Entropy of multivariate Gaussian

$$H(x) = E(-\log P(x)) \quad , \quad X \sim N(\mu, \Sigma)$$

$$P(x) = \frac{1}{(2\pi)^{N/2} \cdot |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right]$$

$$\log P(x) = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)$$

$$-\log P(x) = \frac{N}{2}\log(2\pi) + \frac{1}{2}\log|\Sigma| + \frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)$$

Now calculating the entropy

$$H(x) = E\left(-\log P(x)\right) = \int_{-\infty}^{\infty} -\log P(x) \cdot P(x)\, dx$$

$$= \int_{-\infty}^{\infty} \left(\frac{N}{2} \log 2\pi + \frac{1}{2} \log |\Sigma| + \frac{1}{2} (x-\mu)^T \Sigma^{-1} (x-\mu) \right) P(x)\, dx$$

$$H(x) = \underbrace{\int_{-\infty}^{\infty} \frac{N}{2} \log 2\pi \cdot P(x)\, dx}_{A} + \underbrace{\int_{-\infty}^{\infty} \frac{1}{2} \log |\Sigma|\, P(x)\, dx}_{B} + \underbrace{\int_{-\infty}^{\infty} \frac{1}{2} (x-\mu)^T \Sigma^{-1} (x-\mu)\, P(x)\, dx}_{C}$$

$$\boxed{H(x) = A + B + C}$$

1) $$A = \int_{-\infty}^{\infty} \frac{N}{2} \log 2\pi \; P(x)\, dx$$

$$= \frac{N}{2} \log 2\pi \int_{-\infty}^{\infty} P(x)\, dx$$

$$\boxed{A = \frac{N}{2} \log 2\pi} \quad \text{--- (1)}$$

2) $$B = \int_{-\infty}^{\infty} \frac{1}{2} \log |\Sigma| \; P(x)\, dx$$

$$= \frac{1}{2} \log |\Sigma| \int_{-\infty}^{\infty} P(x)\, dx$$

$$\boxed{B = \frac{1}{2} \log |\Sigma|} \quad \text{--- (ii)}$$

3) $C = \int_{-\infty}^{\infty} \frac{1}{2} (x-\mu)^T \Sigma^{-1} (x-\mu) \cdot P(x) \, dx$

Now $(x-\mu)^T \Sigma^{-1} (x-\mu)$ is in $(1 \times N) \times (N \times N) \times (N \times 1)$

in $\mathbb{R}^1$

Hence

$$(x-\mu)^T \Sigma^{-1} (x-\mu) = \text{trace}\left((x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$

$$\boxed{\text{trace}\left((x-\mu)^T \Sigma^{-1} (x-\mu)\right) = \text{trace}\left(\Sigma^{-1} (x-\mu)(x-\mu)^T\right)}$$

2019

Using the trace trick,

$$C = \frac{1}{2} \int_{-\infty}^{\infty} \text{trace}\left[\left(\Sigma^{-1} (x-\mu)(x-\mu)^T\right) P(x)\right] dx$$

$$C = \frac{1}{2} \text{trace}\left[\int_{-\infty}^{\infty} \Sigma^{-1} (x-\mu)(x-\mu)^T P(x)\, dx\right]$$

$$C = \frac{1}{2} \text{trace}\left[\Sigma^{-1} \int_{-\infty}^{\infty} (x-\mu)(x-\mu)^T P(x)\, dx\right]$$

$$C = \frac{1}{2} \text{trace}\left[\Sigma^{-1} \Sigma\right] = \frac{1}{2} \text{trace}\left[I\right]$$

$$\boxed{C = \frac{N}{2}} \quad - (3)$$

or $\boxed{C = \frac{N}{2}}$ — 3

Combing 1, 2, 3

$$H(x) = \frac{N}{2}\log 2\pi + \frac{1}{2}\log|\Sigma| + \frac{N}{2}\log e$$

$$= \frac{N}{2}\log 2\pi e + \log|\Sigma|^{1/2}$$

$$= \cancel{\frac{1}{2}\log(2\pi e)^{N}}$$

$$\boxed{H(x) = \frac{N}{2}\log 2\pi e + \frac{1}{2}\log|\Sigma|}$$