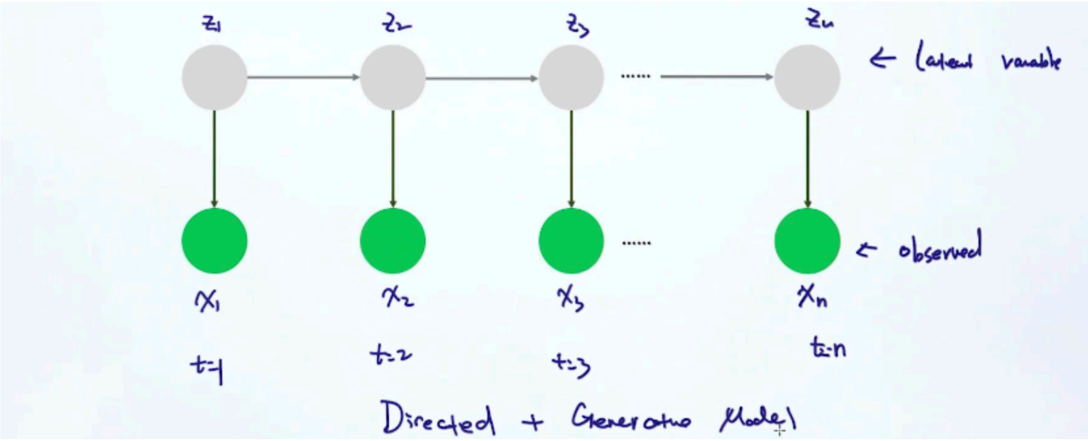
时序模型
HMM/CRF 传统模型
RNN/LSTM 深度学习模型

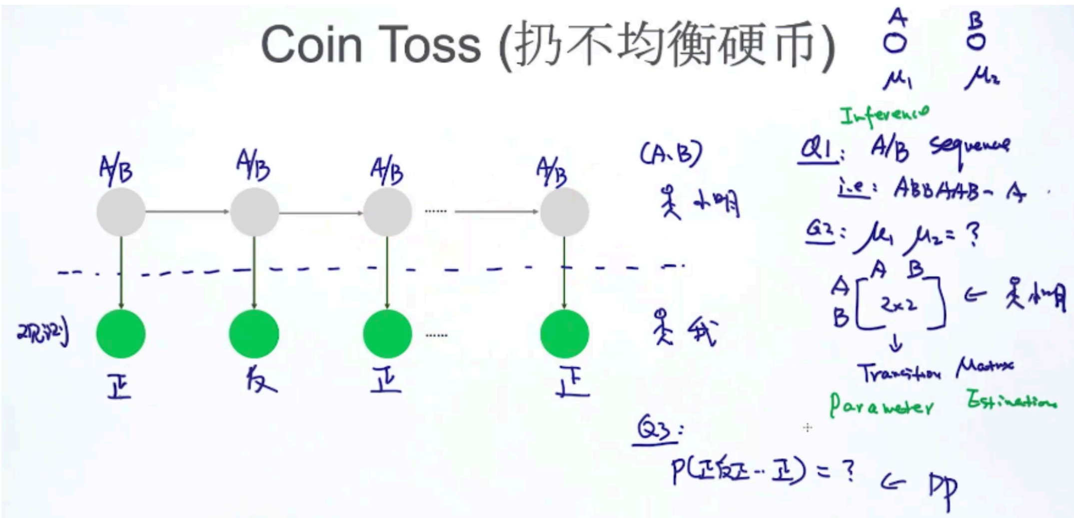
1. HMM 介绍



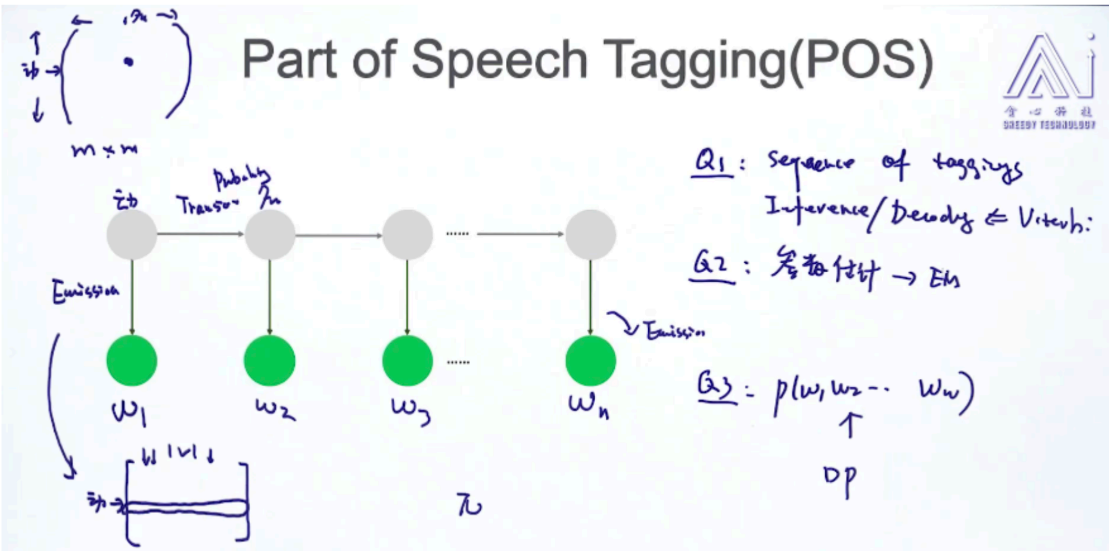
有方向的
生成的，可以生成新的数据
也可以做成判别模型

1.1 应用例子

1）Coin toss

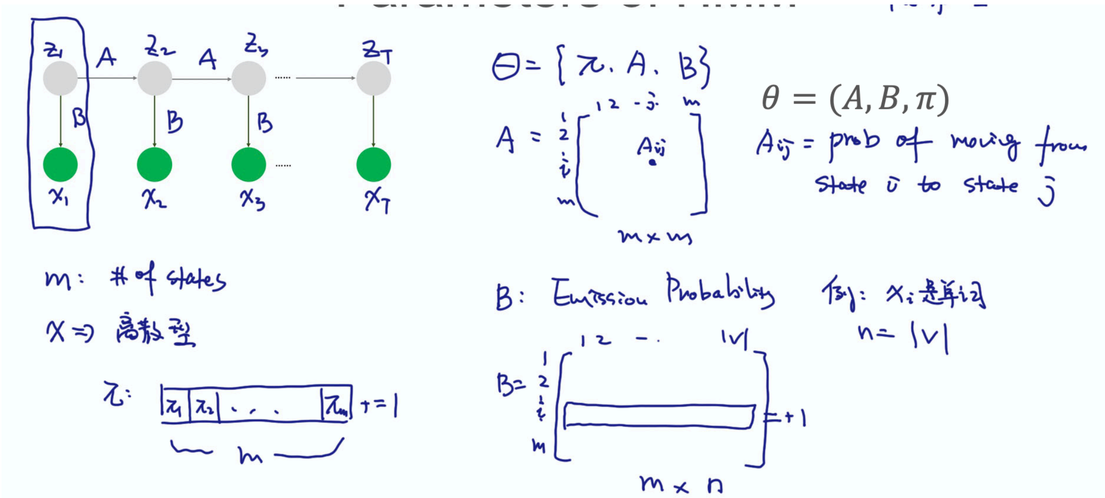


2）speech tagging



HMM 两个较重要的参数：转移概率、初始概率

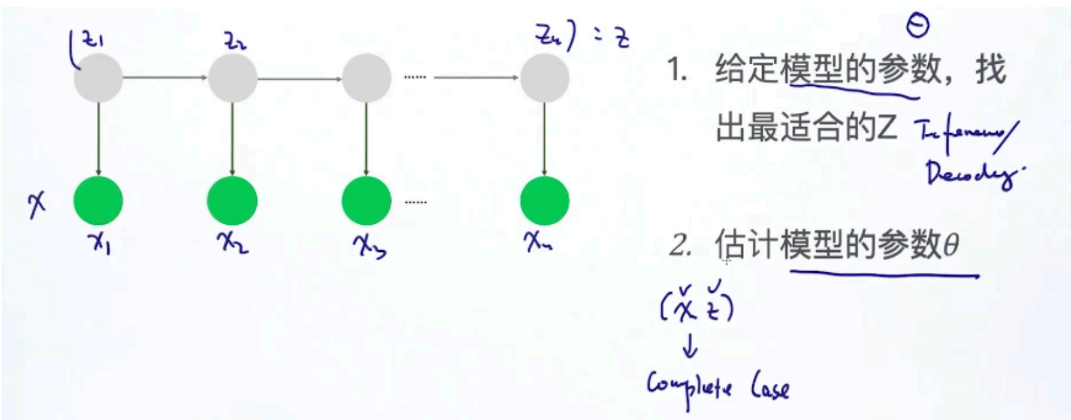
1.2 参数



状态转移的概率不是随机的

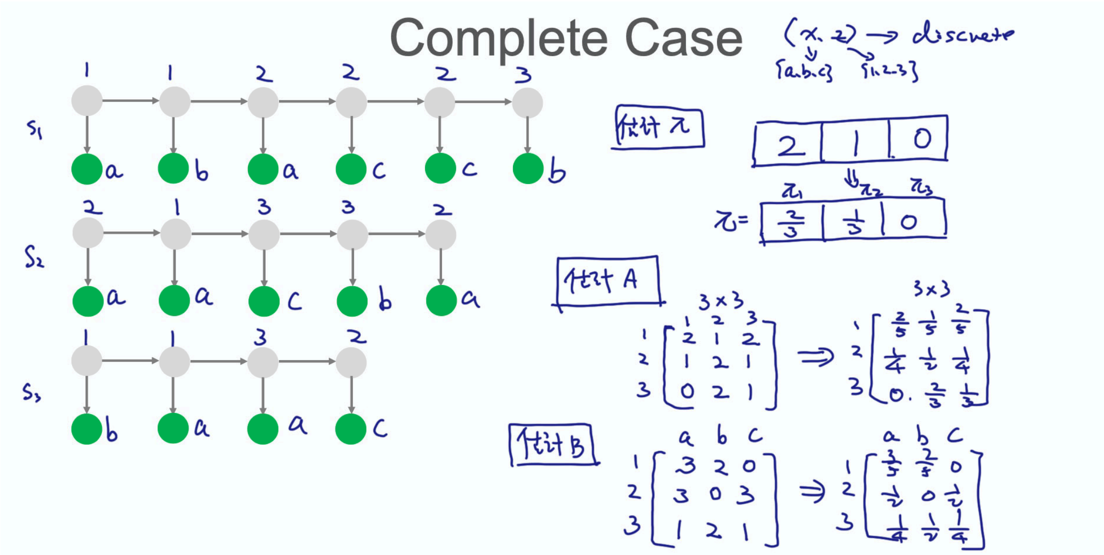
B 相当于生成矩阵（因此，HMM 也可以看成生成模型）

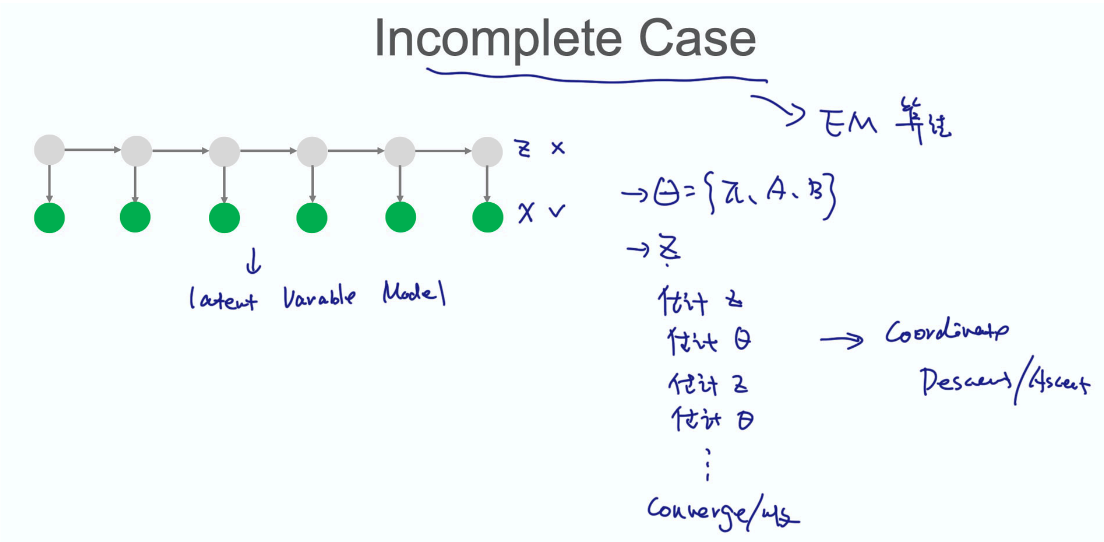
1.3 major tasks



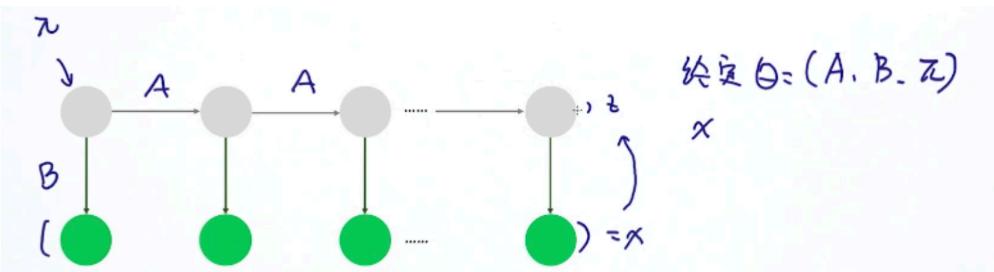
第一种，给定魔性的参数，找出最合适的 Z
第二种，估计魔性的参数 theta

1. complete case，x 和 z 都已知
2. incomplete case，x 已知，z 未知。（需要用到 EM）（3.3）





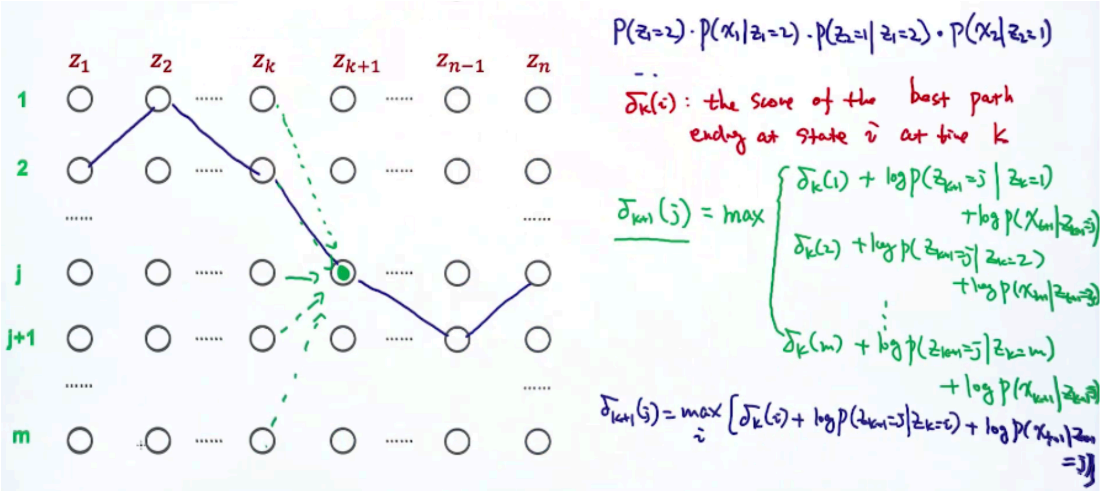
1.4 find best Z

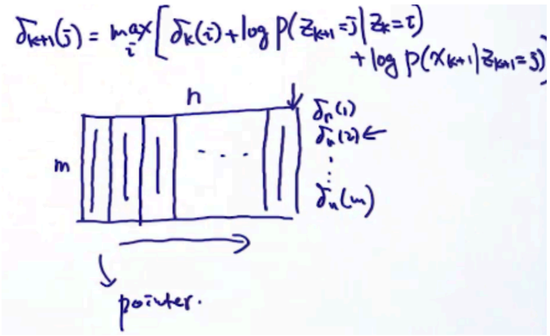


1）遍历算法
把所有可能的 z 序列罗列出来，可以很容易的求得每个概率。
选出最大概率的。

2）viterbi 动态规划

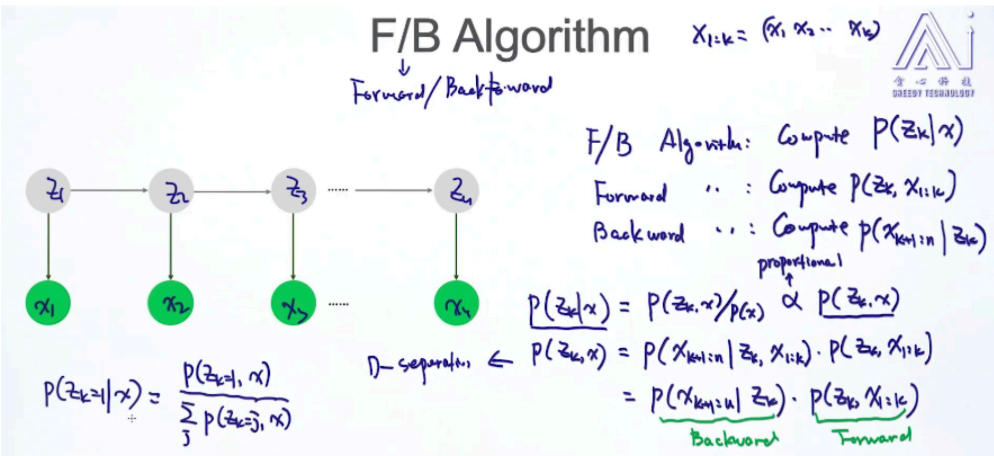
HMM 中有一个条件非常重要---隐式变量 z 只与前后两个变量相关联---这样可以大大减轻计算量。





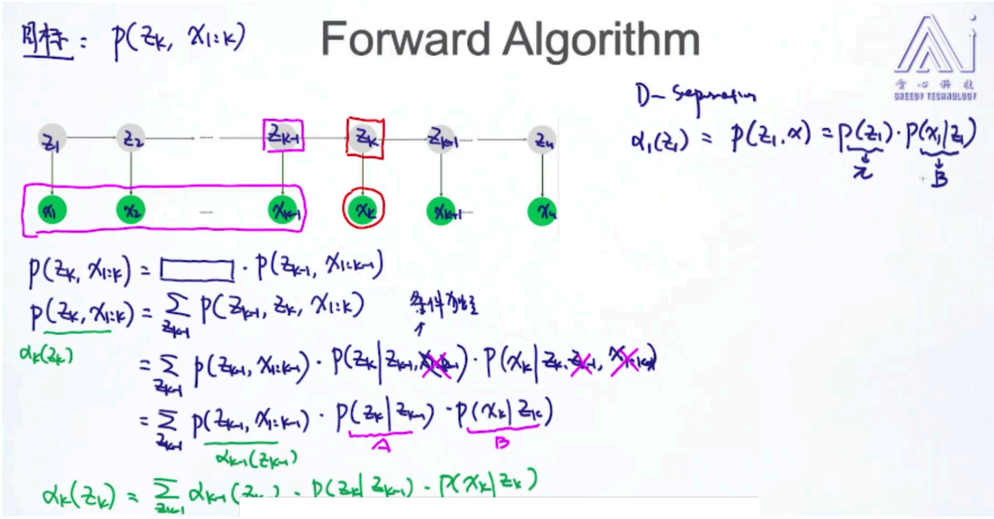
2. HMM 中的 F/B 算法

F/B：forward/backward

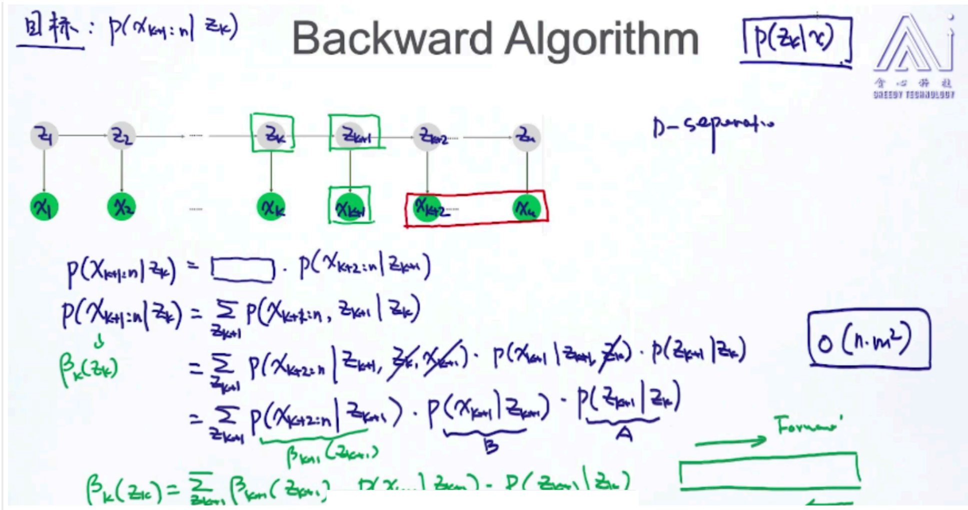


用处：求出模型参数；change detection

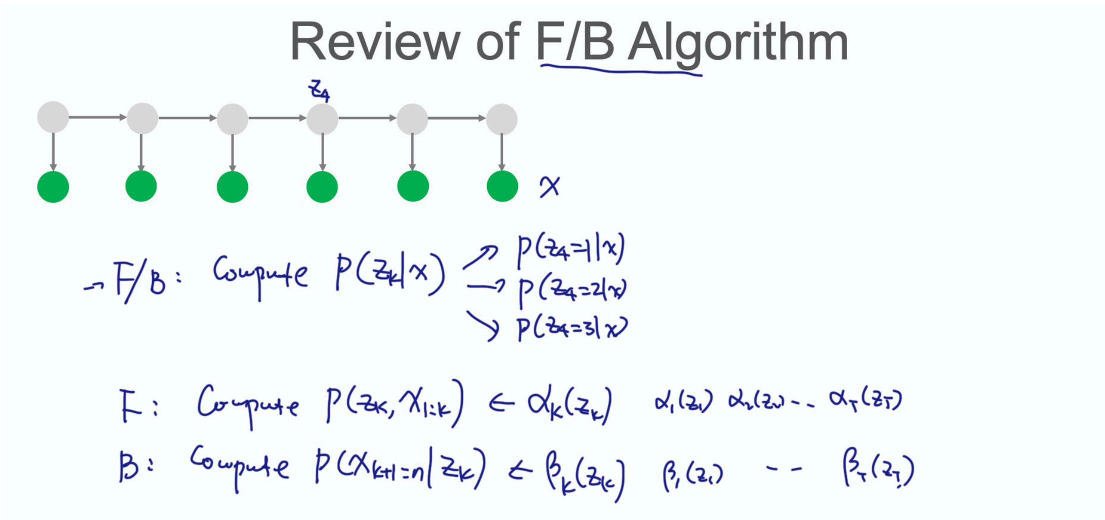
2.1 forward



2.2 backward



2.3 review

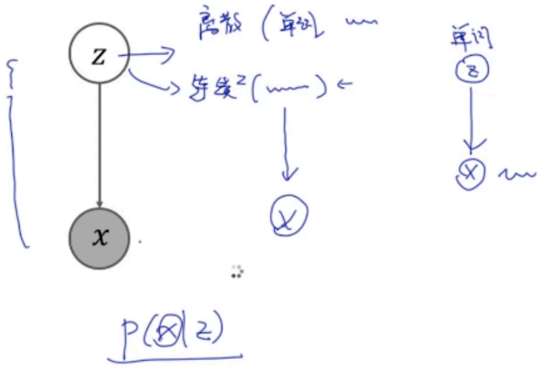


3. Latent Variable Model

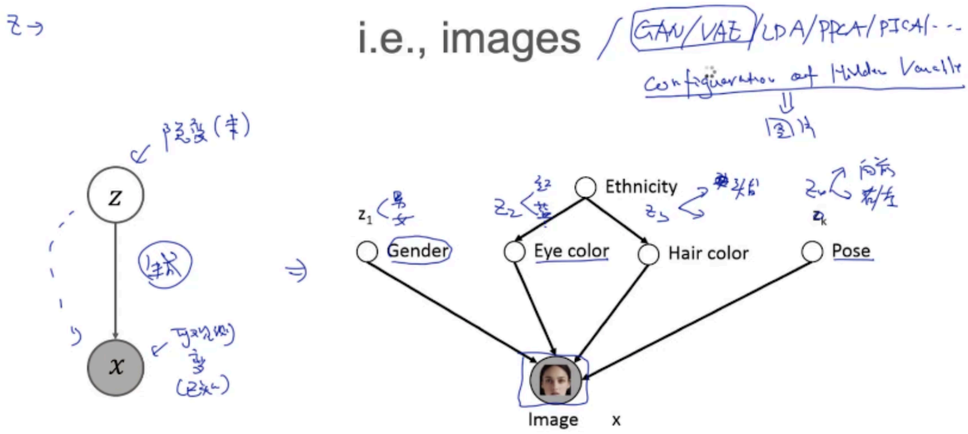
3.1 data representation

更倾向于用低维空间表示信息。---》去掉冗余信息。

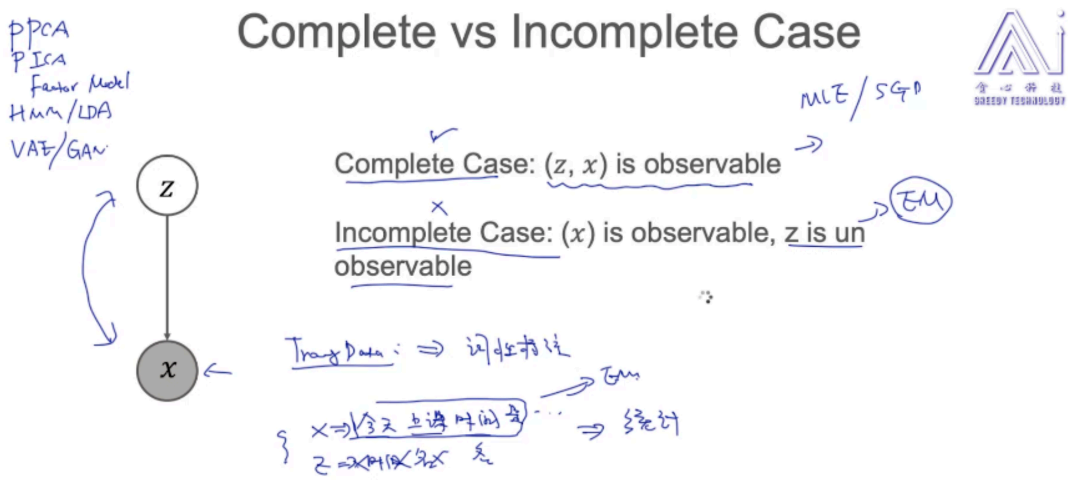
3.2 latent variable models



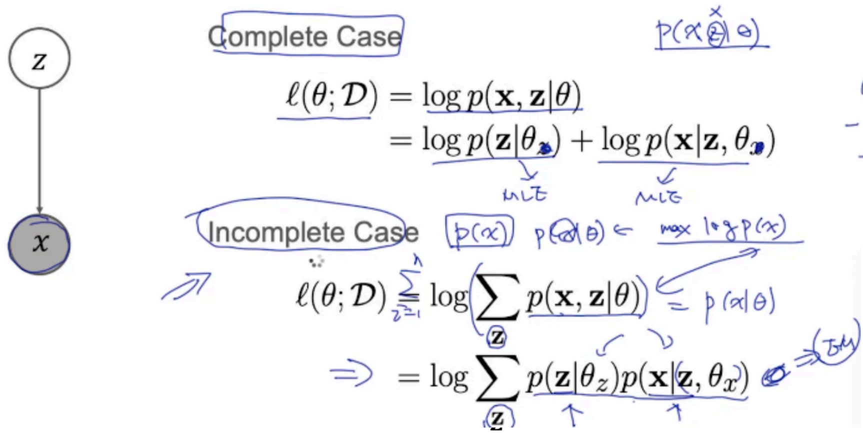
给定隐变量，生成新的数据（生成模型）



3.3 Complete vs Incomplete Case

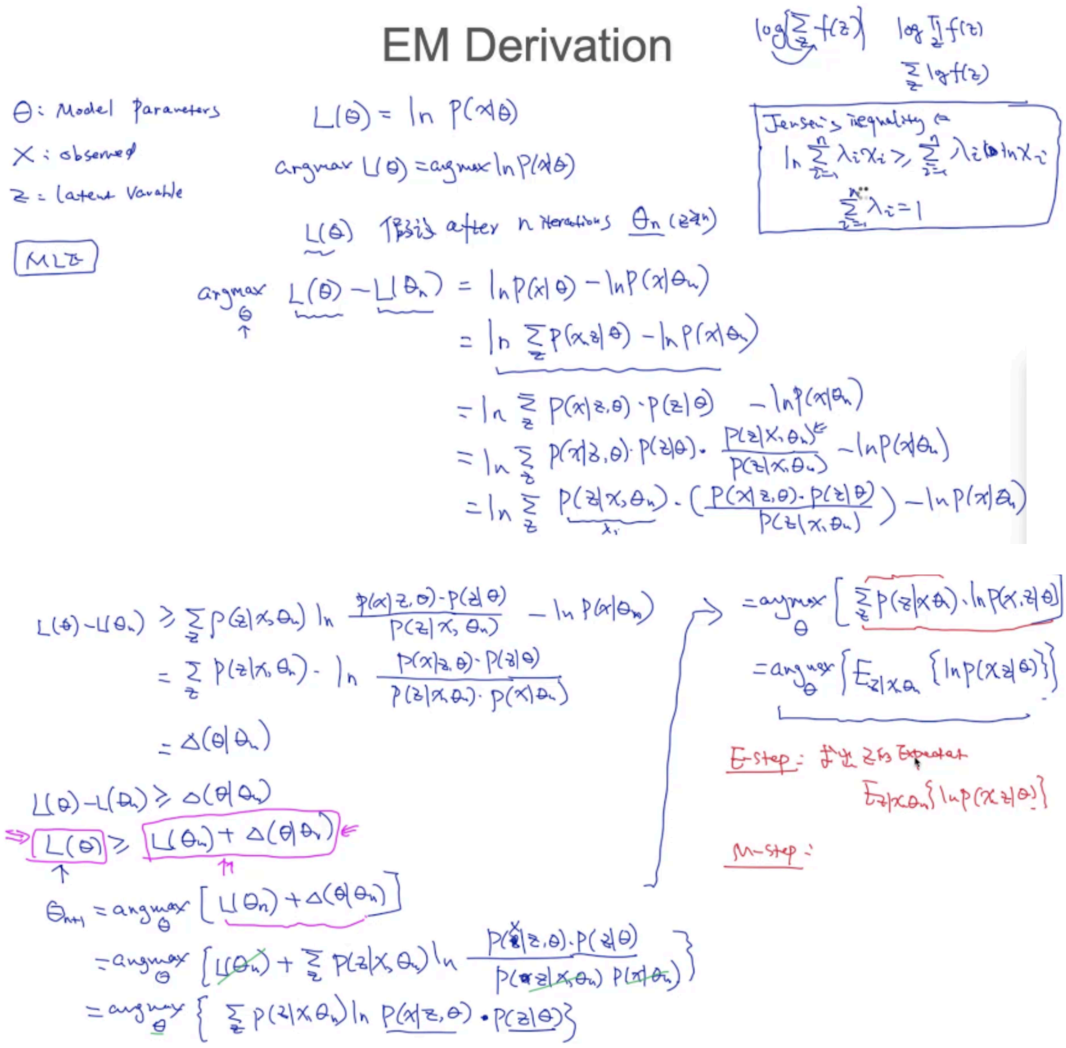


3.4 MLE

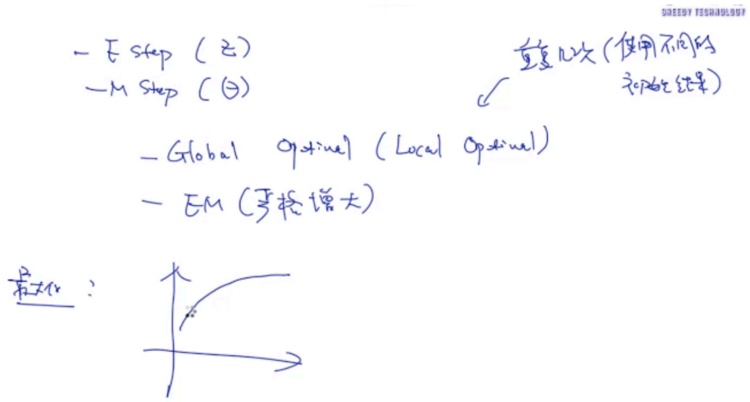


4. EM 算法

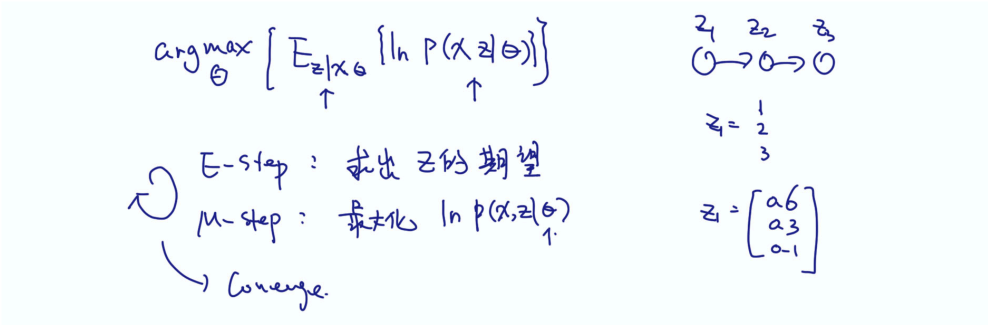
4.1 derivation

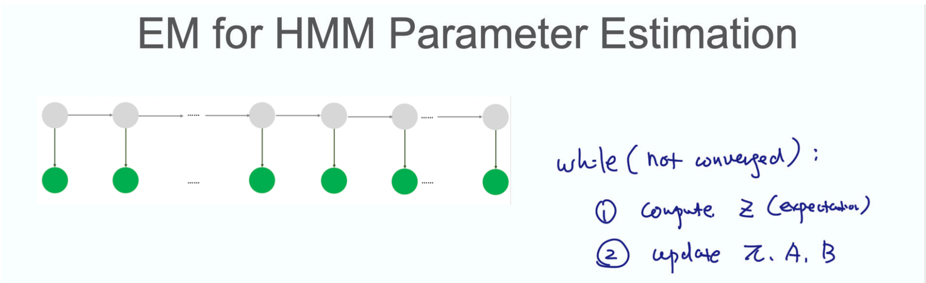


4.2 remarks



不能保证全局最优，每次迭代都能找到比当前更优秀的解

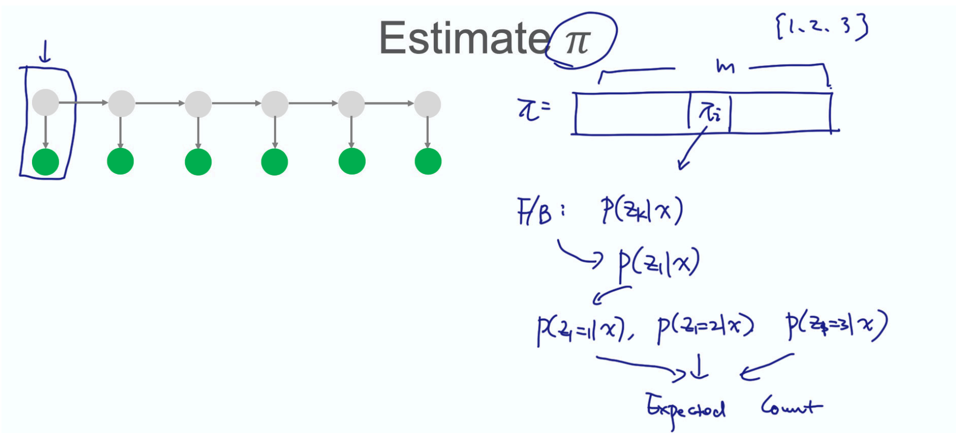


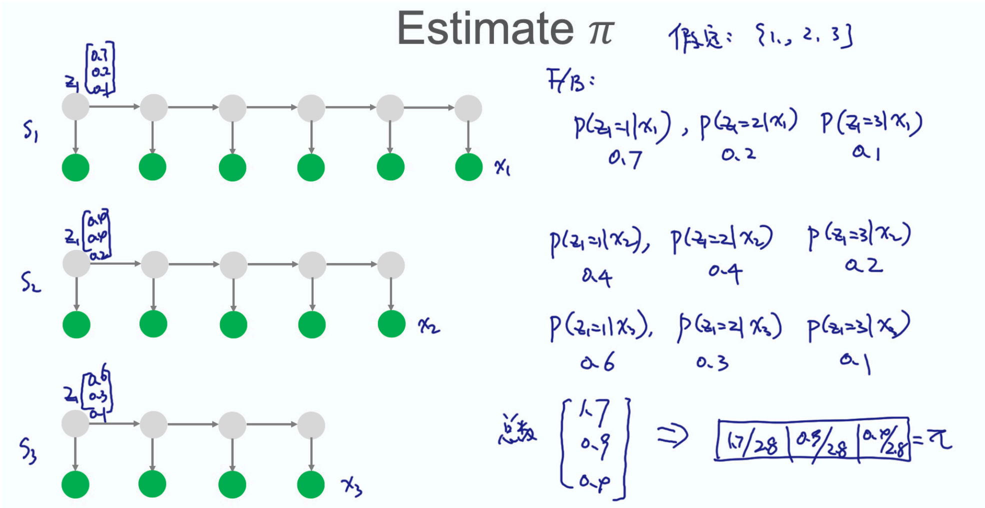


5. HMM 参数估计

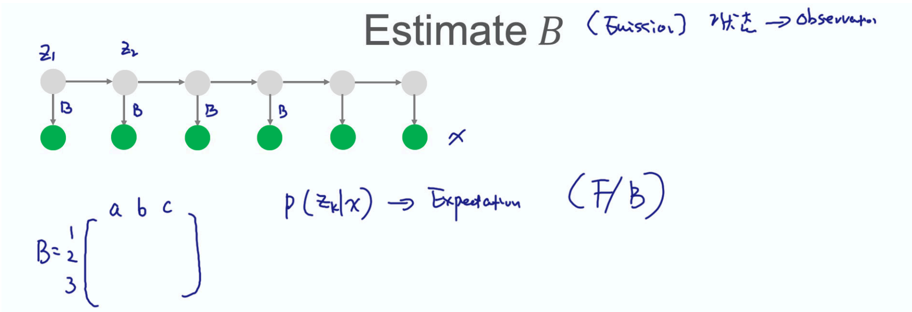
5.1 估计 pi

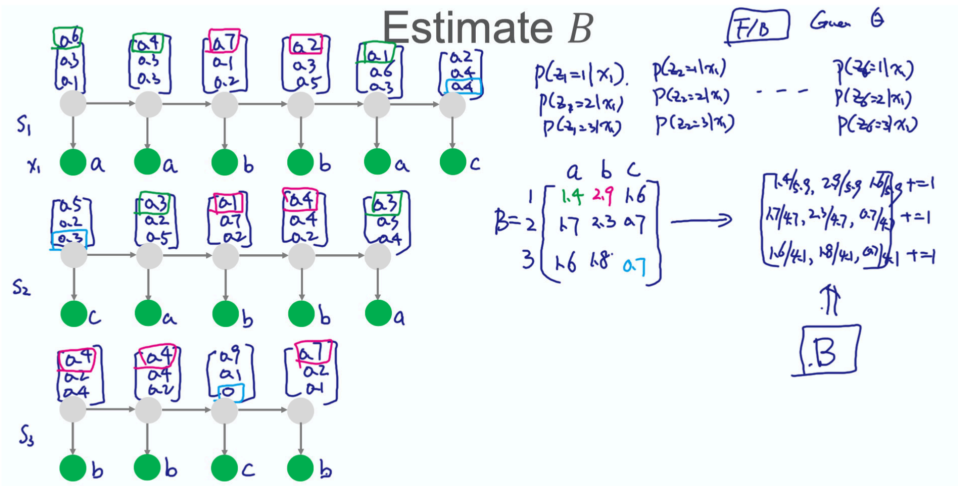
某一个时刻出现某个状态的概率




Estimate π
假定: {1, 2, 3}
F/B:
S1
S2
S3
x1
x2
x3

5.2 估计 B


Estimate B
(Emission) 状态 → observation
p(zk|x) → Expectation (F/B)
a b c
B= 1 2 3


Estimate B
F/B Given θ
S1
S2
S3
B

5.3 估计 A

Estimate A: Review of Language Model

$\boxed{\text{Bi-gram}}$

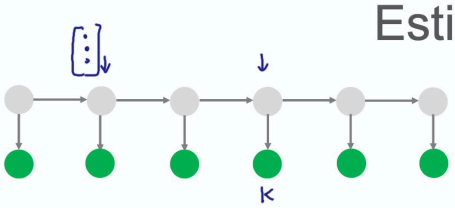
$w_i \to w_j$

$$P = p(w_j|w_i) = \frac{C(w_i, w_j)}{C(w_i)}$$

$$p(我们|今天) = \frac{3}{8}$$

今天□ 今天□
今天我们 今天□
今天□ 今天□
今天我们 今天我们

Estimate A

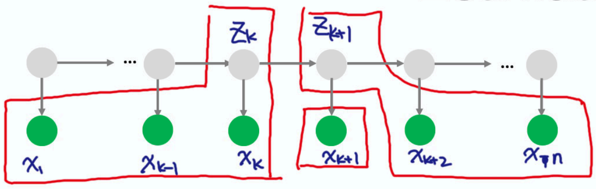


$$\underbrace{p(z_k=i, z_{k+1}=j|x)}_{\uparrow \text{ Forward + Backward}} = ?$$

$$A = \begin{array}{cc} & j \\ i & \begin{bmatrix} & A_{ij} & \\ & & \\ & & \end{bmatrix} \end{array}_{m \times m}$$

$$p(z_{k+1}=j|z_k=i) = \frac{p(z_{k+1}=j, z_k=i)}{p(z_k=i) \leftarrow F/B} = \frac{C(z_k=i, z_{k+1}=j)}{C(z_k=i) \leftarrow}$$

Estimate A $\boxed{p(z_k=i, z_{k+1}=j|x) = }$?



$$1 = \left\{ \begin{array}{l} p(z_k=1, z_{k+1}=1|x) \propto \beta_k(1,1) \\ p(z_k=1, z_{k+1}=2|x) \propto \beta_k(1,2) \\ p(z_k=1, z_{k+1}=3|x) \propto \beta_k(1,3) \\ p(z_k=2, z_{k+1}=1|x) \propto \beta_k(2,1) \\ \vdots \\ p(z_k=3, z_{k+1}=3|x) \propto \beta_k(3,3) \end{array} \right.$$

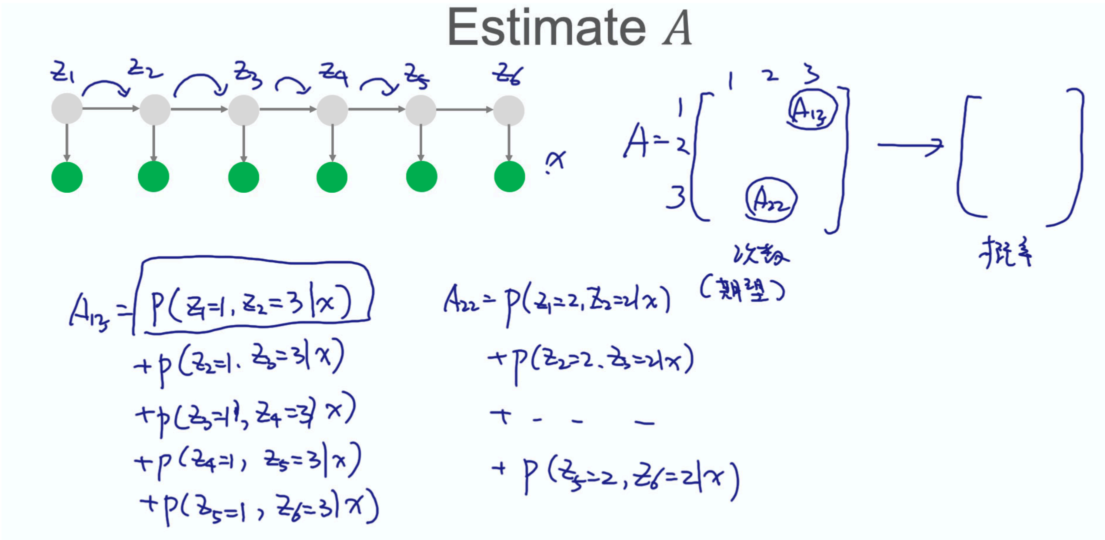
$$p(z_k=1, z_{k+1}=1|x) = \frac{\beta_k(1,1)}{\beta_k(1,1) + \beta_k(1,2) + \dots + \beta_k(3,3)} \text{ Normalize}$$

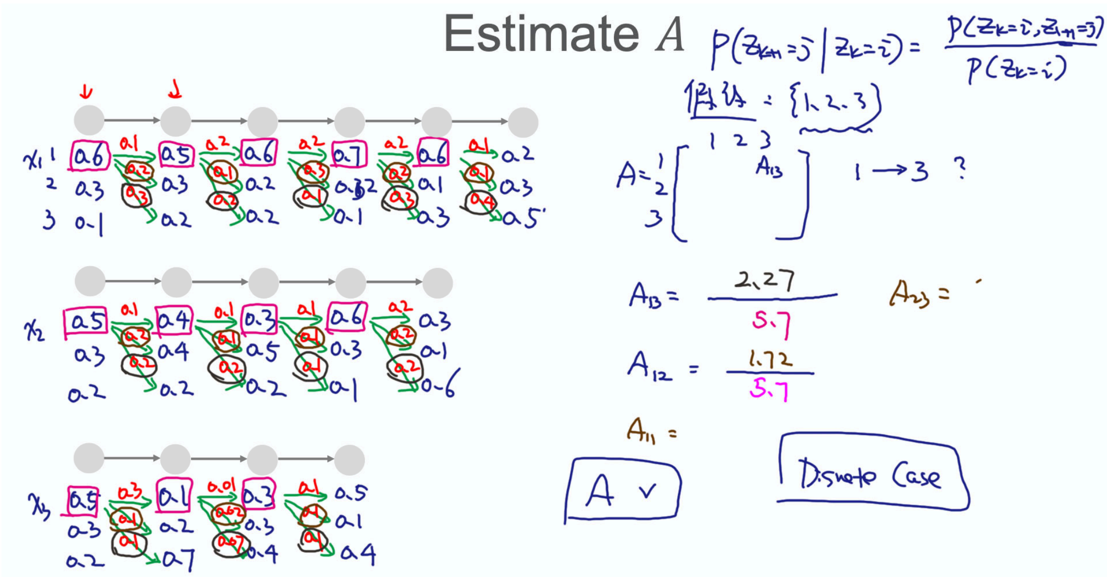
$$p(z_k=i, z_{k+1}=j|x) \propto p(z_k=i, z_{k+1}=j, x)$$

$$= p(z_k=i, z_{k+1}=j, x_{1:k}, x_{k+1}, x_{k+2:n})$$

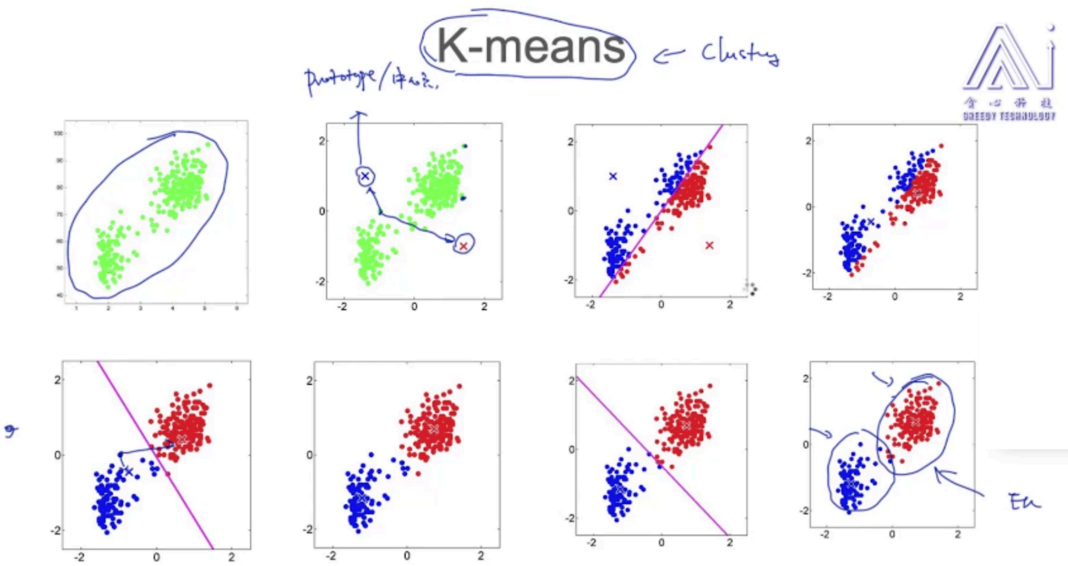
$$= \underbrace{p(z_k, x_{1:k})}_{\text{Forward}} \cdot \underbrace{p(x_{k+2:n}|z_{k+1})}_{\text{Backward}} \cdot \underbrace{p(z_{k+1}|z_k)}_{\text{Transition}} \cdot \underbrace{p(x_{k+1}|z_{k+1})}_{\text{Emission}}$$

$$= \underbrace{\alpha_k(z_k) \cdot \beta_{k+1}(z_{k+1}) \cdot \underset{A}{p(z_{k+1}|z_k)} \cdot \underset{B}{p(x_{k+1}|z_{k+1})}}_{\beta_k(i,j)}$$


Estimate A
$A_{13} = P(z_1=1, z_2=3 | x) + p(z_2=1, z_3=3|x) + p(z_3=1, z_4=3|x) + p(z_4=1, z_5=3|x) + p(z_5=1, z_6=3|x)$
$A_{22} = p(z_1=2, z_2=2|x) + p(z_2=2, z_3=2|x) + \dots + p(z_5=2, z_6=2|x)$
次数
(期望)
概率


Estimate A
$p(z_{k+1}=j | z_k=i) = \frac{p(z_k=i, z_{k+1}=j)}{p(z_k=i)}$
$1 \to 3$?
$A_{13} = \frac{2.27}{5.7}$
$A_{12} = \frac{1.72}{5.7}$
$A_{11} =$
$A_{23} =$
Discrete Case

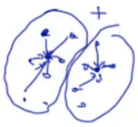
6. K-means


K-means
Clustering
Prototype/中心点

K-means Cost Function

考虑 Objective function:

一个点属于其中某一个 cluster



① 变量: 一个点属于哪一个 cluster?

② 中心点?

K clusters: μ_k → 中心点的坐标

$$r_{nk} = \begin{cases} 1 & \text{if } x_n \text{ belongs to } k\text{th clusters} \\ 0 & \text{otherwise} \end{cases}$$

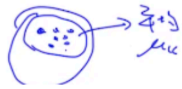
$$x_n \Rightarrow r_{n1}, r_{n2}, r_{n3}, \ldots r_{nk}$$
$$(1, 0, 0, \ldots 0)$$

minimize $\boxed{J = \sum_{n=1}^{N} \sum_{k=1}^{K} \underbrace{r_{nk}}_{\text{Latent}(z)} \| x_n - \underbrace{\mu_k}_{\text{参数}} \|^2} \Rightarrow$ EM

E-step: r_{nk} $\forall n$ (给 μ_k)

minimize $\sum_{k=1}^{K} r_{nk} \| x_n - \mu_k \|^2$

M-step: 估计 μ_k (已知 r_{nk}):



LDA. variational EM