NNDL2025: Exercises & Assignments
Session 0 (13 Mar)
Solutions of mathematical exercises

1. Recall that $\mathbf{U}$ is orthogonal for a symmetric matrix $\mathbf{C}$. We have

$$\begin{aligned}
\mathbf{M}\,\mathbf{M} &= \mathbf{U}\,\mathrm{diag}\left(\sqrt{\lambda_1},\ldots,\sqrt{\lambda_n}\right)\mathbf{U}^T\,\mathbf{U}\,\mathrm{diag}\left(\sqrt{\lambda_1},\ldots,\sqrt{\lambda_n}\right)\mathbf{U}^T \\
&= \mathbf{U}\left(\mathrm{diag}\left(\sqrt{\lambda_1},\ldots,\sqrt{\lambda_n}\right)\mathrm{diag}\left(\sqrt{\lambda_1},\ldots,\sqrt{\lambda_n}\right)\right)\mathbf{U}^T \\
&= \mathbf{U}\,\mathrm{diag}\left(\lambda_1,\ldots,\lambda_n\right)\mathbf{U}^T = \mathbf{C},
\end{aligned}$$

   where the second equality is due to the fact that $\mathbf{U}$ is orthogonal.

2. Consider N observations denoted as $\mathbf{z} = \{z_1, \ldots, z_N\}$. The optimization problem can be formulated as

$$\begin{aligned}
\hat{\alpha} &= \arg\max_{\alpha} \prod_i p(z_i|\alpha) \\
&= \arg\max_{\alpha} \sum_i \log p(z_i|\alpha) \\
&= \arg\max_{\alpha} N \log \alpha - \alpha \sum_i |z_i| \,.
\end{aligned}$$

   Take the derivative of $N \log \alpha - \alpha \sum_i |z_i|$ w.r.t. α, we have

$$\frac{N}{\alpha} - \sum_i |z_i| \,,$$

   Setting the derivative to zero, we have the estimator

$$\hat{\alpha} = \frac{N}{\sum_i |z_i|}.$$

   A more complete derivation would further consider if this is necessarily a maximum, and if there might be other maxima. In fact, we can clearly see that the derivative is greater than 0 when $\alpha < \frac{N}{\sum_i |z_i|}$, and smaller than 0 when $\alpha > \frac{N}{\sum_i |z_i|}$, so this is a maximum and there are no others.

3. (a) The joint distribution is given by just plugging in the quantities in the definition as:

$$p(x, y) = (2\pi)^{-\frac{2}{2}} \left| \begin{bmatrix} 1 & c \\ c & 1 \end{bmatrix} \right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \begin{bmatrix} x \\ y \end{bmatrix}^\top \begin{bmatrix} 1 & c \\ c & 1 \end{bmatrix}^{-1} \begin{bmatrix} x \\ y \end{bmatrix} \right) \tag{1}$$

Using the give formula for the inverse, we can rewrite this as:

$$p(x, y) = (2\pi)^{-1} \left(1 - c^2\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \begin{bmatrix} x \\ y \end{bmatrix}^\top \frac{1}{1 - c^2} \begin{bmatrix} 1 & -c \\ -c & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \right) \tag{2}$$

and doing the matrix-vector multiplications explicitly for the scalar quantities, we can write:

$$p(x, y) = (2\pi)^{-1} \left(1 - c^2\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \frac{x^2 - 2cxy + y^2}{1 - c^2} \right) \tag{3}$$

(b) We need to integrate the above pdf over y. Now, we use the completion of square to give

$$p(x, y) = (2\pi)^{-1} \left(1 - c^2\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \frac{x^2 + (y - cx)^2 - (cx)^2}{1 - c^2} \right) \tag{4}$$

which can be further manipulated (grouping terms in x, using the exponential of a sum) to give

$$p(x, y) = (2\pi)^{-1} \left(1 - c^2\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} x^2 \right) \exp\left(-\frac{1}{2} \frac{(y - cx)^2}{1 - c^2} \right). \tag{5}$$

Thus, the marginal distribution is

$$p(x) = (2\pi)^{-1} \left(1 - c^2\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} x^2 \right) \int \exp\left(-\frac{1}{2} \frac{(y - cx)^2}{1 - c^2} \right) dy \tag{6}$$

where we have moved everything not depending of y to outside the integral.

Now, the trick is to realize that the integrand is, up to a normalizing constant, like a pdf of y where the mean is cx (as if x were fixed) and the variance $1 - c^2$. That normalizing constant is a simple function of c but we don't need to know it. We only need to know it is a constant and does not depend x, because the Gaussian normalization constant $\frac{1}{\sqrt{2\pi\sigma^2}}$ does not depend on the mean but only on the variance here denoted by σ^2. So, the whole integral is a constant that only depends on c.

On the other hand, anything before the integral in (6) is just like a standardized Gaussian distribution, except for a constant that is a function of c. Thus the whole expression is the pdf of a standardized Gaussian distribution up to a global constant which is a function of c. But again, that constant has to be the one that normalizes the pdf to integrate to unity, so again, we don't need to calculate it. Thus, we have shown that $x \sim \mathcal{N}(0,1)$.

(c) First note that by symmetry, we have the marginal distribution $y \sim \mathcal{N}(0,1)$ just like for x. Likewise, we can switch x and y in (4) and get

$$p(x,y) = (2\pi)^{-1}\left(1-c^2\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\frac{y^2+(x-cy)^2-(cy)^2}{1-c^2}\right) \quad (7)$$

Now, we have by definition

$$p(x|y) = \frac{p(x,y)}{p(y)} \quad (8)$$

and plugging in (7) as well as the standardized Gaussian pdf for y, we get

$$\begin{aligned} &p(x|y) = \\ &(2\pi)^{-1}\left(1-c^2\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\frac{y^2+(x-cy)^2-(cy)^2}{1-c^2}\right)(2\pi)^{-1}\exp\left(-\frac{1}{2}y^2\right)^{-1} = \\ &(2\pi)^{-1}\left(1-c^2\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}y^2-\frac{1}{2}\frac{(x-cy)^2}{1-c^2}\right)(2\pi)^{-1}\exp\left(\frac{1}{2}y^2\right) = \\ &\qquad (2\pi)^{-2}\left(1-c^2\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\frac{(x-cy)^2}{1-c^2}\right) \end{aligned} \quad (9)$$

Again, everything in front of the exponential is a constant with respect to x. So, it is just a normalizing constant that makes the integral of this expression over x to be equal to unity, and we can just ignore it. What remains in the pdf of a Gaussian variable with mean cy and variance $1-c^2$. Thus, to $x|y \sim \mathcal{N}(cy, 1-c^2)$.