HW5 Problem 1

Observables O: test scores {low, medium, high}

Hidden states S: health state {healthy, unhealthy}

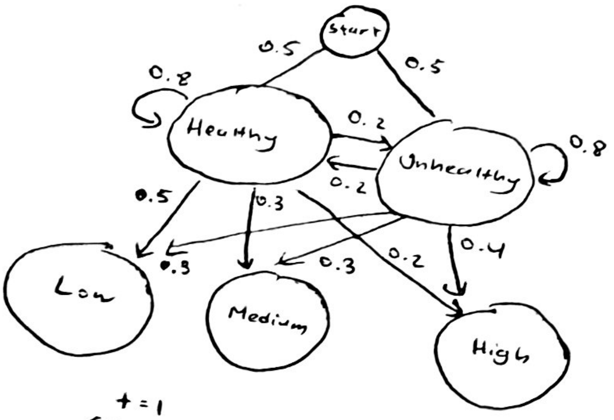
$\lambda = [\pi, a, b]$

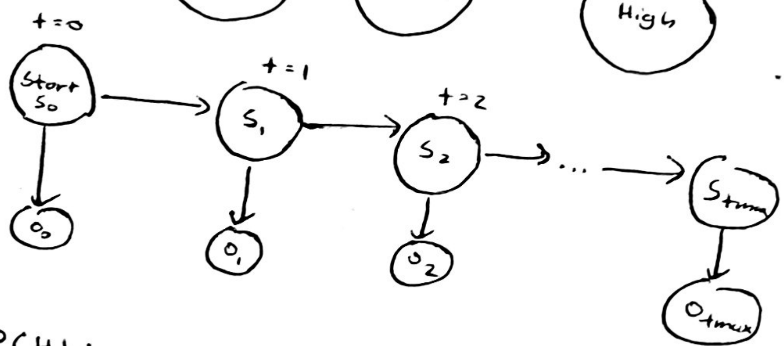
π = Initial state distribution = $[0.5 \quad 0.5]$

a = transition matrix = $\begin{array}{c|cc} & H & U \\ \hline H & 0.8 & 0.2 \\ U & 0.2 & 0.8 \end{array}$

b = emission matrix = $\begin{array}{c|ccc} & \text{low} & \text{med} & \text{high} \\ \hline H & 0.5 & 0.3 & 0.2 \\ U & 0.3 & 0.3 & 0.4 \end{array}$

Graphical Representation.





$P(H \mid low_{t=1}, low_{t=2})$

at t=0

$P(H) = 0.5$

$P(U) = 0.5$

at t=1

$P(H \mid low_{t=1}) = \frac{5}{8}$

$P(U \mid low_{t=1}) = \frac{3}{8}$

at t=2

$\left[\left(\frac{5}{8}\right)(0.8) + \left(\frac{3}{8}\right)(0.2)\right]\frac{5}{8} = 0.359375 \qquad \frac{0.359375}{0.359 + 0.159} = 0.69$

$\left[\left(\frac{3}{8}\right)(0.8) + \left(\frac{5}{8}\right)(0.2)\right]\frac{3}{8} = 0.159375 \qquad \frac{0.159375}{0.359 + 0.159} = 0.3$

$P(H \mid low_{t=1}, low_{t=2}) = 0.693$

H 0.5 $(0.5)(0.8)(0.5) + (0.5)(\overset{0.2}{\cancel{0.8}})(0.5) = 0.25$ $(0.25)(0.8)(0.5) + (0.15)(0.2)(0.5) = 0.115$

U 0.5 $(0.5)(0.8)(0.3) + (0.5)(0.2)(0.3) = 0.15$ $(0.15)(0.8)(0.3) + (0.25)(0.2)(0.3) = 0.05$

t t=0

Obs

t=1 test = low

t=2 test = low

$$\boxed{\text{Most likely path : } \underset{t=0}{H}\ \underset{t=1}{H}\ \underset{t=2}{H}}$$

Problem 2

Show that $\mu_{kj} = \dfrac{\sum_{i=1}^{m} r_k^{(i)} x_j^{(i)}}{\sum_{i=1}^{m} r_k^{(i)}}$

Bernoulli Distribution

$$L = p(x|\mu_k) = \prod_{i=1}^{m} \mu_k^{x} (1-\mu_k)^{(1-x^{(i)})}, \quad p(x|\mu,\pi) = \sum_{k=1}^{K} \pi_k\, p(x|\mu_k)$$

$$\ln L = \sum_{i=1}^{m} \ln\left[\sum_{k=1}^{K} \pi_k\, p(x|\mu_k)\right]$$

$$= \sum_{i=1}^{m}\sum_{k=1}^{K} z(\ln \pi_k + x \ln \mu_k + (1-x)\ln(1-\mu_k))$$

$$r_k^{(i)} = \frac{\pi_k\, p(x^{(i)}; \mu_k, \Sigma_k)}{\sum_{k=1}^{K} \pi_k\, p(x^{(i)}; \mu_k, \Sigma_k)} = E_x(z)$$

$$E_x(\ln L) = r_k\left[\ln \pi_k + \sum_{i=1}^{y}\left[x \ln \mu_k + (1-x)\ln(1-\mu_k)\right]\right]$$

$$\frac{\partial}{\partial \mu_k} E_x(\ln L) = r_k\left[+\frac{x}{\mu_k} - \frac{1-x}{1-\mu_k}\right] = 0$$

$$+\frac{r_k x}{\mu_k} - r_k\frac{1-x}{1-\mu_k} = 0$$

$$+\frac{r_k x}{\mu_k} = r_k\frac{1-x}{1-\mu_k}$$

$$+r_k x(1-\mu_k) = r_k(1-x)\mu_k$$

$$+r_k x - r_k x \mu_k = r_k \mu_k - r_k x \mu_k$$

$$r_k x = r_k \mu_k \longrightarrow \boxed{\frac{\sum_{i=1}^{m} r_k^{(i)} x_j^{(i)}}{\sum_{i=1}^{m} r_k^{(i)}} = \mu_k}$$

$$L = \prod_{i=1}^{m} \left(\sum_{k=1}^{k} \pi_k \, p(x^{(i)} \mid \mu_k) \right) \text{Beta}(\alpha, \beta) \quad (\text{Bayes Theorem})$$

$$\cancel{\ln L = \ln \sum_{k=1}^{k} \pi_k \, p(x \mid \mu_k)}$$

$$\ln L = \sum_{i=1}^{m} \ln \left[\sum_{k=1}^{k} \pi_k \, p(x \mid \mu_k) \right] + (\alpha - 1) \ln \mu_k + (\beta - 1) \ln (1 - \mu_k)$$

$$= \sum_{i=1}^{m} \left(\sum_{k=1}^{k} z \left[\ln \pi_k + x \ln \mu_k + (1-x) \ln (1 - \mu_k) \right] \right] + (\alpha - 1) \ln \mu_k + (\beta - 1) \ln (1 - \mu_k)$$

$$E_x(\ln L) = r_k \left[\ln \pi_k + \sum_{i=1}^{n} \left[x \ln \mu_k + (1-x) \ln (1 - \mu_k) \right] \right] + (\alpha - 1) \ln \mu_k + (\beta - 1) \ln (1 - \mu_k)$$

$$\frac{\partial}{\partial \mu_k} E_x(\ln L) = r_k \left[\frac{x}{\mu_k} - \frac{1-x}{1-\mu_k} \right] + \frac{\alpha - 1}{\mu_k} + - \frac{\beta - 1}{1 - \mu_k}$$

$$= \frac{r_k x}{\mu_k} - r_k \frac{1-x}{1-\mu_k} + \frac{\alpha - 1}{\mu_k} - \frac{\beta - 1}{1 - \mu_k} = 0$$

$$\frac{r_k x + \alpha - 1}{\mu_k} = \frac{r_k (1-x) + \beta - 1}{1 - \mu_k}$$

$$(r_k x + \alpha - 1)(1 - \mu_k) = \mu_k r_k - \mu_k r_k x + \beta \mu_k - \mu_k$$

$$r_k x + \alpha - 1 - \cancel{r_k \mu_k x} - \mu_k \alpha + \mu_k = \mu_k r_k - \cancel{\mu_k r_k x} + \beta \mu_k - \mu_k$$

$$r_k x + \alpha - 1 = \mu_k (r_k + \alpha + \beta - 2)$$

$$\boxed{\mu_k = \frac{\left(\sum_{i=1}^{m} r_k^{(i)} x_j^{(i)} \right) + \alpha - 1}{\left(\sum_{i=1}^{m} r_k^{(i)} \right) + \alpha + \beta - 2}}$$

Problem 3

$f_u(x) = \text{argmin}_{v \in V} \|x - v\|^2$

show that

$\text{argmin}_{u: uu^T = 1} \sum_{i=1}^{m} \|x^{(i)} - f_u(x^{(i)})\|^2$

Define a basis space

$x_n = \sum_{i=1}^{V} a_{ni} u_i$

$a_{ni} = x_n^T u_j$

$x_n = (x_n^T u_j) u_j$ when summed over J

Separating out first N basis vectors

$\hat{x} = \sum_{i=1}^{M} z_{ni} u_i + \sum_{i=M+1}^{V} b_i u_i = f_u(x^{(i)})$

Utilizing the loss function above

$J = \sum_{i=1}^{m} \|x^{(i)} - f_u(x^{(i)})\|^2$

when taking derivative after plugging in

$\frac{\partial J}{\partial z_{ni}} = 0 \Rightarrow z_{ni} = x_n^T u_i$

$\frac{\partial J}{\partial b_i} = 0 \Rightarrow b_i = \bar{x}^T u_i$

$x^{(i)} - f_u(x^{(i)}) = x^{(i)} - (z_{ni} u_i + b_i u_i)$

$= x^{(i)} - (\cancel{x_n^T u_i u_i} + \bar{x}^T u_i u_i)$, $x^{(i)} = (x^{(i)T} u_i) u_i$

in $M+1 \ldots V$

$= [(x^{(i)} - \bar{x}^T) u_i] u_i$

$J = \sum_{i=1}^{m} \sum_{i=M+1}^{V} (x^{(i)T} u - \bar{x}^T u)^2$

$S = \sum_{i=1}^{m} (x^{(i)} - \bar{x})(x^{(i)} - \bar{x})^T$

$J = u_i^T S u_i$ when summed from $M+1$ to V

choosing u_1

$J = u_1^T S u_1 + \lambda_1 (1 - u_1^T u)$

$\frac{\partial J}{\partial u_1} = 2S u_1 - 2\lambda_1 u_1 = 0$

$\therefore S u_1 = \lambda_1 u_1 \rightarrow$ same as variance max solution