TTIC 31200 – Information and Coding Theory – Discussion 3

Kavya Ravichandran

24 January 2025

1 Multivariate Gaussians

Recall the density of a d-dimensional Gaussian random variable X with mean $\mu \in \mathbb{R}^d$ and covariance $\Sigma \in \mathbb{R}^{d \times d}$:

$$f(x) = \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right).$$

Recall that two random variables are independent when the joint density is the product of the marginals.

Exercise 1.1. *Show that when $\Sigma = \mathbb{I}_d$, we have d independent 1-dimensional random variables*

Solution: W hen $\Sigma = \mathbb{I}_d$:

$$\begin{aligned}
f(x) &= \frac{1}{\sqrt{(2\pi)^d}} \exp\left(-\frac{1}{2}(x-\mu)^\top(x-\mu)\right) && (1)\\
&= \frac{1}{\sqrt{(2\pi)^d}} \exp\left(-\frac{1}{2}\sum_{i=1}^{d}(x_i-\mu_i)^2\right) && (2)\\
&= \Pi_{i=1}^{d} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_i-\mu_i)^2\right). && (3)
\end{aligned}$$

Since the joint distribution factors into the marginals, we can say that when $\Sigma = \mathbb{I}_d$, we actually just have d independent 1-dimensional random variables.

Exercise 1.2. *Show that when $\Sigma = diag\begin{pmatrix}\sigma_1^2 & \sigma_2^2 & \dots & \sigma_n^2\end{pmatrix}$, we have d independent 1-dimensional random variables*

Solution: W hen $\Sigma = \operatorname{diag}\begin{pmatrix}\sigma_1^2 & \sigma_2^2 & \dots & \sigma_n^2\end{pmatrix}$:

$$\begin{aligned}
f(x) &= \frac{1}{\sqrt{(2\pi)^d}} \exp\left(-\frac{1}{2}(x-\mu)^\top \begin{pmatrix}\sigma_1^2 & 0 & \dots & 0\\ 0 & \sigma_2^2 & \dots & 0\\ \vdots & & & \\ 0 & 0 & \dots & \sigma_n^2\end{pmatrix}^{-1} (x-\mu)\right) && (4)\\
&= \frac{1}{\sqrt{(2\pi)^d}} \exp\left(-\frac{1}{2}\sum_{i=1}^{d}\frac{(x_i-\mu_i)^2}{\sigma_i^2}\right) && (5)\\
&= \Pi_{i=1}^{d} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x_i-\mu_i)^2}{2\sigma^2}\right). && (6)
\end{aligned}$$

Since the joint distribution factors into the marginals, we can say that when Σ has the stated structure, we actually just have d independent 1-dimensional random variables.

Note that when Σ is a general PSD matrix, the density is a general joint distribution.

2 Differential Entropy of Gaussian

Let us compute the entropy of a multivariate Gaussian. Recall the definition of differential entropy:

Definition 2.1. *(Section 8.1 in Cover and Thomas) The differential entropy $h(X)$ of a continuous random variable X with density $f(x)$ is defined as:*

$$h(X) = -\int_S f(x) \log f(x)\, dx$$

where S is the support set of the random variable.

Exercise 2.1. *Compute $H(X)$ where X is a multivariate Gaussian with mean $\mu \in \mathbb{R}^d$ and covariance $\Sigma \in \mathbb{R}^{d\times d}$.*

Solution: W e start by plugging in the definitions:

$$\begin{aligned}
h(X) &= -\int_{-\infty}^{\infty} \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right) && (7)\\
&\qquad \log\left(\frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right)\right) dx && (8)\\
&= -\int_{-\infty}^{\infty} \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right) && (9)\\
&\qquad \left(\log\left(\frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}}\right) + \frac{1}{\ln 2}\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right)\right) dx && (10)\\
&= \mathbb{E}_f\left[\frac{1}{2}(d\log(2\pi) + \log\det\Sigma)\right] + \frac{1}{2\ln 2}\mathbb{E}_f\left[(x-\mu)^\top \Sigma^{-1}(x-\mu)\right] && (11)\\
&= \frac{d}{2}\log(2\pi) + \log\det\Sigma + \frac{d}{2\ln 2}\,. && (12)
\end{aligned}$$

Note that the last term in Eqn. 11 is the expectation of the ℓ_2 norm of a random variable Y that is a standard Gaussian, which is d. We could also get this by writing out the integral and doing a change of basis.

3 Total Variation Distance Between Two Gaussians

We did not get to this during the discussion, but it is a useful computation to work through, so I am including it in the notes for reference!

Finally, let us study the total variation distance between two Gaussians. Interestingly, there is no known closed form! However, we are lucky to have a tool at our disposal to still try to get a bound – Pinsker's Inequality!

Lemma 3.1. *Pinsker's Inequality (from lecture notes) For two distributions P, Q supported on $\mathcal{X}$:*

$$D(P||Q) \geq \frac{1}{2\ln 2}||P-Q||_1^2\,.$$

We also know that $\delta_{TV}(P,Q) = \frac{1}{2}||P-Q||_1$, so if we can bound $D(P||Q)$, we can get a bound on the total variation distance. So let us focus on that. Let $P \sim \mathcal{N}(\mu_p, \Sigma_p)$, $Q \sim \mathcal{N}(\mu_Q, \Sigma_Q)$.

$$
\begin{aligned}
D(P||Q) &= \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)} && (13)\\
&= \mathbb{E}_{x\sim P}\left[\log \frac{p(x)}{q(x)}\right] && (14)\\
&= \mathbb{E}_{x\sim P}\left[\log\left(\sqrt{\frac{(2\pi)^d \det \Sigma_2}{(2\pi)^d \det \Sigma_1}} \cdot \frac{\left(-\frac{1}{2}(x-\mu_1)^\top \Sigma_1^{-1}(x-\mu_1)\right)}{\left(-\frac{1}{2}(x-\mu_2)^\top \Sigma_2^{-1}(x-\mu_2)\right)}\right)\right] && (15)\\
&= \frac{1}{\ln 2}\left(\frac{1}{2}\ln\frac{\det \Sigma_2}{\det \Sigma_1} + \mathbb{E}_{x\sim P}\left[\ln\left(\exp\left(-\frac{1}{2}\left((x-\mu_1)^\top \Sigma_1^{-1}(x-\mu_1) - (x-\mu_2)^\top \Sigma_2^{-1}(x-\mu_2)\right)\right)\right)\right]\right) && (16)\\
&= \frac{1}{\ln 2}\left(\frac{1}{2}\ln\frac{\det \Sigma_2}{\det \Sigma_1} + \frac{1}{2}(\mu_1-\mu_2)^\top \Sigma_1^{-1}(\mu_1-\mu_2) + \frac{1}{2}\mathrm{tr}(\Sigma_1^{-1}\Sigma_2 - I)\right) && (17)
\end{aligned}
$$

And so now, plugging back into Pinsker's Theorem, we have:

$$
\delta_{TV}(P,Q) \le \sqrt{\frac{\ln 2}{2} D(P||Q)} \le \frac{1}{4}\left(\ln\frac{\det \Sigma_2}{\det \Sigma_1} + (\mu_1-\mu_2)^\top \Sigma_1^{-1}(\mu_1-\mu_2) + \mathrm{tr}(\Sigma_1^{-1}\Sigma_2 - I)\right).
$$