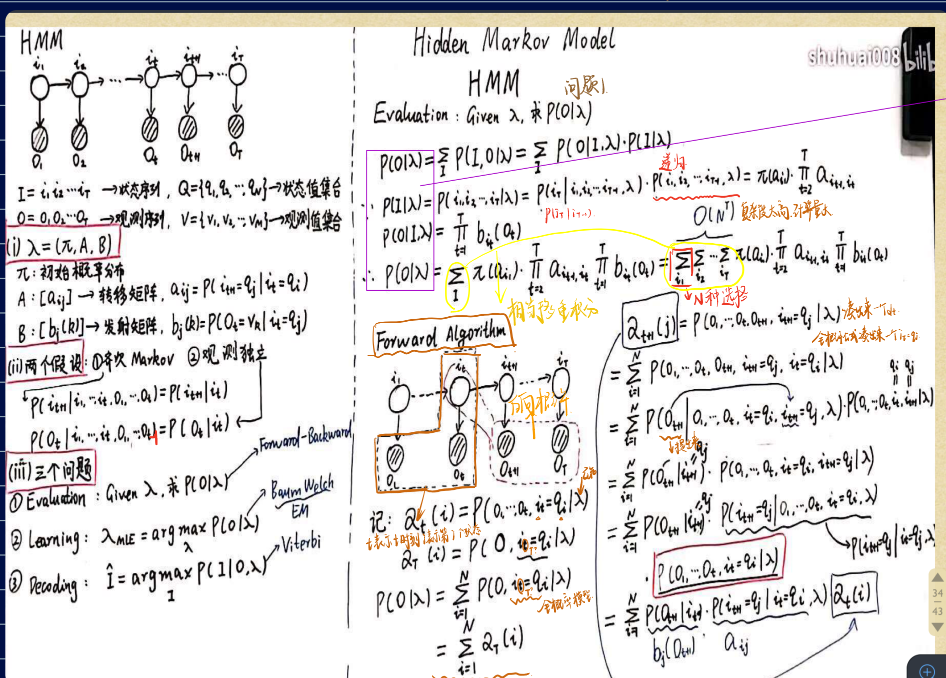
HMM在以前的NLP中使用的很多　　　补充一下概率图模型





⚠在记法上有点问题
$\alpha_T=P(O,i_T=q_i\mid\lambda)$?

这一小部分的回顾:

目的: $P(O\mid\lambda)$—在给定参数的条件下求观测的概率.
这个观测的概率是没有办法直接求的所以
转成了间的西条件来求.

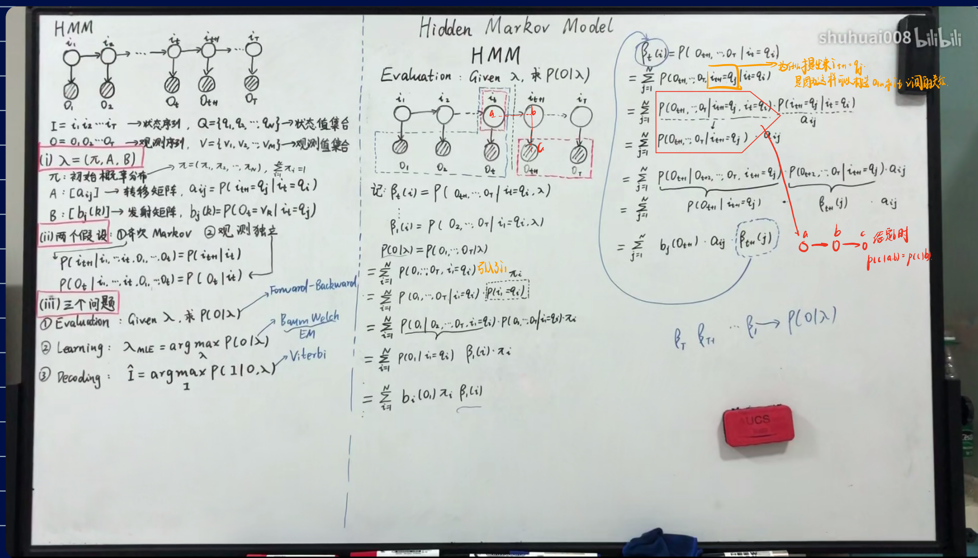
$$=\sum_{i=1}^{N}P(o_{t+1}\mid o_1\cdots o_t,i_t=q_i,i_{t+1}=q_j,\lambda)\cdot$$
$$P(o_1\cdots o_t,i_t=q_i,i_{t+1}=q_j\mid\lambda)$$
$$=\sum_{i=1}^{N}P(o_{t+1}\mid \cancel{o_t,i_t=q_i,}\,i_{t+1}=q_j,\lambda)\cdot P(o_1\cdots o_t,i_t=q_i,i_{t+1}=q_j\mid\lambda)$$
$$=\sum_{i=1}^{N}P(o_{t+1}\mid i_{t+1}\,\cancel{i_t=q_i,i_{t+1}=q_j},\lambda)\,\underline{P(i_{t+1}=q_j\mid o_1\cdots o_t,i_t=q_i,\lambda)}$$
$=P(i_{t+1}=q_j\mid i_t=q_i,\lambda)$
$$\cdot P(o_1\cdots o_t,i_t=q_i\mid\lambda)$$
$$=\sum_{i=1}^{N}P(o_{t+1}\mid i_{t+1})\,P(i_{t+1}=q_j\mid i_t=q_i,\lambda)\cdot\alpha_t(i)$$
$$\Rightarrow\alpha_{t+1}(j)=\sum_{i=1}^{N}P(o_{t+1}\mid i_{t+1})P(i_{t+1}=q_j\mid i_t=q_i,\lambda)\cdot\alpha_t(i)$$
$$=\sum_{i=1}^{N}b_j(o_{t+1})\cdot a_{ij}\cdot\alpha_t(i)$$

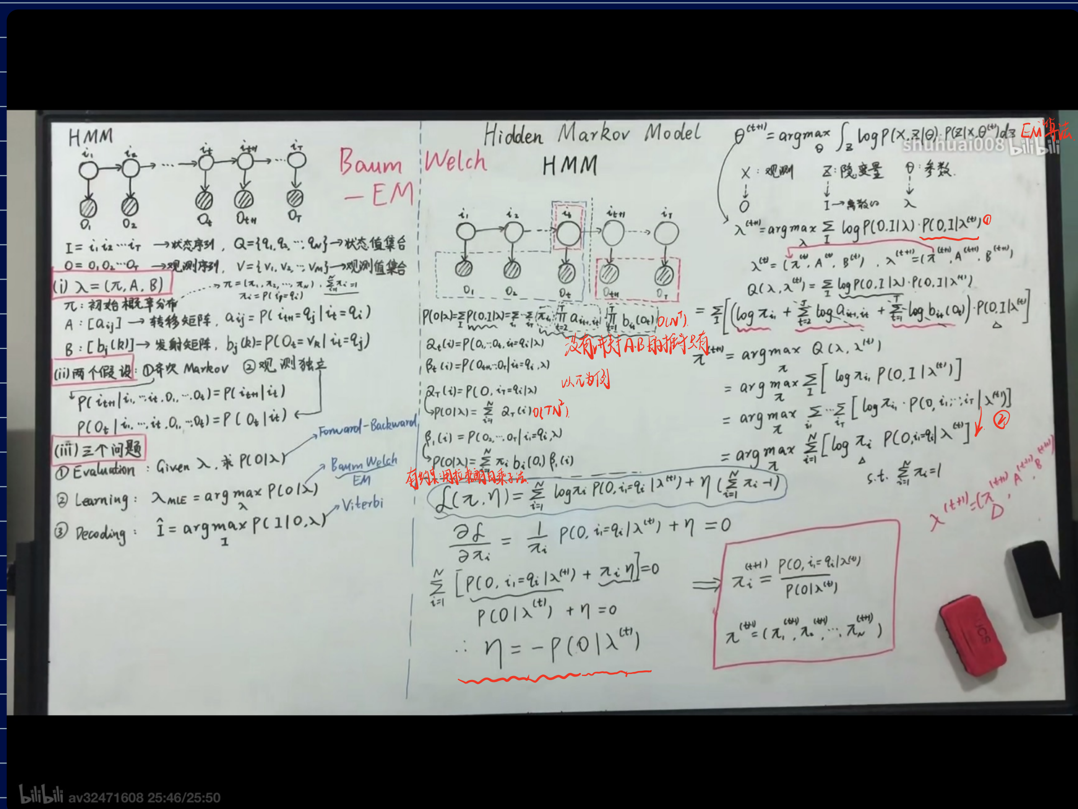
得到了上面的递推式 只用到了两个假设

频率派 → 统计机器学习
→ 优化
①定义出一个模型
②策略,即评估准则
③求解 GD SGD, 牛顿法 逆牛顿法

贝叶斯派 → 概率图模型
→ 推断 → $P(z\mid x)$
后验概率
→ 积分问题
MCMC
数值积分法

前向算法通过定义了一个 $\alpha_t(i)$ $\alpha_t(i) = P(O_1 \cdots O_t, i_t = q_i \mid \lambda)$





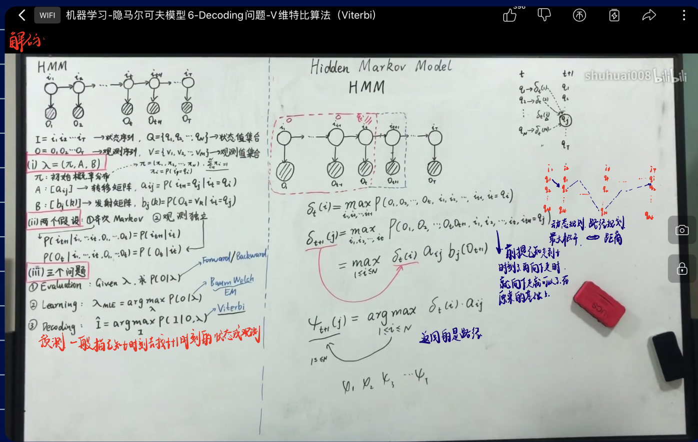
① 原来的 $P(I \mid O, \lambda) = \frac{P(O, I \mid \lambda^{(t)})}{P(O, \lambda^{(t)})}$

但是上面的分母和 λ 没有关系所以 $P(O, \lambda^{(t)})$ 可以去掉

只剩 $P(O, I \mid \lambda^{(t)})$ 也没有关系

② 从 $\sum_{i_1} \sum_{i_2} \cdots \sum_{i_N}$

变成只有 $\sum_{i_1}$ 就是保留 i_1 后对所有的概率进行累加加出来都是 1 所以去掉了 也可以理解成 固定一个把剩下的做全加和 得到的只有一个时间上的隐状态 原始的是 i_1 但是可以换成更加有普适性的如右侧所示.



HMM: mixture + time

混合: GMM. z 1~k 隐变量离散

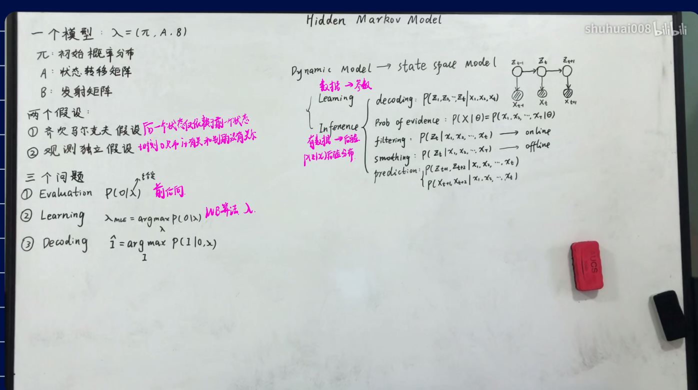
$z \to x$: 高斯分布

观测 x

HMM → GMM + Time

隐状态离散.

发射可以离散也可以连续



没听完 打算后面再听