Name: Parashara Ramesh	**Student Id**: A0285647M	**NUSNET Id**: E1216292

[**NOTE**: 'O' represents the number of Observations.]

1. E step (Expectation)

i. **Finding the scaled alphas and constants**

$\alpha(z_{o1k}) = \pi_k . N(x_{o1}, \mu_k, \sigma_k)$ (here the P(Xon|Zon) is the Gausian represented by the mu_k and sigma_k)

$\tilde{\alpha}(z_{o,n}) = p(x_{on}|z_{on})\Sigma_{z_{o,n-1}}\hat{\alpha}(z_{o,n-1}). A$ (here A is the transition matrix)

$c_{on} = \Sigma_{z_{on}}\tilde{\alpha}(z_{on})$

$\hat{\alpha}(z_{o,n}) = \frac{\tilde{\alpha}(z_{o,n})}{c_{on}}$

ii. Using the constants to find the scaled beta values

$\hat{\beta}(z_{oNk}) = 1$ (For all values of k initialize it to 1)

$\hat{\beta}(z_{on}) = \frac{\Sigma_{z_{n+1}}\hat{\beta}(z_{o,n+1})p(x_{o,n+1}|z_{o,n+1})p(z_{o,n+1}|z_{o,n})}{c_{o,n+1}}$ (Note here the last term in the matrix is the same as A)

iii. **Using the scaled alpha and beta values to find gamma**

$\gamma(z_{onk}) = \hat{\alpha}(z_{onk})\hat{\beta}(z_{onk})$

iv. Using the scaled alpha and beta values to find xi

$\xi(z_{o,n-1}, z_{o,n}) = \frac{\hat{\alpha}(z_{o,n-1}).p(x_{on}|z_{on}).A.\hat{\beta}(z_{on})}{c_{on}}$

2. M step (Maximization) [There are four parameters which needs to be maximized here (A, pi, phi.mu, phi.sigma)]

a. A:

$A_{jk} = \frac{\Sigma_{o=1}^{O}\Sigma_{n=2}^{N}\xi(z_{o,n-1,j}, z_{o,n,k})}{\Sigma_{o=1}^{O}\Sigma_{l=1}^{K}\Sigma_{n=2}^{N}\xi(z_{o,n-1,j}, z_{o,n,l})}$ The shape of xi is (O, N-1, K, K) (as there are N-1 windows of size 2 in a sequence of size N). By summing over timesteps and then over observations results in the numerator being of shape (K,K). In my implementation the denominator is obtained by summing the numerator term over axis=1 to get a shape (K). Then doing a column wise division of the numerator and denominator we obtain a final shape of (K,K)

b. Pi

$\pi_k = \frac{\Sigma_{o=1}^{O}\gamma_{o1k}}{\Sigma_{o=1}^{O}\Sigma_{j=1}^{K}\gamma_{o1j}}$ The numerator is basically the sum across all observation's gamma values for the 1st time step for the kth category normalized by the sum across all categories across all observations.
The shape of both the numerator & denominator finally is (O, K) and summed across all observations results in final shape of (K) in both the numerator & denominator, finally resulting in a vector of shape (K)

c. Phi.mu

$\mu_k = \frac{\Sigma_{o=1}^{O}\Sigma_{n=1}^{N}\gamma_{onk}x_{on}}{\Sigma_{o=1}^{O}\Sigma_{n=1}^{N}\gamma_{onk}}$ In the numerator, the product of the gamma value and x value results in a vector of shape (O,N,K) and after summing across all time steps results in a vector of shape (O,K) . After this is summed across all observations it results in a final shape of (K). Similarly in the denominator the gamma shape is (O,N,K) and after summing across all timesteps followed by summing across all observations results in a vector of shape (K). Dividing both gives a final shape of (K).

d. Phi.sigma

$\Sigma_k = \frac{\Sigma_{o=1}^{O}\Sigma_{n=1}^{N}\gamma(z_{o,n,k})(x_{on}-\mu_k)(x_{on}-\mu_k)^T}{\Sigma_{o=1}^{O}\Sigma_{n=1}^{N}\gamma(z_{o,n,k})}$ In the numerator, the product of all terms results in a vector of shape (O,N,K) and after summing across all time steps results in a vector of shape (O,K) . After this is summed across all observations it results in a vector of shape (K). Even in the denominator after 2 summations we get a final shape of (K) whose division results in a vector of shape(K)

3. Fit HMM

In my implementation, I iteratively call both the E-step (Expectation step) and M-step (Maximization step) functions within an infinite loop to update model parameters (A, pi, phi.mu, phi.sigma) until they converge within a 1e-4 threshold, potentially resulting in parameter estimates in a different order while still being valid.