Tarea 1 - Análisis Multivariado

Fabián Ramírez

Vídeo de comentarios

Dejo un vídeo con comentarios que considero importante de destacar de la tarea

Problemas

Suponga $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$ vectores aleatorios IID desde $\mathcal{N}_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$ y considere:

1. Si $\boldsymbol{\mu}=\boldsymbol{\mu}_0$ conocido. Entonces:

$$\widehat{\boldsymbol{\Sigma}}=\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu}_0)(\boldsymbol{x}_i-\boldsymbol{\mu}_0)^\top$$

Demostración:

Note que si $\boldsymbol{\mu}=\boldsymbol{\mu}_0$ conocido entonces la función de densidad conjunta adopta la siguiente forma:

$$f(\boldsymbol{x};\boldsymbol{\Sigma})=|2\pi\boldsymbol{\Sigma}|^{-n/2}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu}_0)^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}_0)\right\}$$

la verosimilitud viene dada por:

$$L(\boldsymbol{\Sigma})=(2\pi)^{-np/2}|\boldsymbol{\Sigma}|^{-n/2}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu}_0)^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}_0)\right\}$$

Y la log-verosimilitud por:

$$\ell(\boldsymbol{\Sigma})=-\frac{np}{2}\log(2\pi)-\frac{n}{2}\log|\boldsymbol{\Sigma}|-\frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu}_0)(\boldsymbol{x}_i-\boldsymbol{\mu}_0)^\top\right)$$

Note que debemos derivar ℓ con respecto a una matriz, por tanto:

$$\begin{aligned}
\mathrm{d}\ell(\boldsymbol{\Sigma}) &= \mathrm{d}\left[-\frac{np}{2}\log(2\pi) - \frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}\left(\boldsymbol{x}_i-\boldsymbol{\mu}_0\right)\left(\boldsymbol{x}_i-\boldsymbol{\mu}_0\right)^\top\right)\right] \\
&= -\frac{n}{2}\mathrm{d}\left[\log|\boldsymbol{\Sigma}|\right] - \frac{1}{2}\mathrm{d}\left[\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}_0\right)\left(\boldsymbol{x}-\boldsymbol{\mu}_0\right)^\top\right)\right] \\
&= -\frac{n}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\mathrm{d}\boldsymbol{\Sigma}\right) - \frac{1}{2}\operatorname{tr}\left[\mathrm{d}\left\{\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}\left(\boldsymbol{x}_i-\boldsymbol{\mu}_0\right)\left(\boldsymbol{x}_i-\boldsymbol{\mu}_0\right)^\top\right\}\right] \\
&= -\frac{n}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\mathrm{d}\boldsymbol{\Sigma}\right) - \frac{1}{2}\operatorname{tr}\left[-\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}\left(\boldsymbol{x}_i-\boldsymbol{\mu}_0\right)\left(\boldsymbol{x}_i-\boldsymbol{\mu}_0\right)^\top\right] \\
&= -\frac{1}{2}\operatorname{tr}\left(n\boldsymbol{\Sigma}^{-1}\mathrm{d}\boldsymbol{\Sigma}\right) + \frac{1}{2}\operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}\left(\boldsymbol{x}_i-\boldsymbol{\mu}_0\right)\left(\boldsymbol{x}_i-\boldsymbol{\mu}_0\right)^\top\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\right] \\
&= \frac{1}{2}\operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}\left(\boldsymbol{x}_i-\boldsymbol{\mu}_0\right)\left(\boldsymbol{x}_i-\boldsymbol{\mu}_0\right)^\top\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma}) - n\boldsymbol{\Sigma}^{-\mathbf{1}}(\mathrm{d}\boldsymbol{\Sigma})\right] \\
&= \frac{1}{2}\operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\left\{\sum_{i=1}^{n}\left(\boldsymbol{x}_i-\boldsymbol{\mu}_0\right)\left(\boldsymbol{x}_i-\boldsymbol{\mu}_0\right)^\top\boldsymbol{\Sigma}^{-1} - n\right\}(\mathrm{d}\boldsymbol{\Sigma})\right] \\
&= \frac{1}{2}\operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\left\{\sum_{i=1}^{n}\left(\boldsymbol{x}_i-\boldsymbol{\mu}_0\right)\left(\boldsymbol{x}_i-\boldsymbol{\mu}_0\right)^\top - n\boldsymbol{\Sigma}\right\}\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\right]
\end{aligned}$$

Finalmente si $\mathrm{d}\ell(\boldsymbol{\Sigma}) = 0$ se tiene que:

$$\begin{aligned}
\mathrm{d}\ell(\boldsymbol{\Sigma}) = 0 &\iff \frac{1}{2}\operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\left\{\sum_{i=1}^{n}\left(\boldsymbol{x}_i-\boldsymbol{\mu}_0\right)\left(\boldsymbol{x}_i-\boldsymbol{\mu}_0\right)^\top - n\boldsymbol{\Sigma}\right\}\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\right] = 0 \\
&\iff \boldsymbol{\Sigma}^{-1}\left\{\sum_{i=1}^{n}\left(\boldsymbol{x}_i-\boldsymbol{\mu}_0\right)\left(\boldsymbol{x}_i-\boldsymbol{\mu}_0\right)^\top - n\boldsymbol{\Sigma}\right\}\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma}) = 0 \\
&\iff \sum_{i=1}^{n}\left(\boldsymbol{x}_i-\boldsymbol{\mu}_0\right)\left(\boldsymbol{x}_i-\boldsymbol{\mu}_0\right)^\top - n\boldsymbol{\Sigma} = 0 \\
&\iff \boldsymbol{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{x}_i-\boldsymbol{\mu}_0\right)\left(\boldsymbol{x}_i-\boldsymbol{\mu}_0\right)^\top
\end{aligned}$$

Por tanto el estimador máximo verosímil para $\boldsymbol{\Sigma}$ para la distribución normal multivariada con μ conocido viene dado por:

$$\widehat{\boldsymbol{\Sigma}}_{MV} = \frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{x}_i-\boldsymbol{\mu}_0\right)\left(\boldsymbol{x}_i-\boldsymbol{\mu}_0\right)^\top$$

2. Si $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$ conocido. Entonces:

$$\widehat{\boldsymbol{\mu}} = \overline{x}$$

Demostración:

Note que si $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$ conocido entonces la función de densidad conjunta adopta la siguiente forma:

$$f(\boldsymbol{x};\boldsymbol{\mu}) = |2\pi\boldsymbol{\Sigma}_0|^{-n/2}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu})\right\}$$

la verosimilitud viene dada por:

$$L\left(\boldsymbol{\mu}\right) = (2\pi)^{-np/2}|\boldsymbol{\Sigma}_0|^{-n/2}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu})\right\}$$

Y la log-verosimilitud por:

$$\ell(\boldsymbol{\mu}) = -\frac{np}{2}\log(2\pi) - \frac{n}{2}\log|\boldsymbol{\Sigma}_0| - \frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}_0^{-1}\sum_{i=1}^{n}\left(\boldsymbol{x}_i-\boldsymbol{\mu}\right)\left(\boldsymbol{x}_i-\boldsymbol{\mu}\right)^\top\right)$$

Note que debemos derivar ℓ con respecto a una matriz, por tanto:

$$\begin{aligned}
\mathrm{d}\ell(\boldsymbol{\mu}) &= \mathrm{d}\left[-\frac{np}{2}\log(2\pi) - \frac{n}{2}\log|\boldsymbol{\Sigma}_0| - \frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}_0^{-1}\sum_{i=1}^{n}\left(\boldsymbol{x}_i-\boldsymbol{\mu}\right)\left(\boldsymbol{x}_i-\boldsymbol{\mu}\right)^\top\right)\right]\\
&= -\frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}_0^{-1}\mathrm{d}\left[\sum_{i=1}^{n}\left(\boldsymbol{x}_i-\boldsymbol{\mu}\right)\left(\boldsymbol{x}_i-\boldsymbol{\mu}\right)^\top\right]\right)\\
&= -\frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}_0^{-1}\sum_{i=1}^{n}\mathrm{d}\left\{\left(\boldsymbol{x}_i-\boldsymbol{\mu}\right)\left(\boldsymbol{x}_i-\boldsymbol{\mu}\right)^\top\right\}\right)
\end{aligned}$$

Note que:

$$\begin{aligned}
\mathrm{d}\left\{\left(\boldsymbol{x}_i-\boldsymbol{\mu}\right)\left(\boldsymbol{x}_i-\boldsymbol{\mu}\right)^\top\right\} &= \mathrm{d}\left[\left(\boldsymbol{x}_i-\boldsymbol{\mu}\right)\right]\left(\boldsymbol{x}_i-\boldsymbol{\mu}\right)^\top + \left(\boldsymbol{x}_i-\boldsymbol{\mu}\right)\mathrm{d}\left[\left(\boldsymbol{x}_i-\boldsymbol{\mu}\right)^\top\right]\\
&= -(\mathrm{d}\boldsymbol{\mu})\left(\boldsymbol{x}_i-\boldsymbol{\mu}\right)^\top - \left(\boldsymbol{x}_i-\boldsymbol{\mu}\right)(\mathrm{d}\boldsymbol{\mu})^\top
\end{aligned}$$

Luego:

$$\begin{aligned}
\mathrm{d}\ell(\boldsymbol{\mu}) &= -\frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}_0^{-1}\sum_{i=1}^{n}\mathrm{d}\left\{\left(\boldsymbol{x}_i-\boldsymbol{\mu}\right)\left(\boldsymbol{x}_i-\boldsymbol{\mu}\right)^\top\right\}\right)\\
&= \frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}_0^{-1}\sum_{i=1}^{n}\left[(\mathrm{d}\boldsymbol{\mu})\left(\boldsymbol{x}_i-\boldsymbol{\mu}\right)^\top + \left(\boldsymbol{x}_i-\boldsymbol{\mu}\right)(\mathrm{d}\boldsymbol{\mu})^\top\right]\right)\\
&= \frac{n}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}_0^{-1}\left[(\mathrm{d}\boldsymbol{\mu})\left(\overline{\boldsymbol{x}}-\boldsymbol{\mu}\right)^\top + \left(\overline{\boldsymbol{x}}-\boldsymbol{\mu}\right)(\mathrm{d}\boldsymbol{\mu})^\top\right]\right)\\
&= \frac{n}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}_0^{-1}(\mathrm{d}\boldsymbol{\mu})\left(\overline{\boldsymbol{x}}-\boldsymbol{\mu}\right)^\top\right) + \frac{n}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}_0^{-1}\left(\overline{\boldsymbol{x}}-\boldsymbol{\mu}\right)(\mathrm{d}\boldsymbol{\mu})^\top\right)\\
&= \frac{n}{2}\operatorname{tr}\left(\left(\overline{\boldsymbol{x}}-\boldsymbol{\mu}\right)^\top\boldsymbol{\Sigma}_0^{-1}(\mathrm{d}\boldsymbol{\mu})\right) + \frac{n}{2}\operatorname{tr}\left((\mathrm{d}\boldsymbol{\mu})^\top\boldsymbol{\Sigma}_0^{-1}\left(\overline{\boldsymbol{x}}-\boldsymbol{\mu}\right)\right)\\
&= \frac{n}{2}\operatorname{tr}\left((\mathrm{d}\boldsymbol{\mu})^\top\boldsymbol{\Sigma}_0^{-\top}\right)\left(\overline{\boldsymbol{x}}-\boldsymbol{\mu}\right) + \frac{n}{2}\operatorname{tr}\left((\mathrm{d}\boldsymbol{\mu})^\top\boldsymbol{\Sigma}_0^{-1}\left(\overline{\boldsymbol{x}}-\boldsymbol{\mu}\right)\right), \quad \text{como } \boldsymbol{\Sigma}_0^{-\top} = \boldsymbol{\Sigma}_0^{-1}\\
&= n(\mathrm{d}\boldsymbol{\mu})^\top\boldsymbol{\Sigma}_0^{-1}\left(\overline{\boldsymbol{x}}-\boldsymbol{\mu}\right)
\end{aligned}$$

Finalmente si $\mathrm{d}\ell(\boldsymbol{\mu}) = 0$ se tiene que:

$$\begin{aligned}
\mathrm{d}\ell(\boldsymbol{\mu}) = 0 &\iff n(\mathrm{d}\boldsymbol{\mu})^\top\boldsymbol{\Sigma}_0^{-1}\left(\overline{\boldsymbol{x}}-\boldsymbol{\mu}\right) = 0\\
&\iff \boldsymbol{\mu} = \overline{\boldsymbol{x}}
\end{aligned}$$

Por tanto el estimador máximo verosímil para $\boldsymbol{\mu}$ para la distribución normal multivariada con $\boldsymbol{\Sigma}_0$ conocido viene dado por:

$$\widehat{\boldsymbol{\mu}}_{MV} = \overline{\boldsymbol{x}}$$

3. Sea $\boldsymbol{\mu} = \gamma\boldsymbol{a}$, $\gamma \in \mathbb{R}$ y con $\boldsymbol{a} \in \mathbb{R}^p$ conocido. Entonces,

a. Si $\boldsymbol{\Sigma}$ es conocido:

$$\widehat{\gamma}_{\boldsymbol{\Sigma}} = \frac{\boldsymbol{a}^\top\boldsymbol{\Sigma}^{-1}\overline{\boldsymbol{x}}}{\boldsymbol{a}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}}$$

b. Si $\boldsymbol{\Sigma}$ es desconocido:

$$\widehat{\gamma} = \frac{\boldsymbol{a}^\top \boldsymbol{S}^{-1}\overline{\boldsymbol{x}}}{\boldsymbol{a}^\top \boldsymbol{S}^{-1}\boldsymbol{a}}$$

Demostración:

a. Note que la función de log-verosimilitud adopta la siguiente forma:

$$\ell(\gamma) = -\frac{np}{2}\log(2\pi) - \frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}(\boldsymbol{x}_i - \gamma\boldsymbol{a})(\boldsymbol{x}_i - \gamma\boldsymbol{a})^\top\right)$$

Derivando con respecto a γ tenemos que:

$$\mathrm{d}\ell(\gamma) = -\frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}\mathrm{d}\left[(\boldsymbol{x}_i - \gamma\boldsymbol{a})(\boldsymbol{x}_i - \gamma\boldsymbol{a})^\top\right]\right)$$

Note que:

$$\begin{aligned}
\mathrm{d}\left\{(\boldsymbol{x}_i - \gamma\boldsymbol{a})(\boldsymbol{x}_i - \gamma\boldsymbol{a})^\top\right\} &= \mathrm{d}\left[(\boldsymbol{x}_i - \gamma\boldsymbol{a})\right](\boldsymbol{x}_i - \gamma\boldsymbol{a})^\top + (\boldsymbol{x}_i - \gamma\boldsymbol{a})\,\mathrm{d}\left[(\boldsymbol{x}_i - \gamma\boldsymbol{a})^\top\right] \\
&= -(\mathrm{d}\gamma)\boldsymbol{a}(\boldsymbol{x}_i - \gamma\boldsymbol{a})^\top - (\boldsymbol{x}_i - \gamma\boldsymbol{a})\boldsymbol{a}^\top(\mathrm{d}\gamma)^\top
\end{aligned}$$

Luego tenemos que:

$$\begin{aligned}
\mathrm{d}\ell(\gamma) &= -\frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}\mathrm{d}\left[(\boldsymbol{x}_i - \gamma\boldsymbol{a})(\boldsymbol{x}_i - \gamma\boldsymbol{a})^\top\right]\right) \\
&= \frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}\left\{(\mathrm{d}\gamma)\boldsymbol{a}(\boldsymbol{x}_i - \gamma\boldsymbol{a})^\top + (\boldsymbol{x}_i - \gamma\boldsymbol{a})\boldsymbol{a}^\top(\mathrm{d}\gamma)^\top\right\}\right) \\
&= \frac{n}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}(\mathrm{d}\gamma)\boldsymbol{a}(\overline{x} - \gamma\boldsymbol{a})^\top + \boldsymbol{\Sigma}^{-1}(\overline{x} - \gamma\boldsymbol{a})\boldsymbol{a}^\top(\mathrm{d}\gamma)^\top\right) \\
&= \frac{n}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}(\mathrm{d}\gamma)\boldsymbol{a}(\overline{x} - \gamma\boldsymbol{a})^\top\right) + \frac{n}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}(\overline{x} - \gamma\boldsymbol{a})\boldsymbol{a}^\top(\mathrm{d}\gamma)^\top\right) \\
&= \frac{n}{2}\operatorname{tr}\left((\overline{x} - \gamma\boldsymbol{a})\boldsymbol{a}^\top(\mathrm{d}\gamma)^\top\boldsymbol{\Sigma}^{-\top}\right) + \frac{n}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}(\overline{x} - \gamma\boldsymbol{a})\boldsymbol{a}^\top(\mathrm{d}\gamma)^\top\right) \\
&= n\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}(\overline{x} - \gamma\boldsymbol{a})\boldsymbol{a}^\top(\mathrm{d}\gamma)^\top\right) \\
&= n\operatorname{tr}\left(\boldsymbol{a}^\top(\mathrm{d}\gamma)^\top\boldsymbol{\Sigma}^{-1}(\overline{x} - \gamma\boldsymbol{a})\right) \\
&= n\left(\boldsymbol{a}^\top(\mathrm{d}\gamma)^\top\boldsymbol{\Sigma}^{-1}(\overline{x} - \gamma\boldsymbol{a})\right) \\
&= n\left(\boldsymbol{a}^\top(\mathrm{d}\gamma)^\top\boldsymbol{\Sigma}^{-1}(\overline{x} - \gamma\boldsymbol{a})\right) \\
&= n\left(\boldsymbol{a}^\top(\mathrm{d}\gamma)^\top\boldsymbol{\Sigma}^{-1}\overline{x} - \boldsymbol{a}^\top(\mathrm{d}\gamma)^\top\boldsymbol{\Sigma}^{-1}\gamma\boldsymbol{a}\right)
\end{aligned}$$

Finalmente si $\mathrm{d}\ell(\gamma) = 0$ se tiene que:

$$\begin{aligned}
\mathrm{d}\ell(\gamma) = 0 &\iff n\left(\boldsymbol{a}^\top(\mathrm{d}\gamma)^\top\boldsymbol{\Sigma}^{-1}\overline{x} - \boldsymbol{a}^\top(\mathrm{d}\gamma)^\top\boldsymbol{\Sigma}^{-1}\gamma\boldsymbol{a}\right) = 0 \\
&\iff \boldsymbol{a}^\top(\mathrm{d}\gamma)^\top\boldsymbol{\Sigma}^{-1}\overline{x} = \boldsymbol{a}^\top(\mathrm{d}\gamma)^\top\boldsymbol{\Sigma}^{-1}\gamma\boldsymbol{a} \\
&\iff \gamma = \frac{\boldsymbol{a}^\top\boldsymbol{\Sigma}^{-1}\overline{x}}{\boldsymbol{a}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}}
\end{aligned}$$

Note que esto se tiene pues $\mathrm{d}\gamma \in \mathbb{R}^{1\times 1}$. Finalmente el estimador de γ para $\boldsymbol{\Sigma}$ conocido viene dado por:

$$\widehat{\gamma}_{MV} = \frac{\boldsymbol{a}^\top\boldsymbol{\Sigma}^{-1}\overline{x}}{\boldsymbol{a}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}}$$

b. Sea $\boldsymbol{\theta} = (\gamma, \boldsymbol{\Sigma})$. Buscamos el estimador máximo verosímil para $\boldsymbol{\theta}$, note que la función de log-verosimilitud ahora tiene la siguiente estructura:

$$\ell(\gamma, \Sigma) = -\frac{np}{2}\log(2\pi) - \frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}(\boldsymbol{x}_i - \gamma\boldsymbol{a})(\boldsymbol{x}_i - \gamma\boldsymbol{a})^\top\right)$$

Note que la ecuación $\mathrm{d}\ell_\gamma = 0$ tiene como solución $\gamma = \dfrac{\boldsymbol{a}^\top\boldsymbol{\Sigma}^{-1}\overline{x}}{\boldsymbol{a}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{a}}$. Resta encontrar la solución de $\mathrm{d}\ell_{\boldsymbol{\Sigma}}$, pero usando **1.** tenemos el óptimo que es $\boldsymbol{S}_{MV}$ y reemplazando en la estimación de γ tenemos que:

$$\widehat{\gamma}_{MV} = \frac{\boldsymbol{a}^\top\boldsymbol{S}_{MV}^{-1}\overline{x}}{\boldsymbol{a}^\top\boldsymbol{S}_{MV}^{-1}\boldsymbol{a}}$$

4. Sea $\boldsymbol{A}\boldsymbol{\mu} = \boldsymbol{a}$, $\boldsymbol{A} \in \mathbb{R}^{q\times p}$, $\boldsymbol{a} \in \mathbb{R}^q$ matrices conocidas. Luego:

a. Si $\boldsymbol{\Sigma}$ es conocido:

$$\widehat{\boldsymbol{\mu}}_\Sigma = \overline{\boldsymbol{x}} - \boldsymbol{\Sigma}\boldsymbol{A}^\top\left(\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^\top\right)^{-1}(\boldsymbol{A}\overline{\boldsymbol{x}} - \boldsymbol{a})$$

b. Si $\boldsymbol{\Sigma}$ es desconocido:

$$\widehat{\boldsymbol{\mu}} = \overline{\boldsymbol{x}} - \boldsymbol{S}\boldsymbol{A}^\top\left(\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^\top\right)^{-1}(\boldsymbol{A}\overline{\boldsymbol{x}} - \boldsymbol{a})$$

Demostración:

a. Note que queremos optimizar la función:

$$\ell(\boldsymbol{\mu}) = -\frac{np}{2}\log(2\pi) - \frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^\top\right)$$

Sujeto a:

$$\boldsymbol{A}\boldsymbol{\mu} = \boldsymbol{a} \Longleftrightarrow \boldsymbol{\mu}^\top\boldsymbol{A}^T = \boldsymbol{a}^T$$

Por tanto para encontrar el óptimo debemos optimizar mediante multiplicadores de Lagrange. Definamos el Lagrangiano:

$$\mathscr{L}(\boldsymbol{\mu}) = -\frac{np}{2}\log(2\pi) - \frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^\top\right) + \lambda\left[\boldsymbol{\mu}^\top\boldsymbol{A}^T\right]$$

Luego:

$$\begin{aligned}
\mathrm{d}\mathscr{L}(\boldsymbol{\mu}) &= n(\mathrm{d}\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}) + \lambda(\mathrm{d}\boldsymbol{\mu})^\top\boldsymbol{A}^\top \\
&= n(\mathrm{d}\boldsymbol{\mu})^\top\left[\boldsymbol{\Sigma}^{-1}\overline{\boldsymbol{x}} - \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right] + \lambda(\mathrm{d}\boldsymbol{\mu})^\top\boldsymbol{A}^\top \\
&= (\mathrm{d}\boldsymbol{\mu})^\top\left\{n\left(\left[\boldsymbol{\Sigma}^{-1}\overline{\boldsymbol{x}} - \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right]\right) + \lambda\boldsymbol{A}^\top\right\} \\
&= (\mathrm{d}\boldsymbol{\mu})^\top\left\{n\boldsymbol{\Sigma}^{-1}\overline{\boldsymbol{x}} + \boldsymbol{A}^\top\lambda - n\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right\} \\
&= \boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\mu})^\top n\left\{\overline{\boldsymbol{x}} + \frac{\lambda}{n}\boldsymbol{\Sigma}\boldsymbol{A}^\top - \boldsymbol{\mu}\right\}
\end{aligned}$$

Igualando a 0 con la finalidad de obtener el óptimo tenemos que:

$$
\begin{aligned}
\mathrm{d}\mathscr{L}(\boldsymbol{\mu}) = 0 &\Longleftrightarrow \boldsymbol{\Sigma}^{-1}\left(\mathrm{d}\boldsymbol{\mu}\right)^{\top} n\left\{\overline{\boldsymbol{x}} + \frac{\lambda}{n}\boldsymbol{\Sigma}\boldsymbol{A}^{\top} - \boldsymbol{\mu}\right\} = 0 \\
&\Longleftrightarrow \overline{\boldsymbol{x}} + \frac{\lambda}{n}\boldsymbol{\Sigma}\boldsymbol{A}^{\top} - \boldsymbol{\mu} = 0 \quad (\bigstar) \\
&\Longleftrightarrow \boldsymbol{\mu} = \overline{\boldsymbol{x}} + \frac{\lambda}{n}\boldsymbol{\Sigma}\boldsymbol{A}^{\top} \\
&\Longleftrightarrow \boldsymbol{A}\boldsymbol{\mu} = \boldsymbol{A}\overline{\boldsymbol{x}} + \frac{\lambda}{n}\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^{\top} \\
&\Longleftrightarrow n\left(\boldsymbol{a} - \boldsymbol{A}\overline{\boldsymbol{x}}\right) = \lambda\left(\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^{\top}\right) \\
&\Longleftrightarrow \lambda = n\left(\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^{\top}\right)^{-1}\left(\boldsymbol{a} - \boldsymbol{A}\overline{\boldsymbol{x}}\right) \\
&\Longleftrightarrow \lambda = n\left(\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^{\top}\right)^{-1}\left(\boldsymbol{a} - \boldsymbol{A}\overline{\boldsymbol{x}}\right)
\end{aligned}
$$

Reemplazando en $(\bigstar)$ tenemos que:

$$
\overline{\boldsymbol{x}} + \frac{\lambda}{n}\boldsymbol{\Sigma}\boldsymbol{A}^{\top} - \boldsymbol{\mu} = 0 \Longleftrightarrow \boldsymbol{\mu} = \overline{\boldsymbol{x}} + \boldsymbol{\Sigma}\boldsymbol{A}^{\top}\left(\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^{\top}\right)^{-1}\left(\boldsymbol{a} - \boldsymbol{A}\overline{\boldsymbol{x}}\right)
$$

Por tanto $\widehat{\boldsymbol{\mu}}_{MV} = \overline{\boldsymbol{x}} + \boldsymbol{\Sigma}\boldsymbol{A}^{\top}\left(\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^{\top}\right)^{-1}\left(\boldsymbol{a} - \boldsymbol{A}\overline{\boldsymbol{x}}\right)$.

b. Note que el Lagrangiano viene dado por:

$$
\mathscr{L}(\boldsymbol{\mu},\boldsymbol{\Sigma}) = -\frac{np}{2}\log(2\pi) - \frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}\left(\boldsymbol{x}_i - \boldsymbol{\mu}\right)\left(\boldsymbol{x}_i - \boldsymbol{\mu}\right)^{\top}\right) + \lambda\left[\boldsymbol{\mu}^{\top}\boldsymbol{A}^{T}\right]
$$

Note que al derivar el Lagrangiano con respecto a $\boldsymbol{\Sigma}$ e igualarlo a 0 se obtiene el mismo sistema que en el problema 1. por tanto se obtiene de forma directa que $\widehat{\boldsymbol{\Sigma}} = \boldsymbol{S}_{MV}$. Además si derivamos el Lagrangiano con respecto a $\boldsymbol{\mu}$ obtenemos el mismo sistema que en $4a.$ por tanto reemplazando el estimador de $\boldsymbol{\Sigma}$ en esta expresión obtenemos que:

$$
\widehat{\mu}_{MV} = \overline{\boldsymbol{x}} + \boldsymbol{S}_{MV}\boldsymbol{A}^{\top}\left(\boldsymbol{A}\boldsymbol{S}_{MV}\boldsymbol{A}^{\top}\right)^{-1}\left(\boldsymbol{a} - \boldsymbol{A}\overline{\boldsymbol{x}}\right)
$$

5. Suponga $\boldsymbol{\Sigma} = \phi\boldsymbol{V}$ con $\boldsymbol{V} > 0$ conocida y $\phi > 0$. Por tanto:

$$
\widehat{\boldsymbol{\mu}} = \overline{\boldsymbol{x}}, \quad \widehat{\phi} = \frac{1}{p}\operatorname{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{S}\right)
$$

Demostración:

Note que la función de log-verosimilitud adopta la forma:

$$
\ell(\boldsymbol{\mu},\phi) = -\frac{np}{2}\log(2\pi) - \frac{n}{2}\log|\phi\boldsymbol{V}| - \frac{1}{2}\operatorname{tr}\left(\left(\phi\boldsymbol{V}\right)^{-1}\sum_{i=1}^{n}\left(\boldsymbol{x}_i - \boldsymbol{\mu}\right)\left(\boldsymbol{x}_i - \boldsymbol{\mu}\right)^{\top}\right)
$$

Note que derivando ℓ con respecto a $\boldsymbol{\mu}$ y asumiendo $\boldsymbol{\Sigma} = \phi\boldsymbol{V}$ constante se tiene que:

$$
\mathrm{d}\ell_{\boldsymbol{\mu}} = n(\mathrm{d}\boldsymbol{\mu})^{\top}\left(\phi\boldsymbol{V}\right)^{-1}\left(\overline{\boldsymbol{x}} - \boldsymbol{\mu}\right)
$$

Igualando a 0 tenemos que:

$$
\begin{aligned}
\mathrm{d}\ell_{\boldsymbol{\mu}} = 0 &\Longleftrightarrow n(\mathrm{d}\boldsymbol{\mu})^{\top}\left(\phi\boldsymbol{V}\right)^{-1}\left(\overline{\boldsymbol{x}} - \boldsymbol{\mu}\right) = 0 \\
&= \overline{\boldsymbol{x}} = \boldsymbol{\mu}
\end{aligned}
$$

Note que para derivar con respecto a ϕ es conveniente notar que la log-verosimilitud se puede escribir de la siguiente forma:

$$
\begin{aligned}
\ell(\boldsymbol{\mu},\phi) &= -\frac{np}{2}\log(2\pi) - \frac{n}{2}\log|\phi\boldsymbol{V}| - \frac{1}{2}\operatorname{tr}\left((\phi\boldsymbol{V})^{-1}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})^{\top}\right) \\
&= -\frac{np}{2}\log(2\pi) - \frac{np}{2}\left(\log(\phi)\right) - \frac{n}{2}\left(\log|\boldsymbol{V}|\right) + \frac{1}{\phi}\left\{-\frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}^{-1}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})^{\top}\right)\right\}
\end{aligned}
$$

Luego derivando con respecto a ϕ tenemos que:

$$
\begin{aligned}
\mathrm{d}\ell_{\phi} &= -\frac{np}{2}\frac{1}{\phi} + \left(-\frac{1}{\phi^2}\right)\left\{-\frac{1}{2}\operatorname{tr}\left(\boldsymbol{V}^{-1}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})^{\top}\right)\right\} \\
&= -\frac{np}{2}\frac{1}{\phi} + \left(\frac{1}{2\phi^2}\right)\left\{\operatorname{tr}\left(\boldsymbol{V}^{-1}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})^{\top}\right)\right\} \\
&= -\phi\frac{np}{2\phi^2} + \left(\frac{1}{2\phi^2}\right)\left\{\operatorname{tr}\left(\boldsymbol{V}^{-1}(n)\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})^{\top}\right)\right\} \\
&= \left(\frac{np}{2\phi^2}\right)\left\{-\phi + \frac{1}{p}\operatorname{tr}\left(\boldsymbol{V}^{-1}\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})^{\top}\right)\right\}
\end{aligned}
$$

Haciendo $\mathrm{d}\ell_{\phi} = 0$ y reemplazando $\boldsymbol{\mu} = \overline{\boldsymbol{x}}$ tenemos que:

$$
\begin{aligned}
\phi &= \frac{1}{p}\operatorname{tr}\left(\boldsymbol{V}^{-1}\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})^{\top}\right) \\
&= \frac{1}{p}\operatorname{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{S}_{MV}\right)
\end{aligned}
$$

Por lo tanto el estimador de ϕ por maxima verosimilitud es $\widehat{\phi} = \frac{1}{p}\operatorname{tr}\left(\boldsymbol{V}^{-1}\boldsymbol{S}_{MV}\right)$