2.2. Given:

$$[\sigma(\theta)]_i = \frac{e^{\theta_i}}{\sum_{j=0}^{M-1} e^{\theta_j}} \quad \cdots ①$$

$$\rho(a,b) = \sum_{i=0}^{M-1} -a_i \log b_i \quad \cdots ②$$

$$L(\theta) = \sum_{n=0}^{N-1} \rho(x_n, \sigma(\theta)) \quad \cdots ③$$

Substituting ② in ③ $(a = x_n$ and $b = \sigma(\theta))$

$$L(\theta) = \sum_{n=0}^{N-1} \sum_{i=0}^{M-1} \left[-x_{n,i} \log\left([\sigma(\theta)]_i\right) \right]$$

Substituting ① in the above equation

$$L(\theta) = \sum_{n=0}^{N-1} \sum_{i=0}^{M-1} \left[-x_{n,i} \log \left(\frac{e^{\theta_i}}{\sum_{j=0}^{M-1} e^{\theta_j}} \right) \right]$$

$$L(\theta) = \sum_{n=0}^{N-1} \sum_{i=0}^{M-1} \left[-x_{n,i} \left(\log e^{\theta_i} - \log \left(\sum_{j=0}^{M-1} e^{\theta_j} \right) \right) \right]$$

$$L(\theta) = \sum_{i=0}^{M-1} \left[\underbrace{\sum_{n=0}^{N-1} x_{n,i}}_{\text{From 2.1}} \left(-\theta_i + \log \left(\sum_{j=0}^{M-1} e^{\theta_j} \right) \right) \right]$$

$$L(\theta) = \sum_{i=0}^{M-1} \left[N_i \left(-\theta_i + \log \left(\sum_{j=0}^{M-1} e^{\theta_j} \right) \right) \right]$$