Problem 1:

Wednesday, October 16, 2024 3:26 PM

$X_n \in \mathcal{X} = \{0,1\}$: Markov chain

$$A = \begin{bmatrix} \theta & 1-\theta \\ 1-\theta & \theta \end{bmatrix}$$

$X_0 = 1 \Rightarrow \pi(X_0 = 1) = 1$ and $\pi(X_0 = 0) = 0$

$Y_{1:8} = 11110?11$

① $P(X_1 = Y_1, \ldots, X_8 = Y_8 \mid \theta) = P(X_8 = Y_8 \mid X_7 = Y_7, \ldots, X_1 = Y_1, \theta) \cdot P(X_7 = Y_7, \ldots, X_1 = Y_1 \mid \theta)$

(By Markov Property) $= P(X_8 = Y_8 \mid X_7 = Y_7, \theta) \cdot P(X_7 = Y_7 \mid X_5 = Y_5, \ldots, X_1 = Y_1, \theta)\, P(X_5 = Y_5, \ldots, X_1 = Y_1 \mid \theta)$

(By Markov Property) $= P_\theta(X_8 = Y_8 \mid X_7 = Y_7) \cdot P_\theta(X_7 = Y_7 \mid X_5 = Y_5) \prod_{i=1}^{4} P_\theta(X_{i+1} = Y_{i+1} \mid X_i = Y_i) \cdot P_\theta(X_1 = Y_1 \mid X_0 = 1)$

$= \theta \cdot P_\theta(X_7 = 1 \mid X_5 = 0) \cdot (1-\theta) \cdot \theta^4$

$= \theta^5 (1-\theta) \cdot \sum_{X_6 \in \mathcal{X}} P(X_7 = 1 \mid X_6) \cdot P(X_6 \mid X_5 = 0)$

$= \theta^5 (1-\theta) \cdot [P(X_7 = 1 \mid X_6 = 0) \cdot P(X_6 = 0 \mid X_5 = 0) + P(X_7 = 1 \mid X_6 = 1) \cdot P(X_6 = 1 \mid X_5 = 0)]$

$= \theta^5 (1-\theta) \cdot [(1-\theta) \cdot \theta + \theta \cdot (1-\theta)]$

$$\boxed{\mathcal{L}(Y) = 2\theta^6 (1-\theta)^2}$$

② ML estimator of $Y \mid \theta$:

$$\frac{\partial \mathcal{L}}{\partial \theta} = 2[6\theta^5 (1-\theta)^2 - 2\theta^6 (1-\theta)] = 0$$

$\Rightarrow 3(1-\theta) - \theta = 0$

$\Rightarrow \boxed{\theta = 3/4}$

③ E-M algorithm for estimating θ using X as complete data.

Complete data log-likelihood $= l_{cd}(\theta, X) = \ln q_\theta(X)$

$= \ln p(X_1, \ldots, X_8)$

$= \ln p(X_8 \mid X_7) \cdot p(X_7 \mid X_6) \cdots p(X_1 \mid X_0)$

$= \sum_{i=0}^{7} \ln p(X_{i+1} \mid X_i)$

$= \sum_{i=0}^{7} \{X_i X_{i+1} + (1-X_i)(1-X_{i+1})\} \ln \theta + \{(1-X_i) X_{i+1} + X_i (1-X_{i+1})\} \ln(1-\theta)$

E-step:

$Q(\theta \mid \hat{\theta}^{(k)}) = E_{\hat{\theta}^{(k)}}[l_{cd}(\theta, X) \mid Y = y]$

$= E_{\hat{\theta}^{(k)}}[\sum_{i=0}^{7} \ln p(X_{i+1} \mid X_i) \mid Y = Y_{1:8}]$

$= \sum_{i=0}^{7} E_{\hat{\theta}^{(k)}}[\ln p(X_{i+1} \mid X_i) \mid Y = Y_{1:8}]$

$= \sum_{i=0}^{7} E_{\hat{\theta}^{(k)}}[\mathbb{1}_{\{X_i = X_{i+1}\}} \ln \theta + \mathbb{1}_{\{X_i \neq X_{i+1}\}} \cdot \ln(1-\theta) \mid Y_{1:8}]$

where $\mathbb{1}_{\{\cdot\}}$ is an indicator function

$= \ln \theta \cdot \sum_{i=0}^{7} E_{\hat{\theta}^{(k)}}[\mathbb{1}_{\{X_i = X_{i+1}\}} \mid Y_{1:8}] + \ln(1-\theta) \sum_{i=0}^{7} E_{\hat{\theta}^{(k)}}[\mathbb{1}_{\{X_i \neq X_{i+1}\}} \mid Y_{1:8}]$

$= \ln \theta \{4 + E_{\hat{\theta}^{(k)}}[\mathbb{1}_{\{X_6 = 0\}}] + E_{\hat{\theta}^{(k)}}[\mathbb{1}_{\{X_6 = 1\}}] + 1\} + \ln(1-\theta) \{1 + E_{\hat{\theta}^{(k)}}[\mathbb{1}_{\{X_6 \neq 0\}}] + E_{\hat{\theta}^{(k)}}[\mathbb{1}_{\{X_6 \neq 1\}}]\}$

$= 5 \ln \theta + \ln(1-\theta) + (\ln \theta + \ln(1-\theta)) \{E_{\hat{\theta}^{(k)}}[\mathbb{1}_{\{X_6 = 0\}}] + E_{\hat{\theta}^{(k)}}[\mathbb{1}_{\{X_6 = 1\}}]\}$

$= 5 \ln \theta + \ln(1-\theta) + (\ln \theta + \ln(1-\theta)) \{P[X_6 = 0 \mid \hat{\theta}^{(k)}] + P[X_6 = 1 \mid \hat{\theta}^{(k)}]\}$

$= 6 \ln \theta + 2 \ln(1-\theta)$

M-step:

$$\hat{\theta}^{(k+1)} = \operatorname*{argmax}_\theta Q(\theta \mid \hat{\theta}^{(k)})$$

$$\frac{\partial Q(\theta \mid \hat{\theta}^{(k)})}{\partial \theta} = 0$$

$\Rightarrow \frac{6}{\theta} - \frac{2}{1-\theta} = 0$

$\Rightarrow 6 - 6\theta - 2\theta = 0$

$\Rightarrow \theta = \frac{3}{4}$

$\therefore \boxed{\hat{\theta}^{(k+1)} = 3/4}$

$\therefore$ E-M estimate converges to the ML estimate at the first iteration, irrespective of the initialization.

Problem 2

Sunday, October 20, 2024 10:17 PM

$Y_i \in \mathbb{R}^2$, $1 \le i \le n$

$Y_i \sim p_\theta(y)$ where $p_\theta(y) = \pi_1 \phi_1(y) + (1-\pi_1)\cdot\phi_2(y)$

$$\phi_j(y) = \frac{1}{(2\pi|\Sigma_j|)^{1/2}} \exp\left(-\frac{1}{2}(y-\mu_j)^T \Sigma_j^{-1}(y-\mu_j)\right) ; j=1,2$$

$\theta = (\pi_1, \mu_1, \Sigma_1, \mu_2, \Sigma_2)$

Let the complete data space contain $Z = (Y, J)$, where J is a random variable denoting the Gaussian component from which Y is sampled.

$$\begin{aligned} l_{cd}(\theta, Z) &= \sum_{i=1}^n \ln q_\theta(z) \\ &= \sum_{i=1}^n \ln \kappa(y, j) \\ &= \sum_{i=1}^n \ln \pi(j_i) + \ln p(y_i|j_i) \\ &= \sum_{i=1}^n \left\{ \ln \pi(j_i) - \frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(|\Sigma_{j_i}|) - \frac{1}{2}(y_i-\mu_{j_i})^T \Sigma_{j_i}^{-1}(y_i-\mu_{j_i}) \right\} \end{aligned}$$

E-step:

$$\begin{aligned} Q(\theta|\hat\theta^{(k)}) &= \mathbb{E}_{\hat\theta^{(k)}}[l_{cd}(\theta, z)|Y=y] \\ &= \mathbb{E}_{\hat\theta^{(k)}}\Big[\text{Const} + \sum_{i=1}^n \Big\{ \ln \pi(j_i) - \frac{1}{2}\ln(|\Sigma_{j_i}|) - \frac{1}{2}(y_i-\mu_{j_i})^T \Sigma_{j_i}^{-1}(y_i-\mu_{j_i})\Big\} \Big| Y=y\Big] \\ &= \text{const} + \sum_{j_i} \sum_{i=1}^n \Big\{ \ln \pi(j_i) - \frac{1}{2}\ln(|\Sigma_{j_i}|) - \frac{1}{2}(y_i-\mu_{j_i})^T \Sigma_{j_i}^{-1}(y_i-\mu_{j_i})\Big\} \times p_{\hat\theta^{(k)}}(j_i|Y_i=y_i) \end{aligned}$$

Let $w_j^{(i)} = p_{\hat\theta^{(k)}}(j_i = j | Y_i = y_i)$

$$= \frac{p_{\hat\theta^{(k)}}(Y_i=y_i|j_i=j)\cdot p_{\hat\theta^{(k)}}(j_i=j)}{\sum_{u=1}^2 p_{\hat\theta^{(k)}}(Y_i=y_i|j_i=u)\cdot p_{\hat\theta^{(k)}}(j_i=u)}$$

$$p_{\hat\theta^{(k)}}(Y_i=y_i|j_i=u) = \frac{1}{(2\pi|\hat\Sigma_u^{(k)}|)^{1/2}} \exp\left(-\frac{1}{2}(y-\hat\mu_u^{(k)})^T \hat\Sigma_u^{-1(k)} (y-\hat\mu_u^{(k)})\right)$$

$p_{\hat\theta^{(k)}}(j_i=u) = \hat\pi_u^{(k)}$ where $\hat\pi_2^{(k)} = 1-\hat\pi_1^{(k)}$

M-step:

$$\hat\theta^{(k+1)} = \underset{\theta}{\text{argmax}}\ Q(\theta|\hat\theta^{(k)})$$

Maximizing w.r.t. μ_j first as it occurs in only 1 term:

$$\frac{\partial Q(\theta|\hat\theta^{(k)})}{\partial \mu_j} = 0$$

$$\Rightarrow \frac{\partial}{\partial \mu_j} \mathbb{E}_{\hat\theta^{(k)}}\left[\sum_{i=1}^n (y_i-\mu_{j_i})^T \Sigma_{j_i}^{-1}(y_i-\mu_{j_i}) \Big| Y=y\right] = 0$$

$$\Rightarrow \sum_{i=1}^n \frac{\partial}{\partial \mu_j} \mathbb{E}_{\hat\theta^{(k)}}\left[(y_i-\mu_{j_i})^T \Sigma_{j_i}^{-1}(y_i-\mu_{j_i}) | Y_i=y_i\right] = 0$$

$$\Rightarrow \sum_{i=1}^n \frac{\partial}{\partial \mu_j} \sum_{j_i} (y_i-\mu_{j_i})^T \Sigma_{j_i}^{-1}(y_i-\mu_{j_i})\cdot p_{\hat\theta^{(k)}}(j_i|Y_i=y_i) = 0$$

$$\Rightarrow \sum_{i=1}^n p_{\hat\theta^{(k)}}(j_i=j|Y=y)\cdot \Sigma_j^{-1}(y_i-\mu_j) = 0$$

$$\Rightarrow \hat\mu_j^{(k+1)} = \frac{\sum_{i=1}^n y_i \cdot p_{\hat\theta^{(k)}}(j_i=j|Y_i=y)}{\sum_{i=1}^n p_{\hat\theta^{(k)}}(j_i=j|Y=y)}$$

$$\therefore \boxed{\hat\mu_j^{(k+1)} = \frac{\sum_{i=1}^n y_i\, w_j^{(i)}}{\sum_{i=1}^n w_j^{(i)}}}$$

Next, maximizing wrt Σ_j:

$$\frac{\partial}{\partial \Sigma_j} Q(\theta|\hat\theta^{(k)}) = 0$$

$$\Rightarrow \frac{\partial}{\partial \Sigma_j} \mathbb{E}_{\hat\theta^{(k)}}\left[\sum_{i=1}^n \ln|\Sigma_{j_i}| + (y_i-\mu_{j_i})^T \Sigma_{j_i}^{-1}(y_i-\mu_{j_i}) | Y=y\right] = 0$$

$$\Rightarrow \sum_{i=1}^n \frac{\partial}{\partial \Sigma_j} \sum_{j_i} \ln|\Sigma_{j_i}| + (y_i-\mu_{j_i})^T \Sigma_{j_i}^{-1}(y_i-\mu_{j_i})\cdot p_{\hat\theta^{(k)}}(j_i|Y_i=y_i) = 0$$

$$\Rightarrow \sum_{i=1}^n w_j^{(i)}\left\{\frac{\partial}{\partial \Sigma_j}\ln|\Sigma_j| + \frac{\partial}{\partial \Sigma_j}(y_i-\mu_j)^T \Sigma_j^{-1}(y_i-\mu_j)\right\} = 0$$

$$\Rightarrow \sum_{i=1}^n w_j^{(i)}\left\{\Sigma_j^{-1} - \Sigma_j^{-1}(y_i-\mu_j)(y_i-\mu_j)^T\Sigma_j^{-1}\right\} = 0$$

$$\Rightarrow \boxed{\hat\Sigma_j^{(k+1)} = \frac{\sum_{i=1}^n w_j^{(i)}\cdot(y_i-\mu_j)\cdot(y_i-\mu_j)^T}{\sum_{i=1}^n w_j^{(i)}}}$$

Maximizing wrt π_1 next:

$$\frac{\partial}{\partial \pi_1} Q(\theta|\hat\theta^{(k)}) = 0$$

$$\Rightarrow \frac{\partial}{\partial \pi_1} \mathbb{E}_{\hat\theta^{(k)}}\left[\sum_{i=1}^n \ln \pi(j_i) | Y=y\right] = 0$$

$$\Rightarrow \sum_{i=1}^n \frac{\partial}{\partial \pi_1} \sum_{j_i} \ln \pi(j_i)\cdot p(j_i|Y_i=y_i) = 0$$

$$\Rightarrow \sum_{i=1}^n \frac{\partial}{\partial \pi_1}\left\{\ln \pi_1 \cdot p(j_i=1|Y_i=y_i) + \ln(1-\pi_1)\cdot p(j_i=2|Y_i=y_i)\right\} = 0$$

$$\Rightarrow \sum_{i=1}^n w_j^{(i)} \times \frac{1}{\pi_1} - \frac{1}{1-\pi_1}(1-w_j^{(i)}) = 0$$

$$\Rightarrow \frac{1}{\pi_1}\sum_{i=1}^n w_j^{(i)} = \frac{1}{1-\pi_1}\left(n - \sum_{i=1}^n w_j^{(i)}\right)$$

$$\Rightarrow (1-\pi_1)\sum_{i=1}^n w_j^{(i)} = \pi_1\left[n - \sum_{i=1}^n w_j^{(i)}\right]$$

$$\Rightarrow \boxed{\hat\pi_1^{(k+1)} = \frac{\sum_{i=1}^n w_j^{(i)}}{n}}$$

The code for my implementation is at https://github.com/ishcha/ece566_hw/blob/main/hw3/em.py

The final estimates of θ are:

$\hat\pi_1^* = 0.32$

$\hat\mu_1^* = [-2.59, -4.26]$

$\hat\mu_2^* = [1.13, 1.36]$

$$\hat\Sigma_1^* = \begin{bmatrix} 0.85 & -0.02 \\ -0.02 & 0.88 \end{bmatrix} ; \quad \hat\Sigma_2^* = \begin{bmatrix} 3.12 & -1.61 \\ -1.61 & 2.89 \end{bmatrix}$$

I choose $\hat\theta^{(0)}$ randomly. Specifically, I fix $\hat\pi_1^{(0)} = 0.5$ and vary $\hat\mu_1^{(0)}, \hat\mu_2^{(0)}$ randomly, while $\hat\Sigma_1^{(0)}$ and $\hat\Sigma_2^{(0)}$ are initialized randomly from std. normal distribution, to have values close to 0.

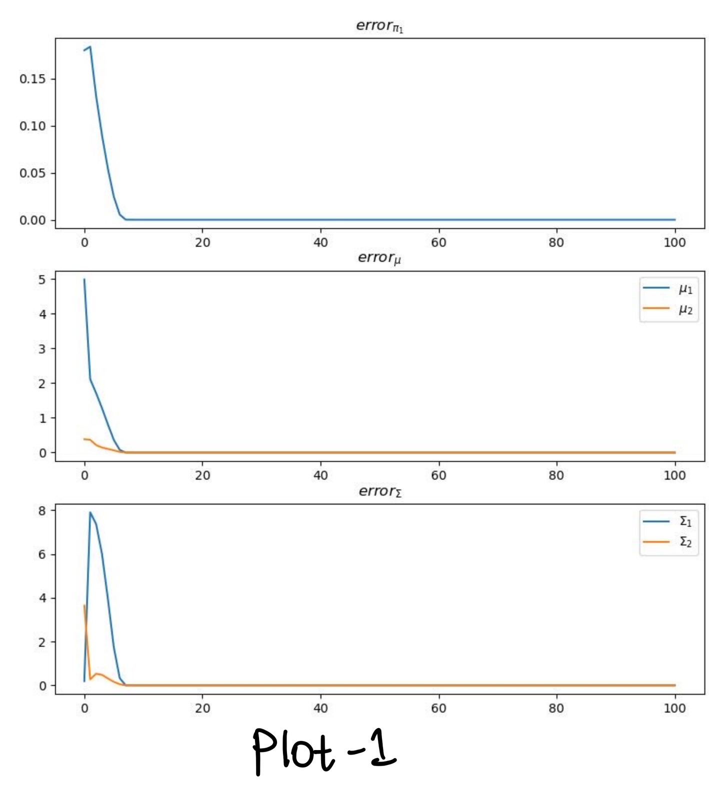
I check the Mean Square Errors of the estimates at each iteration with the final estimates for each parameter in θ next, wrt different initializations of $\hat\theta^{(0)}$.

(I use MSE as some parameters such as μ, Σ have multiple elements.)

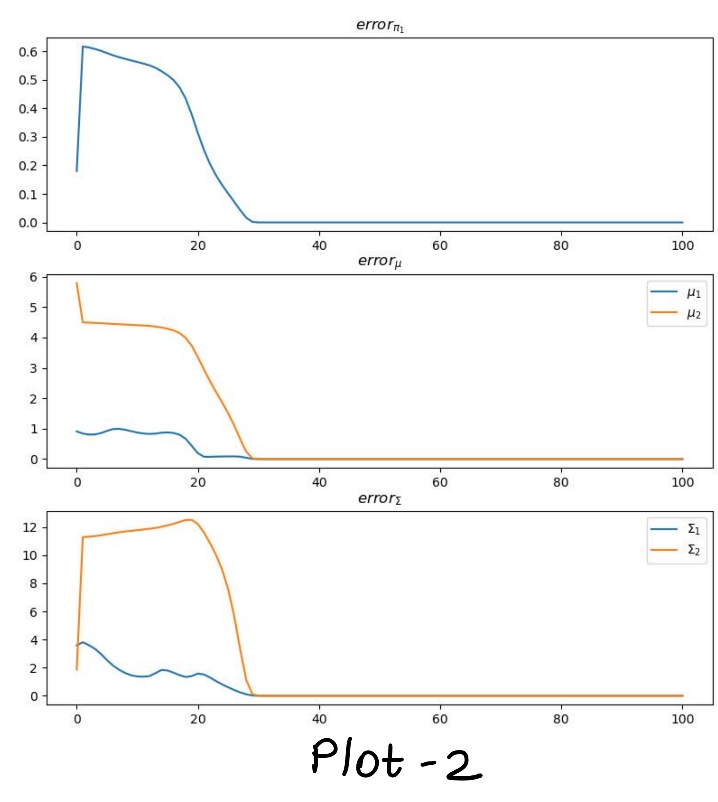
The below plots are for the MSE of the estimates with iterations for different random initializations.

x-axis is iterations, y-axis is error

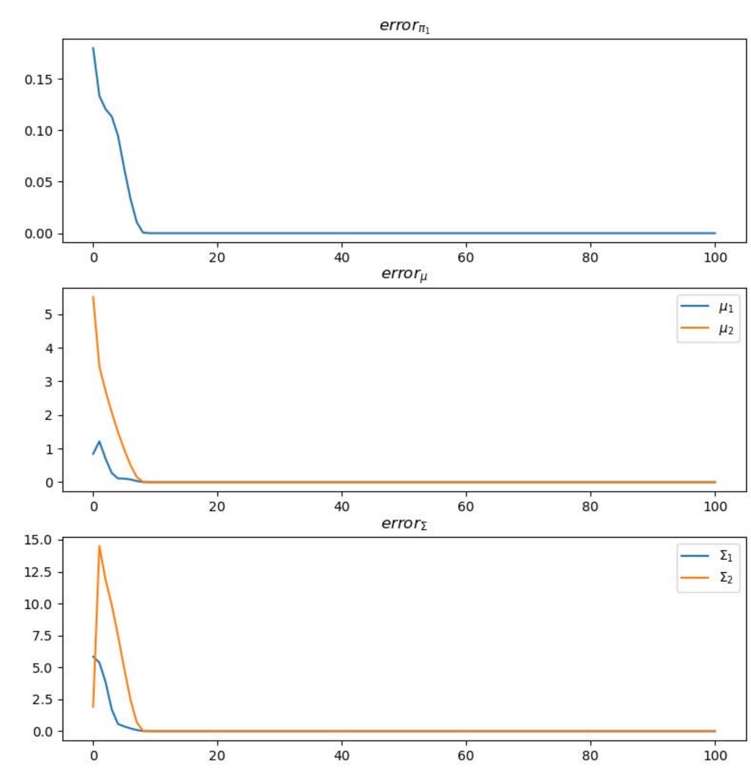
We observe that the E-M algorithm converges to the final estimate (error goes to 0) in less than 30-40 iterations for any random initializations.



Plot-1



Plot-2



Plot-3

Problem 3

Sunday, October 20, 2024 11:56 PM

$x \in \{0,1,2\}^5 \;;\; \mathcal{X} = \{0, 1, 2\}$

$$\mathcal{E}(u) = \begin{cases} \sum_t u_t & ; \prod_t u_t \neq 0 \\ -\infty & ; \text{else} \end{cases}$$

$$u_t = t - 3 + |x_{t+1} - x_t| \;;\; t = 1,2,3,4$$

$\prod_t u_t = 0$ occurs when $u_t = 0$

$$u_t = t - 3 + |x_{t+1} - x_t| = 0$$

$$\Rightarrow |x_{t+1} - x_t| = 3 - t$$

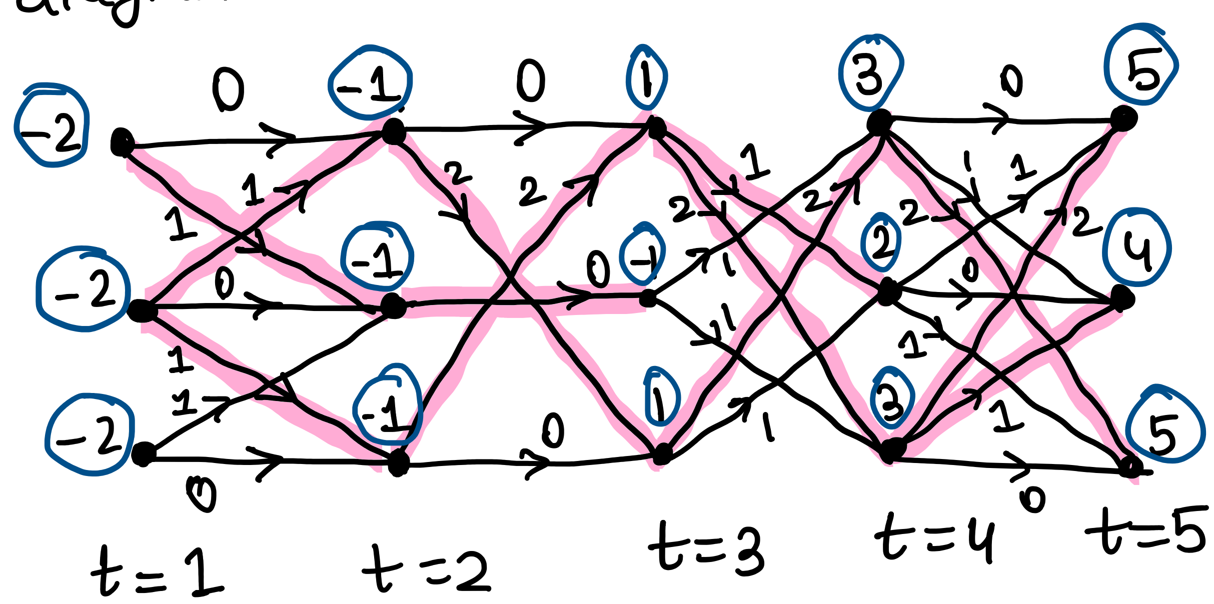
For $t=1$: $|x_{t+1} - x_t| = 2$
$t=2$: $|x_{t+1} - x_t| = 1$
$t=3$: $|x_{t+1} - x_t| = 0$
} these transitions are prohibited.

$$\max_{\substack{x \in \mathcal{X}^5 \\ \prod_t u_t \neq 0}} \mathcal{E}(u) = \sum_t u_t = -2 + \sum_{t=1}^{4} |x_{t+1} - x_t|$$

$$= f(x_1) + \sum_{t=1}^{4} g(x_t, x_{t+1})$$

where $f(x_1) = -2$; $g(x_t, x_{t+1}) = |x_{t+1} - x_t|$

Trellis diagram:



All surviving paths are highlighted in pink.

n : indicates value function of sequence till time t.

$\therefore$ max $\mathcal{E}(u)$ occurs at sequence : $x = (1, 2, 0, 2, 0)$

This is not a unique solution, but one of many maxima.

Another solution occurs at $x = (1, 0, 2, 0, 2)$

Problem 4:

Monday, October 21, 2024 12:15 AM

$$A^{(0)}(x,x') = 0$$

To prove: $A^{(k)}(x,x') = 0 \quad \forall\, k \geq 0$

Proof by induction on $k \in \{0,1,2,\dots\}$

Base case: $k=0$: $\hat{A}^{(0)}(x,x') = 0$ [$\because$ given]

Induction hypothesis: Let $\hat{A}^{(k)}(x,x') = 0$ for $k \geq 0$

Induction step:

$$\hat{A}^{(k+1)}(x,x') = \frac{\sum_{t=1}^{n-1} P_{\hat{\theta}^{(k)}}[X_t = x, X_{t+1} = x' \mid Y=y]}{\sum_{t=1}^{n-1} P_{\hat{\theta}^{(k)}}[X_t = x \mid Y=y]}$$

$$= \frac{\sum_{t=1}^{n-1} P_{\hat{\theta}^{(k)}}[X_{t+1} = x' \mid X_t = x, Y=y] \cdot P_{\hat{\theta}^{(k)}}[X_t = x \mid Y=y]}{\sum_{t=1}^{n} P_{\hat{\theta}^{(k)}}[X_t = x \mid Y=y]}$$

$$= \frac{\sum_{t=1}^{n-1} P_{\hat{\theta}^{(k)}}[X_{t+1} = x' \mid X_t = x] \cdot P_{\hat{\theta}^{(k)}}[X_t = x \mid Y=y]}{\sum_{t=1}^{n} P_{\hat{\theta}^{(k)}}[X_t = x \mid Y=y]}$$

$$= \frac{\hat{A}^{(k)}(x,x') \sum_{t=1}^{n} P_{\hat{\theta}^{(k)}}[X_t = x \mid Y=y]}{\sum_{t=1}^{n} P_{\hat{\theta}^{(k)}}[X_t = x \mid Y=y]}$$

$$= 0$$

Hence proved by induction.