Expectation-Maximization :

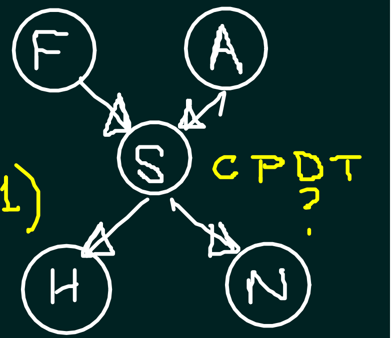
SUMMARY

① Bayes Net : Fully Known
Data : Fully Observed

$\hat{\theta} = \underset{\theta}{\text{argmax}} \left[P[\text{Data} | \theta] \right]$ MLE

$$\theta_{s|ij} = P(S=1 | F=i, A=j)$$

$$= \frac{\sum_k \delta(f_k = i, a_k = j, S_k = 1)}{\sum_k \delta(f_k = i, a_k = j)}$$



② Bayes Net : Known
Data : Partially Obs.

→ E-step : $\text{Exp}[Z | X, \theta]$

unobserved observed

$\hat{\theta}' = \underset{\theta'}{\text{argmax}} \, Q(\theta', \theta)$

→ M-step : $Q(\theta', \theta) = E_{Z|X,\theta} \log [P(X, Z | \theta')]$

current Mstep

↳ Bayesian Classifiers

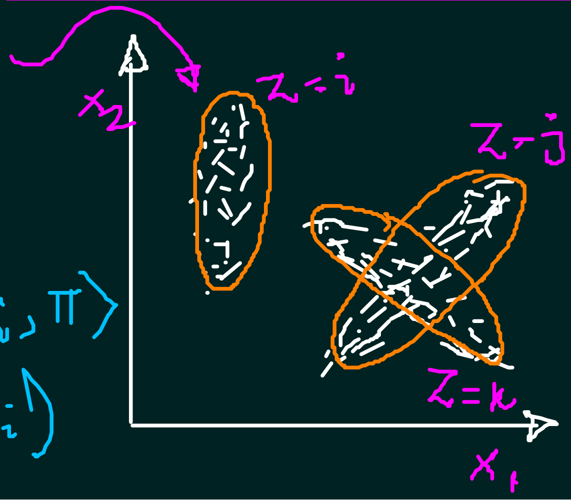
↳ Unsupervised Learning → clustering (where all O/P labels ??)

→ Gaussian Mixture Models :

Gaussian Dist

$$P(X_1 \ldots X_N) = \sum_i P(Z=i) P(X_1 \ldots X_N | Z=i)$$

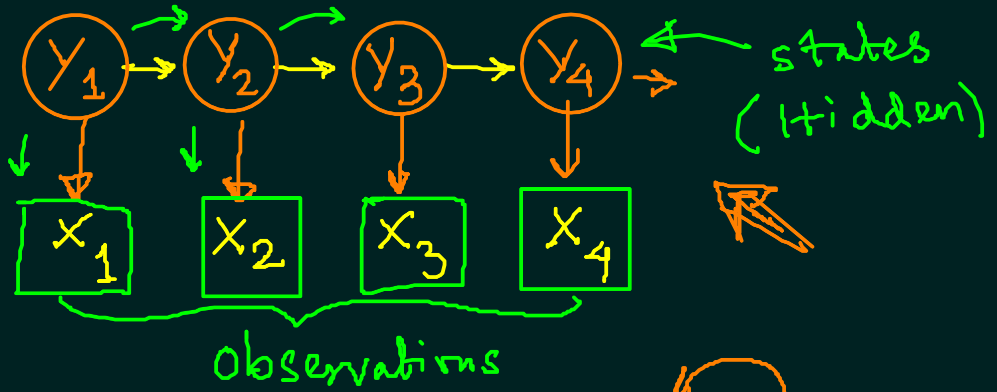
$\mathcal{N}(\mu_{ji}, \sigma)$



$\theta = \langle \mu_{ji}, \pi \rangle$

→ E-step : $P(Z(n) | X(n), \theta)$

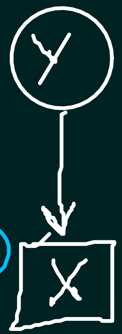
→ M-step : $\pi \leftarrow \frac{1}{N} \sum_n E[Z(n)]$ ... (similarly μ_{ji})



Hidden Markov Models

$Y = y_1$
$= y_2$

(i) $Y \sim$ multinomial Distribution



(ii) $X | Y \sim$ Gaussian (cont) Distribution

Multinomial (Discrete)

Markov Process: $P(Y_{t+1} | Y_t \overbrace{X_{t-1} \dots Y_1})$

$= P(Y_{t+1} | Y_t)$

(iii) $\boxed{\text{HMM}}$

- → Initial distribution: → $\boxed{P(Y_t)}$ ← Prior [multinomial]
- → Transition Dist: → $\boxed{P(Y_t | Y_{t-1})}$ ~ multinomial
- → Emission Dist: → $\boxed{P(X_t | Y_t)}$ → Gaussian (cont)
  - → Multinomial (Discrete)

Ex: $P(Y_{1\dots t} X_{1---t}) \Rightarrow$ JPDT

$$\underbrace{P(Y_{1\dots t} X_{1\dots t})} = \underbrace{P(Y_1)} \prod_{i=1}^{t-1} \underbrace{P(Y_{i+1} | Y_i)} \cdot \prod_{i=1}^{t} \underbrace{P(X_i | Y_i)}$$

$x_{1\dots} x_t \rightarrow Y_{t+K}$ Predict

$x_{1\dots} x_t \rightarrow Y_{K<t}$ Hindsight

$\boxed{\text{Ftr or Rec}}$ → Monitor

$X_{1\dots t}$

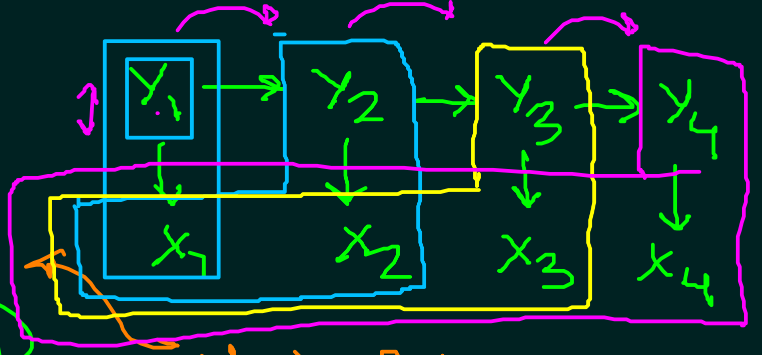
Y_{t+1} ?

Monitoring : M^t $\boxed{P(Y_t \mid X_{1\ldots t})} \propto P(X_t \mid Y_t, X_{1\ldots t-1})\, \underline{P(Y_t \mid X_{1\ldots t-1})}$

Bayes Theorem

$$= \underset{\text{Cond. Ind}}{P(X_t \mid Y_t)} \cdot \sum_{Y_{t-1}} \underbrace{P(Y_t\, Y_{t-1} \mid X_{1\ldots t-1})}_{\text{marginal}}$$

$$= P(X_t \mid Y_t) \cdot \sum_{Y_{t-1}} P(Y_t \mid Y_{t-1}, X_{1\ldots t-1}) \cdot \underbrace{P(Y_{t-1} \mid X_{1\ldots t-1})}$$



Chain Rule

$$= P(X_t \mid Y_t) \sum_{Y_{t-1}} P(Y_t \mid Y_{t-1})\; \underset{\text{Rec}}{M^{(t-1)}}$$

$\boxed{\text{Monitoring}}$

(forward - Algo)

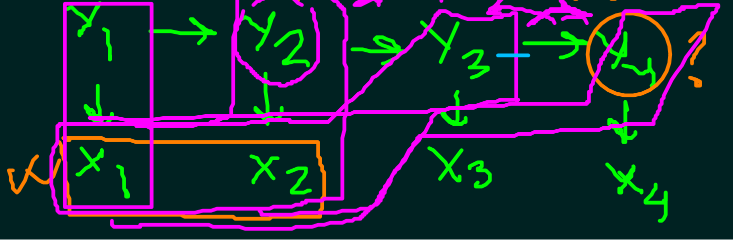
Prediction : P^k $P(Y_{t+k} \mid X_{1\ldots t})$ (where $k > 1$)

$$= \sum_{Y_{t+k-1}} P(Y_{t+k}\, Y_{t+k-1} \mid X_{1\ldots t}) \leftarrow \text{marginal}$$

Forward Algo

$$= \sum_{Y_{t+k-1}} P(Y_{t+k} \mid Y_{t+k-1}, X_{1\ldots t}) \cdot P(Y_{t+k-1} \mid X_{1\ldots t}) \leftarrow \text{chain Rule}$$

$$= \sum_{Y_{t+k-1}} P(Y_{t+k} \mid Y_{t+k-1}) \cdot P^{(k-1)}$$



Hindsight: $P(Y_k \mid X_{1---t})$ $\boxed{k<t}$

$= P(Y_k \cancel{X_{k+1\cdots t}} \mid X_{1\ldots k})$ ← Cond. Ind. ?

$= \underbrace{P(Y_k \mid X_{1--k})}_{M^k} \cdot P(X_{k+1\cdots t} \mid Y_k)$ H^k

Fwd →

← Chain Rule

$P(Y_k \mid X_{1--k}\,\overbrace{k+1---t}^{\circledast})$

$= P(Y_k \circledast \mid X_{1\ldots k})$

$\Rightarrow P(X_{k+1--t} \mid Y_k)$

$= \sum_{Y_{k+1}} P(Y_{k+1}, X_{k+1,\cdots t} \mid Y_k)$ ← marginals

$= \sum_{Y_{k+1}} P(Y_{k+1} \mid Y_k) \cdot P(X_{k+1} \mid Y_{k+1}) \cdot \overbrace{P(X_{k+2\ldots t} \mid Y_{k+1})}$ H^{k+1}

Cond. Ind.

Chain Rule

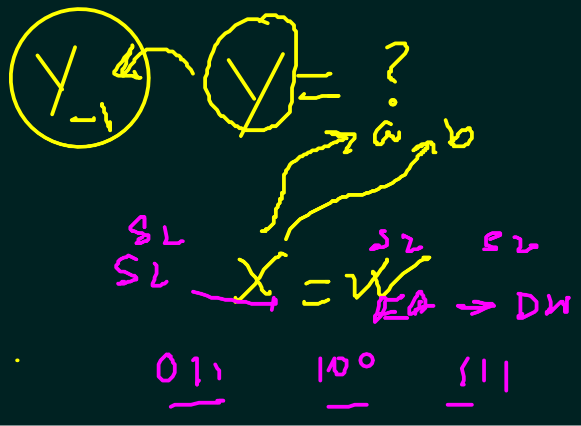
←

back

↳ Forward-Backward Algo.

→ $\max_{Y_{1---t}} P(\overset{abaac}{Y_{1\ldots t}} \mid X_{1---t})$

most likely exp. ←



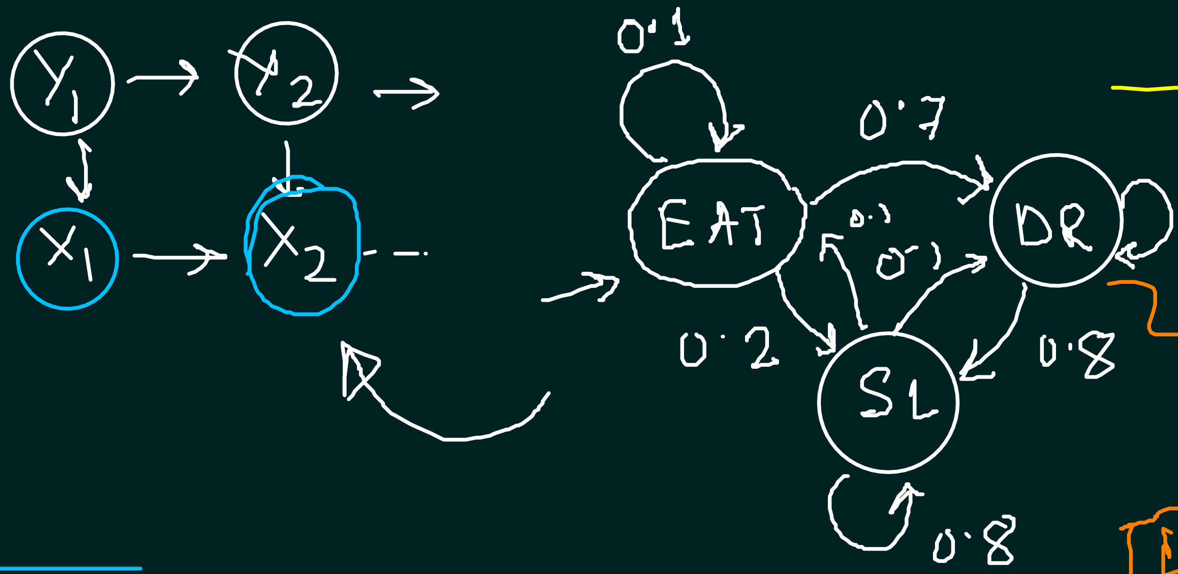
$$\max_{\underbrace{Y_{1--t}}} P(Y_{1--t} \mid X_{1---t}) = \max_{\underbrace{Y_t}} P(X_t \mid Y_t) . \max_{Y_{1..t-1}} P(Y_{1--t-1} \mid X_{1..t})$$

(Dynamic Prog.)

↳ Viterbi Algo

$Y_1 = a?$ $Y_1 = b?$ → $Y_2 = a?$ $Y_2 = b?$ → $Y_3 = a?$ $Y_3 = b?$

→ $P(X_1 = 0)$ → $P(X_2 = 1)$ →



$P(x_1 \mid DR) = ?$

$P(x_2 \mid DR) = ?$

HMM

→ Exp

→ Max

} EM

$P(y)$ ✓ →

$P(Y_t \mid X_{t-1})$ ✓ →

$P(X_t \mid Y_t)$ ✓ →

(Partiall Fited)

} MLE