2.34 The log likelihood function is:

$$\ln p(\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = -\frac{ND}{2}\ln(2\pi) - \frac{N}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{n=1}^{N}(\mathbf{x}_n - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}_n - \boldsymbol{\mu})$$

Taking derivative w.r.t $\boldsymbol{\Sigma}$, we get:

$$\frac{\partial}{\partial \boldsymbol{\Sigma}}\ln p(\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = 0 - \frac{N}{2}(\boldsymbol{\Sigma}^{-1})^T - \frac{1}{2}\sum_{n=1}^{N}(-(\boldsymbol{\Sigma}^{-1})^T)(\mathbf{x}_n - \boldsymbol{\mu})^T(\mathbf{x}_n - \boldsymbol{\mu})(\boldsymbol{\Sigma}^{-1})^T$$

$$= -\frac{N}{2}(\boldsymbol{\Sigma}^{-1})^T + \frac{1}{2}\sum_{n=1}^{N}(\boldsymbol{\Sigma}^{-1})^T(\mathbf{x}_n - \boldsymbol{\mu})^T(\mathbf{x}_n - \boldsymbol{\mu})(\boldsymbol{\Sigma}^{-1})^T$$

Setting it to 0, we get:

$$0 = -\frac{N}{2}(\boldsymbol{\Sigma}^{-1})^T + \frac{1}{2}\sum_{n=1}^{N}(\boldsymbol{\Sigma}^{-1})^T(\mathbf{x}_n - \boldsymbol{\mu})(\mathbf{x}_n - \boldsymbol{\mu})^T(\boldsymbol{\Sigma}^{-1})^T$$

$$\implies \frac{N}{2}(\boldsymbol{\Sigma}^T)^{-1} = \frac{1}{2}\sum_{n=1}^{N}(\boldsymbol{\Sigma}^T)^{-1}(\mathbf{x}_n - \boldsymbol{\mu})(\mathbf{x}_n - \boldsymbol{\mu})^T(\boldsymbol{\Sigma}^T)^{-1}$$

$$\implies N(\boldsymbol{\Sigma}^T)^{-1} = \sum_{n=1}^{N}(\boldsymbol{\Sigma}^T)^{-1}(\mathbf{x}_n - \boldsymbol{\mu})(\mathbf{x}_n - \boldsymbol{\mu})^T(\boldsymbol{\Sigma}^T)^{-1}$$

Multiplying both sides on the right by $\boldsymbol{\Sigma}^T$, we get (Using equations 57 and 61 from the Matrix cookbook):

$$\implies N = \sum_{n=1}^{N}(\boldsymbol{\Sigma}^T)^{-1}(\mathbf{x}_n - \boldsymbol{\mu})(\mathbf{x}_n - \boldsymbol{\mu})^T$$

$$\implies N = (\boldsymbol{\Sigma}^T)^{-1}\sum_{n=1}^{N}(\mathbf{x}_n - \boldsymbol{\mu})(\mathbf{x}_n - \boldsymbol{\mu})^T$$

$$\implies \boldsymbol{\Sigma}^T = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n - \boldsymbol{\mu})(\mathbf{x}_n - \boldsymbol{\mu})^T$$

$$\implies \boldsymbol{\Sigma} = \frac{1}{N}\sum_{n=1}^{N}\left((\mathbf{x}_n - \boldsymbol{\mu})(\mathbf{x}_n - \boldsymbol{\mu})^T\right)^T$$

$$\implies \boldsymbol{\Sigma} = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n - \boldsymbol{\mu})(\mathbf{x}_n - \boldsymbol{\mu})^T$$

We can substitute $\boldsymbol{\mu}_{ML}$ for $\boldsymbol{\mu}$ here to get the maximum likelihood solution given by 2.122.

$$\implies \boldsymbol{\Sigma}_{ML} = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n - \boldsymbol{\mu}_{ML})(\mathbf{x}_n - \boldsymbol{\mu}_{ML})^T$$