Problem 2.13

The Gaussian Distribution:

$$\mathcal{N}(x|\mu,\Sigma) = \frac{1}{(2\pi)^{D/2}}\frac{1}{(|\Sigma|)^{1/2}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

KL Divergence:

$$KL(p(x)||q(x)) = \int p(x)ln\{\frac{p(x)}{q(x)}\}dx$$

With two above equations, we obtain:

$$KL(p(x)||q(x)) = \int p(x)\frac{1}{2}(ln(\frac{|L|}{|\Sigma|}) - (x-\mu)^T\Sigma^{-1}(x-\mu) + (x-m)^TL^{-1}(x-m))dx$$

We have property of expectation:

$$\begin{aligned} E[(x-\mu)^TA(x-\mu)] = Tr(A\Sigma) + (x-\mu)^TA(x-\mu) \\ E[x] = \int p(x)xdx \\ \int p(x)dx = 1 \end{aligned}$$

Hence:

$$\begin{aligned} KL(p(x)||q(x)) &= \frac{1}{2}(ln(\frac{|L|}{|\Sigma|}) + E[(x-\mu)^T\Sigma^{-1}(x-\mu)] - E[(x-m)^TL^{-1}(x-m)]) \\ &= \frac{1}{2}(ln(\frac{|L|}{|\Sigma|}) + Tr(\Sigma^{-1}\Sigma) + (x-\mu)^T\Sigma^{-1}(x-\mu) - Tr(L^{-1}\Sigma) - (x-m)^TL^{-1}(x-m)) \\ &= \frac{1}{2}(ln(\frac{|L|}{|\Sigma|}) + Tr(I_D) + (\mu-\mu)^T\Sigma^{-1}(\mu-\mu) - Tr(L^{-1}\Sigma) - (\mu-m)^TL^{-1}(\mu-m)) \\ &= \frac{1}{2}(ln(\frac{|L|}{|\Sigma|}) + D - Tr(L^{-1}\Sigma) - (\mu-m)^TL^{-1}(\mu-m)) \end{aligned}$$