Tutorial Questions

March 21, 2025

Exercise 6.6.1

Consider two centered random vectors $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$ with covariance matrices Σ_{XX}, Σ_{YY}, and cross-covariance Σ_{XY}. Show that the canonical correlation problem:

$$\max_{a\in\mathbb{R}^p, b\in\mathbb{R}^q} \frac{a^\top \Sigma_{XY} b}{\sqrt{(a^\top \Sigma_{XX} a)(b^\top \Sigma_{YY} b)}}$$

is equivalent to the generalized eigenvalue problem:

$$\begin{pmatrix} 0 & \Sigma_{XY} \\ \Sigma_{YX} & 0 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \lambda \begin{pmatrix} \Sigma_{XX} & 0 \\ 0 & \Sigma_{YY} \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}$$

Solution

First, we equivalently write the problem with normalization constraints:

$$\max_{a,b} \ a^\top \Sigma_{XY} b \ \text{ subject to } \ a^\top \Sigma_{XX} a = 1, \ b^\top \Sigma_{YY} b = 1$$

This is because the standard CCA normalization of a and b is $\alpha = \Sigma_{XX}^{1/2} a$ and $\beta = \Sigma_{YY}^{1/2} b$. Thus,

$$a^\top \Sigma_{XX} a = 1 \iff 1 = a^\top \Sigma_{XX}^{1/2} \Sigma_{XX}^{1/2} a = \alpha^\top \alpha,$$

and

$$b^\top \Sigma_{YY} b = 1 \iff 1 = b^\top \Sigma_{YY}^{1/2} \Sigma_{YY}^{1/2} b = \beta^\top \beta.$$

Substitute the above back into the previous optimization problem, we see that the previous optimization problem becomes the standard CCA optimization of the form:

$$\max_{\alpha,\beta} \ \alpha^\top M \beta \ \text{ subject to } \ \|\alpha\| = 1, \ \|\beta\| = 1,$$

where $M = \Sigma_{XX}^{-1/2} \Sigma_{XY} \Sigma_{YY}^{-1/2}$.

Now we introduce Lagrange multipliers $\lambda, \mu \in \mathbb{R}$:

$$\mathcal{L}(a, b, \lambda, \mu) = a^\top \Sigma_{XY} b - \frac{\lambda}{2}(a^\top \Sigma_{XX} a - 1) - \frac{\mu}{2}(b^\top \Sigma_{YY} b - 1)$$

Take partial derivatives and set to zero:

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial a} = \Sigma_{XY} b - \lambda \Sigma_{XX} a = 0 \\ \frac{\partial \mathcal{L}}{\partial b} = \Sigma_{YX} a - \mu \Sigma_{YY} b = 0 \end{cases}$$

Notice that $\Sigma_{YX} = \Sigma_{XY}^\top$. Multiply the first equation by $a^\top$ and second by $b^\top$:

$$\begin{cases} a^\top \Sigma_{XY} b = \lambda a^\top \Sigma_{XX} a = \lambda \\ b^\top \Sigma_{YX} a = \mu b^\top \Sigma_{YY} b = \mu \end{cases}$$

Since $a^\top \Sigma_{XY} b = b^\top \Sigma_{YX} a$, we get $\lambda = \mu$.

Rewrite the system as:

$$\begin{cases} \Sigma_{XY} b = \lambda \Sigma_{XX} a \\ \Sigma_{YX} a = \lambda \Sigma_{YY} b \end{cases}$$

This can be expressed in block matrix form:

$$\begin{pmatrix} 0 & \Sigma_{XY} \\ \Sigma_{YX} & 0 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \lambda \begin{pmatrix} \Sigma_{XX} & 0 \\ 0 & \Sigma_{YY} \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}$$

In addition, we can solve the above eigenvalue problems by considering

$$\begin{pmatrix} \Sigma_{XX} & 0 \\ 0 & \Sigma_{YY} \end{pmatrix}^{-1} \begin{pmatrix} 0 & \Sigma_{XY} \\ \Sigma_{YX} & 0 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \lambda \begin{pmatrix} a \\ b \end{pmatrix},$$

which is equivalent to

$$\begin{pmatrix} -\lambda I_p & \Sigma_{XX}^{-1} \Sigma_{XY} \\ \Sigma_{YY}^{-1} \Sigma_{YX} & -\lambda I_q \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = 0.$$

Therefore,

$$\det(\Sigma_{XX}^{-1} \Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{YX} - \lambda^2 I_p) = 0.$$

Thus, λ is the canonical correlation coefficient. If we let $\lambda' = \lambda^2$, and multiply the above determinant by Σ_{XX} on both sides, we get

$$\det(\Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{YX} - \lambda' \Sigma_{XX}) = 0,$$

which is what Exercise 6.6.3 is looking for.

Exercise 6.6.4

Suppose that $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$ are jointly Gaussian random vectors with mean zero and covariance matrix:

$$\begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix}.$$

Show that the canonical correlation variables $(\eta_i, \phi_i) = (a_i^\top X, b_i^\top Y)$ are jointly Gaussian and derive their joint distribution explicitly.

Solution

Since X and Y are jointly Gaussian, the stacked vector $\begin{pmatrix} X \\ Y \end{pmatrix}$ follows a multivariate normal distribution:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix}\right).$$

The canonical variables $\eta_i = a_i^\top X$ and $\phi_i = b_i^\top Y$ are linear transformations of X and Y. A fundamental property of multivariate Gaussian distributions is that *any linear combination of jointly Gaussian variables remains jointly Gaussian.* Therefore, (η_i, ϕ_i) are jointly Gaussian.

Mean Vector

Both variables have zero mean:

$$\mathbb{E}[\eta_i] = a_i^\top \mathbb{E}[X] = 0, \quad \mathbb{E}[\phi_i] = b_i^\top \mathbb{E}[Y] = 0.$$

Covariance Matrix

Compute the covariance components:

Variances:

$$\begin{aligned} \mathrm{Var}(\eta_i) &= a_i^\top \Sigma_{XX} a_i = 1 \quad \text{(CCA unit norm constraint)}, \\ \mathrm{Var}(\phi_i) &= b_i^\top \Sigma_{YY} b_i = 1 \quad \text{(CCA unit norm constraint)}. \end{aligned}$$

Cross-Covariance:

$$\mathrm{Cov}(\eta_i, \phi_i) = \mathbb{E}[a_i^\top X Y^\top b_i] = a_i^\top \Sigma_{XY} b_i = \lambda_i,$$

where λ_i is the i-th canonical correlation coefficient by definition.

The joint distribution is therefore:

$$\begin{pmatrix} \eta_i \\ \phi_i \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} 1 & \lambda_i \\ \lambda_i & 1 \end{pmatrix}\right).$$

Exercise 6.6.6

Let $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$ be random vectors with canonical correlations $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq 0$, where $r = \min(p, q)$. Show that if X and Y are independent, then all canonical correlations satisfy $\sigma_i = 0$ for $i = 1, 2, \ldots, r$.

Solution

If X and Y are independent, their cross-covariance matrix satisfies:

$$\Sigma_{XY} = \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])^\top\right] = 0.$$

This follows because independence implies $\mathbb{E}[XY^\top] = \mathbb{E}[X]\mathbb{E}[Y]^\top$, and for centered vectors ($\mathbb{E}[X] = \mathbb{E}[Y] = 0$), we have $\Sigma_{XY} = 0$.

Canonical correlations $\{\sigma_i\}$ are the singular values of the matrix:

$$M = \Sigma_{XX}^{-1/2}\Sigma_{XY}\Sigma_{YY}^{-1/2}.$$

Substituting $\Sigma_{XY} = 0$ into M gives:

$$M = \Sigma_{XX}^{-1/2} \cdot 0 \cdot \Sigma_{YY}^{-1/2} = 0.$$

The singular values of the zero matrix are all zero. Hence, $\sigma_i = 0$ for all i.

Exercise 6.6.7

Given two centered random vectors $X \in \mathbb{R}^2$ and $Y \in \mathbb{R}^2$ with joint covariance matrix:

$$\Sigma = \begin{bmatrix} 1 & 0.8 & 0.5 & 0.3 \\ 0.8 & 1 & 0.6 & 0.4 \\ 0.5 & 0.6 & 1 & 0.9 \\ 0.3 & 0.4 & 0.9 & 1 \end{bmatrix},$$

compute the canonical correlations between X and Y numerically.

Detailed Solution

$$\Sigma_{XX} = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}, \quad \Sigma_{XY} = \begin{bmatrix} 0.5 & 0.3 \\ 0.6 & 0.4 \end{bmatrix}, \quad \Sigma_{YY} = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}, \quad \Sigma_{YX} = \Sigma_{XY}^\top$$

Compute Σ_{XX}^{-1}:

$$\det(\Sigma_{XX}) = (1)(1) - (0.8)^2 = 0.36$$

$$\Sigma_{XX}^{-1} = \frac{1}{0.36}\begin{bmatrix} 1 & -0.8 \\ -0.8 & 1 \end{bmatrix} \approx \begin{bmatrix} 2.7778 & -2.2222 \\ -2.2222 & 2.7778 \end{bmatrix}$$

Compute Σ_{YY}^{-1}:

$$\det(\Sigma_{YY}) = (1)(1) - (0.9)^2 = 0.19$$

$$\Sigma_{YY}^{-1} = \frac{1}{0.19}\begin{bmatrix} 1 & -0.9 \\ -0.9 & 1 \end{bmatrix} \approx \begin{bmatrix} 5.2632 & -4.7368 \\ -4.7368 & 5.2632 \end{bmatrix}$$

Compute $\Sigma_{XY}\Sigma_{YY}^{-1}$:

$$\begin{bmatrix} 0.5 & 0.3 \\ 0.6 & 0.4 \end{bmatrix}\begin{bmatrix} 5.2632 & -4.7368 \\ -4.7368 & 5.2632 \end{bmatrix} \approx \begin{bmatrix} 1.2106 & -0.7894 \\ 1.2632 & -0.7368 \end{bmatrix}$$

Compute $\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}$:

$$\begin{bmatrix} 1.2106 & -0.7894 \\ 1.2632 & -0.7368 \end{bmatrix}\begin{bmatrix} 0.5 & 0.6 \\ 0.3 & 0.4 \end{bmatrix} \approx \begin{bmatrix} 0.3685 & 0.4106 \\ 0.4106 & 0.4632 \end{bmatrix}$$

$$M = \Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX} \approx \begin{bmatrix} 2.7778 & -2.2222 \\ -2.2222 & 2.7778 \end{bmatrix} \begin{bmatrix} 0.3685 & 0.4106 \\ 0.4106 & 0.4632 \end{bmatrix} \approx \begin{bmatrix} 0.1110 & 0.1093 \\ 0.3205 & 0.3734 \end{bmatrix}$$

Solve $\det(M - \lambda I) = 0$:

$$\begin{vmatrix} 0.1110 - \lambda & 0.1093 \\ 0.3205 & 0.3734 - \lambda \end{vmatrix} = 0$$

$$(0.1110 - \lambda)(0.3734 - \lambda) - (0.1093)(0.3205) = 0$$

$$\lambda^2 - 0.4844\lambda + 0.0064 = 0$$

Using quadratic formula:

$$\lambda = \frac{0.4844 \pm \sqrt{(0.4844)^2 - 4(0.0064)}}{2} \approx \frac{0.4844 \pm 0.4574}{2}$$

$$\lambda_1 \approx 0.4708, \quad \lambda_2 \approx 0.0137$$

$$\sigma_1 = \sqrt{\lambda_1} \approx \sqrt{0.4708} \approx 0.686$$

$$\sigma_2 = \sqrt{\lambda_2} \approx \sqrt{0.0137} \approx 0.117$$

The canonical correlations are:

$$\boxed{0.69 \quad \text{and} \quad 0.12}$$

Exercise 6.6.8

Show that canonical correlation analysis (CCA) reduces to finding the Pearson correlation coefficient when $p = q = 1$.

Solution

For scalar X and Y, the canonical correlation problem seeks scalars $a \in \mathbb{R}$ and $b \in \mathbb{R}$ to maximize the correlation between aX and bY:

$$\text{Corr}(aX, bY) = \frac{\mathbb{E}[aX \cdot bY] - \mathbb{E}[aX]\mathbb{E}[bY]}{\sqrt{\text{Var}(aX)\text{Var}(bY)}}.$$

Thus,

$$\text{Corr}(aX, bY) = \frac{ab \cdot \text{Cov}[XY]}{\sqrt{a^2\sigma_X^2 \cdot b^2\sigma_Y^2}} = \frac{ab \cdot \sigma_{XY}}{|a|\sigma_X \cdot |b|\sigma_Y}.$$

Substitute $\sigma_{XY} = \rho\sigma_X\sigma_Y$:

$$\text{Corr}(aX, bY) = \frac{ab \cdot \rho\sigma_X\sigma_Y}{|a|\sigma_X \cdot |b|\sigma_Y} = \frac{ab}{|ab|}\rho = \text{sign}(ab) \cdot \rho.$$

To maximize this correlation, choose $\text{sign}(ab) = \text{sign}(\rho)$, giving:

$$\max_{a,b} \text{Corr}(aX, bY) = |\rho|.$$

Thus, $\sigma = |\rho|$.