Cross Entropy of Two Normal Distribution

by Arun Iyer

Given, $p(x) = \mathcal{N}(x|\mu, \Sigma)$ and $q(x) = \mathcal{N}(x|m, L)$.

$$
\begin{aligned}
H(p,q) &= -\int p(x) \ln q(x) dx \\
&= \int \mathcal{N}(x|\mu, \Sigma) \frac{1}{2}(D \ln(2\pi) + \ln |L| + (x-m)^\top L^{-1}(x-m)) dx
\end{aligned}
$$

Pushing the integral inside,

$$
\begin{aligned}
&= \frac{1}{2}(D \ln(2\pi) \int p(x) dx + \ln |L| \int p(x) dx + \int (x-m)^\top L^{-1}(x-m) p(x) dx) \\
&= \frac{1}{2}(D \ln(2\pi) + \ln |L| + \int (x-m)^\top L^{-1}(x-m) p(x) dx)
\end{aligned}
$$

Expanding the third term,

$$
\begin{aligned}
&= \frac{1}{2}(D \ln(2\pi) + \ln |L| + \int (x^\top L^{-1} x - m^\top L^{-1} x - x^\top L^{-1} m + m^\top L^{-1} m) p(x) dx) \\
&= \frac{1}{2}(D \ln(2\pi) + \ln |L| + E_{x \sim p(x)}[x^\top L^{-1} x] - \\
&\quad E_{x \sim p(x)}[m^\top L^{-1} x] - E_{x \sim p(x)}[x^\top L^{-1} m] + E_{x \sim p(x)}[m^\top L^{-1} m]) \\
&= \frac{1}{2}(D \ln(2\pi) + \ln |L| + E_{x \sim p(x)}[x^\top L^{-1} x] - \\
&\quad m^\top L^{-1} E_{x \sim p(x)}[x] - E_{x \sim p(x)}[x^\top] L^{-1} m + m^\top L^{-1} m) \\
&= \frac{1}{2}(D \ln(2\pi) + \ln |L| + E_{x \sim p(x)}[x^\top L^{-1} x] - m^\top L^{-1} \mu - \mu^\top L^{-1} m + m^\top L^{-1} m)
\end{aligned}
$$

Now, only that middle expectation is left. Note that, $x^\top L^{-1} x$ is a scalar real, therefore, $x^\top L^{-1} x = Tr(x^\top L^{-1} x)$. Therefore,

$$
\begin{aligned}
E[x^\top L^{-1} x] &= E[Tr(x^\top L^{-1} x)] \\
&= E[Tr(L^{-1} x x^\top)] \\
&= Tr(E[L^{-1} x x^\top]) \cdots \text{since both Tr and E are linear} \\
&= Tr(L^{-1} E[x x^\top])
\end{aligned}
$$

Now, $Cov(x) = E[xx^\top] - E[x]E[x]^\top$, i.e $\Sigma = E[xx^\top] - \mu\mu^\top$. Therefore,

$$
E[x^\top L^{-1} x] = Tr(L^{-1}(\mu\mu^\top + \Sigma))
$$

Substituting this back in the earlier expression gives us the result.