Maximum likelihood of Gaussian.

$X \sim N(\mu, \delta)$

$$P(x|\theta) = \frac{1}{\sqrt{2\pi}\delta} \exp\left[-\frac{1}{2\delta^2}(x-\mu)^2\right]$$

for $x_1 \cdots x_N$

$$P(x_1 \cdots x_N|\theta) = \prod_{i=1}^{N} \frac{1}{(2\pi\delta^2)^{1/2}} \exp\left(-\frac{1}{2\delta^2}(x_i-\mu)^2\right]$$

$$= \frac{1}{(2\pi\delta^2)^{N/2}} \exp\left(-\sum_{i=1}^{N} \frac{1}{2\delta^2}(x_i-\mu)^2\right]$$

Given log liklihood.

$$\ell(\theta, x) = \log(P(x|\theta))$$

$$= \log\left[\frac{1}{(2\pi\delta^2)^{N/2}} \exp\left(-\sum_{i=1}^{N} \frac{1}{2\delta^2}(x_i-\mu)^2\right)\right]$$

$$= -\frac{N}{2}(\log(2\pi) + \log\delta^2) - \sum_{i=1}^{N} \frac{1}{2\delta^2}(x_i-\mu)^2$$

$$= -\frac{N}{2}\log 2\pi - \frac{N}{2}\log\delta^2 - \sum_{i=1}^{N} \frac{1}{2\delta^2}(x_i-\mu)^2$$

$$\frac{\partial \ell}{\partial \mu} = 0 \Rightarrow \sum_{i=1}^{N} \frac{1}{\delta^2}(x_i-\mu) = 0$$

$$\sum_{i=1}^{N} (x_i-\mu) = 0$$

$$\hat{\mu}_{ML} = \frac{\sum_{i=1}^{N} x_i}{N}$$

$$\frac{\partial \ell}{\partial \delta^2} = 0 \Rightarrow \left(\frac{1}{x}\right)' = -\frac{1}{x^2} \quad \begin{cases} \ln x = \frac{x}{1} \\ \log x = \frac{1}{x} \cdot \log e \end{cases}$$

$$-\frac{N}{2\delta^2} + \frac{1}{2} \cdot \frac{1}{\delta^4} \sum_{i=1}^{N} (x_i-\mu)^2 = 0$$

$$\Rightarrow \frac{1}{2\delta^4} \sum_{i=1}^{N} (x_i-\mu)^2 = \frac{N}{2\delta^2}$$

$$\frac{1}{\delta^2} \sum_{i=1}^{N} (x_i-\mu)^2 = N$$

$$\delta^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i-\mu)^2$$

$$\delta^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i-\mu_{ML})^2.$$

$$N(x|\mu,\Sigma) = \frac{1}{(2\pi)^{D/2}\,|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)$$

multivariate Gaussian.

$\theta = (\theta_1 \cdots \theta_n)$

$X\theta = Y$

$\rightarrow n$, $\downarrow N$

$$\begin{pmatrix} x_1 & \cdots & x_n \\ x & \cdots & x \\ x_{N1} & \cdots & x_{Nn} \end{pmatrix} \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}$$

$X_1\,(x_1 \cdots x_n) \rightarrow y_1$

$X_2\,(x_{12} \cdots x_2) \rightarrow y_2$

$\vdots$

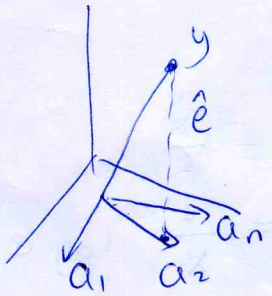
$X_N\,(x_1 \cdots x_{1N}) \rightarrow y_N$

$$E(\theta) = (y - A\theta)^T (y - A\theta)$$
$$= (y^T - \theta^T A^T)(y - A\theta)$$
$$= y^T y + \theta^T A^T A\theta - y^T A\theta - \theta^T A^T y$$
$$= y^T y + \theta^T A^T A\theta - 2\,y^T A\theta \quad (\text{scalar})$$
$$-2\,\theta^T A^T y$$

$$\frac{\partial E}{\partial \theta} = 2A^T A\theta - 2y^T A = 0$$

$$A^T A\,\theta = y^T A \Rightarrow \hat{\theta} = (A^T A)^{-1} y^T A$$

$$\frac{\partial E}{\partial \theta} = 0 \Rightarrow \hat{\theta} = (A^T A)^T A^T y$$



$$\vec{a}_1\theta_1 + \vec{a}_2\theta_2 \cdots + \vec{a}_n\theta_n = y$$

$$A^T(y - A\hat{\theta}) = 0$$
$$A^T y - A^T A\hat{\theta} = 0$$
$$\hat{\theta} = (A^T A)^{-1} A^T y$$

$$\begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} - \begin{pmatrix} A\theta_1 \\ \vdots \\ A\theta_N \end{pmatrix}$$

Stochastic process

A collection of random variable. $X_1 \cdots X_n$

Chapman-Kolmogorov Equation.

Suppose that $\{f_i\}$ is an indexed collection of random variables. that is, a stochastic process

Let $P_{i_1,\cdots i_n}(f_1 \cdots f_n)$ be the joint probability density function of the values of the random variables $f_1 \cdots$ to f_n. Then, the Chapman-Kolmogorov equation is:

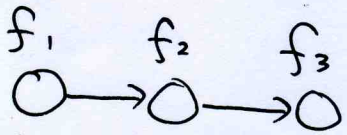
$$P_{i_1 \cdots i_{n-1}}(f_1 \cdots f_{n-1}) = \int_{-\infty}^{\infty} P_{i_1 \cdots i_n}(f_1 \cdots f_n)\, df_n$$

Given a Markov chain:

$$P_{i_1 \cdots i_n}(f_1 \cdots f_n) = P_{i_1}(f_1)\, P_{i_2 i_1}(f_2|f_1) \cdots P_{i_n, i_{n-1}}(f_n|f_{n-1})$$

$P_{ij}(f_i, f_j)$ is called the transition probability.

$$P_{i_3 i_1}(f_3|f_1) = \int_{-\infty}^{\infty} P_{i_3 i_2}(f_3|f_2)\, P_{i_2 i_1}(f_2|f_1)\, df_2$$



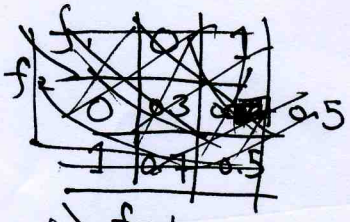
$P(f_3=1 \mid f_1=0) = ?$

$P(f_3=1|f_2=0)\,P(f_2=0|f_1=0) + P(f_3=1|f_2=1)\,P(f_2=1|f_1=0)$

$= 0.3 \times 0.9 + 0.7 \times 0.2$

$= 0.27 + 0.14 = 0.41$

$\Rightarrow$ ~~P_{03}~~ =



f_1 \ f_2	0	1
0	0.3	0.7
1	0.4	0.6

f_2 \ f_3	0	1
0	0.1	0.9
1	0.8	0.2

Example of Markov chain.

$$P(X_n | X_{n-1} \cdots X_1) = P(X_n | X_{n-1})$$

$$P_{ij}(t) \triangleq P(X_{t+1} = j \mid X_t = i)$$

is the transition probability from state i to state j at time t. If the transition probability does not depend on time.

$$P_{ij}(t) \rightarrow P_{ij}$$

we say that the Markov chain is time-homogenous.

From DNA to transitioned matrix

2-gram language model.

{ ACATCCTTTGGCGTC 4×4

. I am a student of Chinese university ..... |V|×|V|

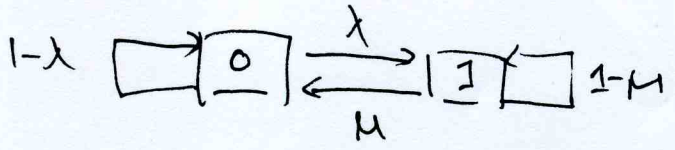
PATH probability:

the conditional probability of a path, conditional on the first value, is the product of the transitional probabilities between successive states of the path.

$$P(X_1, X_2 \cdots X_t) = (x_1 x_2 \cdots x_t)) =$$
$$P(X_1 = x_1) P_{x_1 x_2} P_{x_2 x_3} \cdots P_{x_{t-1} x_t}$$

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = P(X_1 = x_1) P_{x_1 x_2} P_{x_2} P_{x_3}$$
$$= P_{x_1} \cdot P_{x_1} P_{x_2} P_{x_2 x_3}$$

$$P(X_3 = x_3 \mid X_1 = x_1) = \sum_{x_2} P_{x_2} P_{x_1 x_2} P_{x_2 x_3}$$

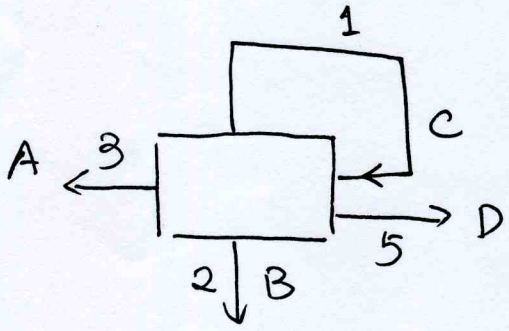


$$A = \begin{pmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{pmatrix} = \begin{pmatrix} 1-\lambda & \lambda \\ M & 1-M \end{pmatrix}.$$

$P(X_1 = 0) = 0.5$. $P(0,0,0,1,1,0) = ?$

$$= 0.5 P_{00} P_{00} P_{01} P_{11} P_{10} = 0.5(1-\lambda)(1-\lambda)\lambda(1-M)M$$

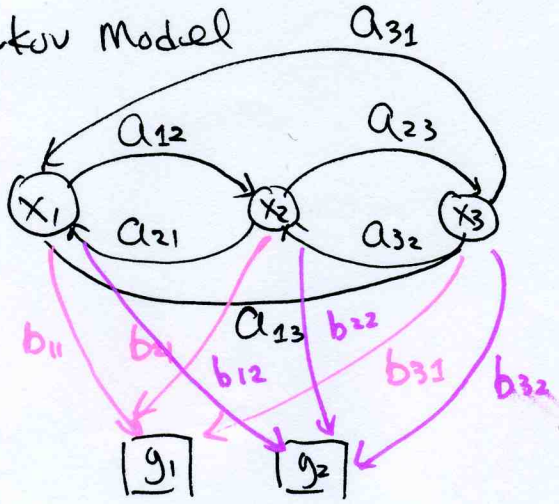
Similarly, we can calculate the probability of a sequence of DNA sequence GGCTTTCAAGCT ... with $(1/4, 1/4, 1/4, 1/4)$



$$AVE = \frac{1}{4}\times 3 + \frac{1}{4}\times 2 + \frac{1}{4}\times 5 + \frac{1}{4}\times (AVE+1)$$

$$\Rightarrow AVE = ??$$

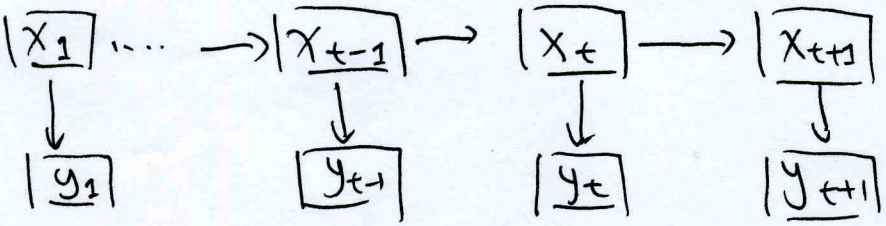
Hidden markov model



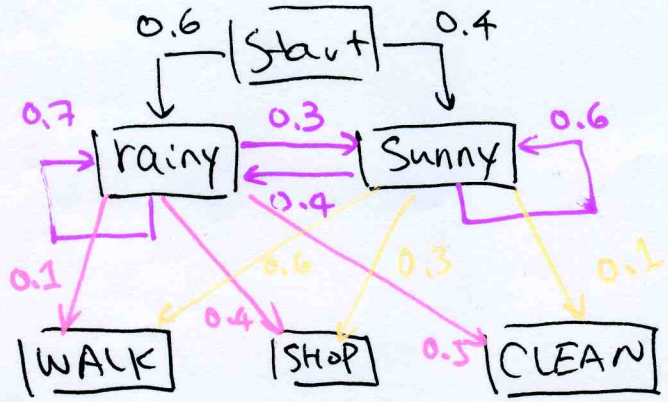
$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

$$B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{pmatrix}$$

Given $(y_t\ y_{t-1} \cdots y_1)$ how to estimate A and B.



Most likely sequence can be



Viterbi Algorithm

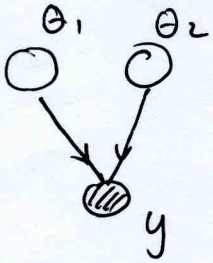
given a HMM with state X, initial probability π_i of being in state i and transition probabilities a_{ij} from state i to state j. Observe the outputs $y_0 \cdots y_T$ what is the most likly sequence of states $(x_0 \cdots x_T)$?

$$\begin{cases} V_{0k} = P(y_0 \mid k)\,\pi_k \\ V_{tk} = P(y_t \mid k)\cdot \max_{x \in X}(a_{x,k}\, V_{t-1,x}) \end{cases} \qquad \begin{cases} \pi_0 = \{0.6, 0.4\} \\ X = \{\text{rainy, sunny}\} \end{cases}$$

V_{tk} is the probability of the most probable state sequence responsible the first $t+1$ observations that has the k as the final state. Let $Ptr(k,t)$ be the function that returns the value of x used to compute $V_{t,k}$ if $t>0$, or k if $t=0$. Then:

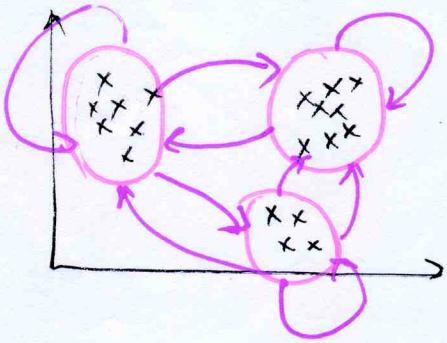
$$\begin{cases} X_T = \arg\max_{x \in X}(V_{T,x}) \\ X_{t-1} = Ptr(X_t, t) \end{cases}$$

The similarity to mixture model. HMM is a dynamic mixture model.



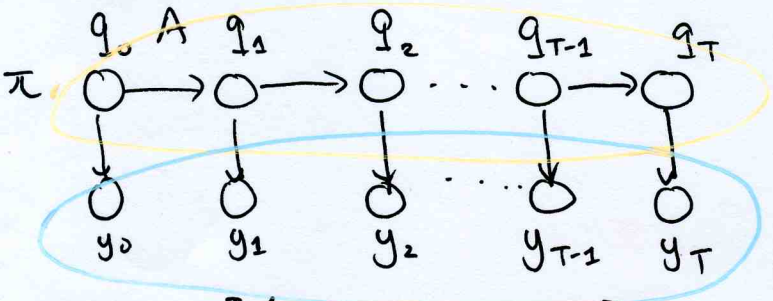
θ_1 - State 1
θ_2 - State 2.

{ emission probability is the direct probability distribution.



It has the state transition probability. And the data is generated in sequence.

New Notation of Jordan

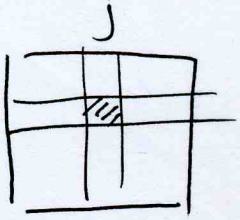


q - state
y - output
A - homogeneous transition matrix.

$$P(q,y) = P(q_0)\prod_{t=0}^{T-1} P(q_{t+1} \mid q_t)\prod_{t=0}^{T} P(y_t \mid q_t)$$

$$a_{q_t, q_{t+1}} \triangleq \prod_{i,j=1}^{M} [a_{ij}]^{q_t^i q_{t+1}^j} \Rightarrow$$

$$P(q,y) = \pi_{q_0} \prod_{t=0}^{T-1} a_{q_t q_{t+1}} \prod_{t=0}^{T} P(y_t \mid q_t)$$



$$P(y) = \boxed{\sum_q P(q,y)}^{1} = \sum_{q_0}\sum_{q_1}\cdots\sum_{q_T} \pi_{q_0} \prod_{t=0}^{T} a_{q_t q_{t+1}} \prod_{t=0}^{T} P(y_t|q_t)$$

$$P(q_t|y) = \frac{\boxed{P(y|q_t)P(q_t)}^{2}}{P(y)}$$

$$= \frac{P(y_0\cdots y_t|q_t)\,P(q_t)\,P(y_{t+1}\cdots y_T|q_t)}{P(y)}$$

$$= \frac{\alpha(q_t)\beta(q_t)}{P(y)}$$

$$\boxed{\alpha(q_t) \triangleq P(y_0\cdots y_t, q_t)}^{3}$$
$$\boxed{\beta(q_t) \triangleq P(y_{t+1}\cdots y_T|q_t)}$$

$\alpha(q_t)$ is a joint probability while $\beta(q_t)$ is a conditional probability. 1, 2 and 3:

$$\boxed{P(y) = \sum_{q_t} \alpha(q_t)\beta(q_t)}^{4} \qquad P(y) = \sum_{q_t} P(y, q_t)$$

$$\gamma(q_t) \triangleq \frac{\alpha(q_t)\beta(q_t)}{P(y)}$$

A new notation of posterior probability.

$$\alpha(q_{t+1}) = P(y_0,\cdots y_{t+1}, q_{t+1})$$
$$= P(y_0,\cdots y_{t+1}|q_{t+1})\,P(q_{t+1})$$
$$= P(y_0,\cdots y_t|q_{t+1})\,P(y_{t+1}|q_{t+1})\,P(q_{t+1})$$
$$= P(y_0,\cdots y_t, q_{t+1})\,P(y_{t+1}|q_{t+1})$$
$$= \sum_{q_t} P(y_0\cdots y_t, q_t, q_{t+1})\,P(y_{t+1}|q_{t+1})$$
$$= \sum_{q_t} P(y_0,\cdots y_t, q_{t+1}|q_t)\,P(q_t)\,P(y_{t+1}|q_{t+1})$$
$$= \sum_{q_t} P(y_0,\cdots y_t|q_t)\,P(q_{t+1}|q_t)\,P(q_t)\,P(y_{t+1}|q_{t+1})$$
$$= \sum_{q_t} P(y_0,\cdots y_t, q_t)\,P(q_{t+1}|q_t)\,P(y_{t+1}|q_{t+1})$$
$$= \sum_{q_t} \alpha(q_t)\,a_{q_t q_{t+1}}\,P(y_{t+1}|q_{t+1}) \qquad b(q_{t+1})$$

$$\alpha(q_0) = P(y_0, q_0) = P(y_0|q_0)P(q_0) = P(y_0|q_0)\pi_{q_0}$$

$$\beta(q_t) = P(y_{t+1}\cdots y_T|q_t) = \sum_{q_{t+1}} P(y_{t+1}\cdots y_T, q_{t+1}|q_t)$$
$$= \sum_{q_{t+1}} P(y_{t+1}\cdots y_T|q_{t+1}, q_t)\,P(q_{t+1}|q_t)$$
$$= \sum_{q_{t+1}} P(y_{t+1}|q_{t+1})\,P(y_{t+2}\cdots y_T|q_{t+1})\,P(q_{t+1}|q_t)$$
$$= \sum_{q_{t+1}} \beta(q_{t+1})\,a_{q_t q_{t+1}}\,P(y_{t+1}|q_{t+1}) \qquad b(q_{t+1})$$