Some Derivations Involving Matrix Calculus

Anthony Platanios

Gaussian & Mean Prior

$$\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\mu}_0, \Sigma_0)$$

$$\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \overset{iid}{\sim} \mathcal{N}(\boldsymbol{\mu}, \Sigma)$$

Remember that: $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma) \Rightarrow p(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu})}$

Gaussian & Mean Prior

$$\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\mu}_0, \Sigma_0)$$
$$\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \overset{iid}{\sim} \mathcal{N}(\boldsymbol{\mu}, \Sigma)$$

Write down the log-posterior $\mathcal{L}(\boldsymbol{\mu}) = \log p(\boldsymbol{\mu}|\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$:

$$p(\boldsymbol{\mu}|\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \propto p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n|\boldsymbol{\mu})p(\boldsymbol{\mu}) = p(\boldsymbol{\mu})\prod_{i=1}^{n} p(\boldsymbol{x}_i|\boldsymbol{\mu}) \Rightarrow$$

$$\mathcal{L}(\boldsymbol{\mu}) = \log p(\boldsymbol{\mu}) + \sum_{i=1}^{n} \log p(\boldsymbol{x}_i|\boldsymbol{\mu}) + \mathrm{C}$$

$$= -\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)^\top \Sigma_0^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_0) - \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}) + \mathrm{C}$$

Remember that: $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma) \Rightarrow p(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^d|\Sigma|}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$

Gaussian & Mean Prior

$$\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\mu}_0, \Sigma_0)$$
$$\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \stackrel{iid}{\sim} \mathcal{N}(\boldsymbol{\mu}, \Sigma)$$

Derive the first derivative of the log-posterior.

$$\nabla_{\boldsymbol{\mu}} \mathcal{L}(\boldsymbol{\mu}) = \frac{\partial}{\partial \boldsymbol{\mu}} \left(-\frac{1}{2} (\boldsymbol{\mu} - \boldsymbol{\mu}_0)^\top \Sigma_0^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0) - \frac{1}{2} \sum_{i=1}^{n} (\boldsymbol{x}_i - \boldsymbol{\mu})^\top \Sigma^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}) \right)$$

$$= -\Sigma_0^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0) - \frac{1}{2} \sum_{i=1}^{n} \frac{\partial}{\partial \boldsymbol{\mu}} (\boldsymbol{x}_i - \boldsymbol{\mu})^\top \Sigma^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu})$$

$$= -\Sigma_0^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0) - \Sigma^{-1} \sum_{i=1}^{n} (\boldsymbol{x}_i - \boldsymbol{\mu}) \frac{\partial}{\partial \boldsymbol{\mu}} (\boldsymbol{x}_i - \boldsymbol{\mu})$$

$$= -\Sigma_0^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0) + \Sigma^{-1} \sum_{i=1}^{n} (\boldsymbol{x}_i - \boldsymbol{\mu})$$

$$\frac{\partial \boldsymbol{x}^\top \mathrm{A} \boldsymbol{x}}{\partial \boldsymbol{x}} = (\mathrm{A} + \mathrm{A}^\top) \boldsymbol{x}$$

Gaussian & Mean Prior

$$\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\mu}_0, \Sigma_0)$$

$$\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \overset{iid}{\sim} \mathcal{N}(\boldsymbol{\mu}, \Sigma)$$

Derive the Hessian of the log-posterior.

$$\begin{aligned}
\nabla^2_{\boldsymbol{\mu}} \mathcal{L}(\boldsymbol{\mu}) &= \frac{\partial}{\partial \boldsymbol{\mu}} \left(-\Sigma_0^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_0) + \Sigma^{-1} \sum_{i=1}^{n} (\boldsymbol{x}_i - \boldsymbol{\mu}) \right) \\
&= -\Sigma_0^{-1} + \Sigma^{-1} \sum_{i=1}^{n} \frac{\partial}{\partial \boldsymbol{\mu}} (\boldsymbol{x}_i - \boldsymbol{\mu}) \\
&= -\Sigma_0^{-1} - n\Sigma^{-1}
\end{aligned}$$

Gaussian & Mean Prior

$$\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\mu}_0, \Sigma_0)$$

$$\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \overset{iid}{\sim} \mathcal{N}(\boldsymbol{\mu}, \Sigma)$$

Derive the MAP solution.

$$\nabla_{\boldsymbol{\mu}} \mathcal{L}(\boldsymbol{\mu}) = 0 \Rightarrow$$

$$-\Sigma_0^{-1}(\boldsymbol{\mu}^* - \boldsymbol{\mu}_0) + \Sigma^{-1} \sum_{i=1}^{n} (\boldsymbol{x}_i - \boldsymbol{\mu}^*) = 0 \Rightarrow$$

$$\left(\Sigma_0^{-1} + n\Sigma^{-1}\right) \boldsymbol{\mu}^* = \Sigma^{-1} \sum_{i=1}^{n} \boldsymbol{x}_i + \Sigma_0^{-1} \boldsymbol{\mu}_0 \Rightarrow$$

$$\boldsymbol{\mu}^* = \left(\Sigma_0^{-1} + n\Sigma^{-1}\right)^{-1} \left(\Sigma^{-1} \sum_{i=1}^{n} \boldsymbol{x}_i + \Sigma_0^{-1} \boldsymbol{\mu}_0\right)$$

Gaussian & Joint Prior

$$\boldsymbol{\mu}, \Sigma \sim \mathcal{N}(\boldsymbol{\mu}|\boldsymbol{\mu}_0, \beta^{-1}\Sigma)\mathcal{W}^{-1}(\Sigma|\Psi, \nu)$$
$$\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \overset{iid}{\sim} \mathcal{N}(\boldsymbol{\mu}, \Sigma)$$

$$\boldsymbol{\mu}, \Sigma \sim \mathrm{NIW}(\boldsymbol{\mu}_0, \beta, \Psi, \nu) = \mathcal{N}(\boldsymbol{\mu}|\boldsymbol{\mu}_0, \beta^{-1}\Sigma)\mathcal{W}^{-1}(\Sigma|\Psi, \nu)$$
$$\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \overset{iid}{\sim} \mathcal{N}(\boldsymbol{\mu}, \Sigma)$$

Note: $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma) \Rightarrow p(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^d|\Sigma|}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$, and:

$$\Sigma \sim \mathcal{W}^{-1}(\Psi, \nu) \Rightarrow p(\Sigma) = \frac{|\Psi|^{\nu/2}}{2^{\nu d/2}\Gamma_d\left(\frac{\nu}{2}\right)} |\Sigma|^{-(\nu+d+1)/2} e^{-\frac{1}{2}\mathrm{Trace}(\Psi\Sigma^{-1})} ,$$

where Γ_d is the multivariate Gamma function.

Gaussian & Joint Prior

$$\boldsymbol{\mu}, \Sigma \sim \mathcal{N}(\boldsymbol{\mu}|\boldsymbol{\mu}_0, \beta^{-1}\Sigma)\mathcal{W}^{-1}(\Sigma|\Psi, \nu)$$
$$\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \overset{iid}{\sim} \mathcal{N}(\boldsymbol{\mu}, \Sigma)$$

Write down the log-posterior $\mathcal{L}(\boldsymbol{\mu}, \Sigma) = \log p(\boldsymbol{\mu}, \Sigma|\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$:

Note: $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma) \Rightarrow p(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^d|\Sigma|}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$, and:

$$\Sigma \sim \mathcal{W}^{-1}(\Psi, \nu) \Rightarrow p(\Sigma) = \frac{|\Psi|^{\nu/2}}{2^{\nu d/2}\Gamma_d\left(\frac{\nu}{2}\right)}|\Sigma|^{-(\nu+d+1)/2} e^{-\frac{1}{2}\mathrm{Trace}(\Psi\Sigma^{-1})},$$

where Γ_d is the multivariate Gamma function.

Gaussian & Joint Prior

$$\boldsymbol{\mu}, \Sigma \sim \mathcal{N}(\boldsymbol{\mu}|\boldsymbol{\mu}_0, \beta^{-1}\Sigma)\mathcal{W}^{-1}(\Sigma|\Psi, \nu)$$

$$\boldsymbol{x}_1, \dots, \boldsymbol{x}_n \overset{iid}{\sim} \mathcal{N}(\boldsymbol{\mu}, \Sigma)$$

$$p(\boldsymbol{\mu}, \Sigma|\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) \propto p(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n|\boldsymbol{\mu}, \Sigma)p(\boldsymbol{\mu}|\Sigma)p(\Sigma)$$

$$\propto p(\boldsymbol{\mu}|\Sigma)p(\Sigma)\prod_{i=1}^{n} p(\boldsymbol{x}_i|\boldsymbol{\mu}) \Rightarrow$$

Note: $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma) \Rightarrow p(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^d|\Sigma|}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$, and:

$$\Sigma \sim \mathcal{W}^{-1}(\Psi, \nu) \Rightarrow p(\Sigma) = \frac{|\Psi|^{\nu/2}}{2^{\nu d/2}\Gamma_d\left(\frac{\nu}{2}\right)}|\Sigma|^{-(\nu+d+1)/2}e^{-\frac{1}{2}\mathrm{Trace}(\Psi\Sigma^{-1})},$$

where Γ_d is the multivariate Gamma function.

Gaussian & Joint Prior

$$\boldsymbol{\mu}, \Sigma \sim \mathcal{N}(\boldsymbol{\mu}|\boldsymbol{\mu}_0, \beta^{-1}\Sigma)\mathcal{W}^{-1}(\Sigma|\Psi, \nu)$$
$$\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \overset{iid}{\sim} \mathcal{N}(\boldsymbol{\mu}, \Sigma)$$

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\mu}, \Sigma) &= \log p(\boldsymbol{\mu}|\Sigma) + \log p(\Sigma) + \sum_{i=1}^{n} \log p(\boldsymbol{x}_i|\boldsymbol{\mu}, \Sigma) + \mathrm{C} \\
&= -\frac{1}{2}\log|\Sigma| - \frac{\beta}{2}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)^\top \Sigma^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_0) \\
&\quad - \frac{\nu + d + 1}{2}\log|\Sigma| - \frac{1}{2}\mathrm{Trace}(\Psi\Sigma^{-1}) \\
&\quad - \frac{n}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}) + \mathrm{C}
\end{aligned}$$

Note: $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma) \Rightarrow p(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^d|\Sigma|}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$, and:

$$\Sigma \sim \mathcal{W}^{-1}(\Psi, \nu) \Rightarrow p(\Sigma) = \frac{|\Psi|^{\nu/2}}{2^{\nu d/2}\Gamma_d\left(\frac{\nu}{2}\right)}|\Sigma|^{-(\nu+d+1)/2} e^{-\frac{1}{2}\mathrm{Trace}(\Psi\Sigma^{-1})},$$

where Γ_d is the multivariate Gamma function.

Gaussian & Joint Prior

$$\boldsymbol{\mu}, \Sigma \sim \mathcal{N}(\boldsymbol{\mu}|\boldsymbol{\mu}_0, \beta^{-1}\Sigma)\mathcal{W}^{-1}(\Sigma|\Psi, \nu)$$
$$\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \overset{iid}{\sim} \mathcal{N}(\boldsymbol{\mu}, \Sigma)$$

Derive $\nabla_{\boldsymbol{\mu}}\mathcal{L}(\boldsymbol{\mu}, \Sigma)$.

$$\begin{aligned}
\nabla_{\boldsymbol{\mu}}\mathcal{L}(\boldsymbol{\mu}, \Sigma) &= \frac{\partial}{\partial \boldsymbol{\mu}}\left(-\frac{\beta}{2}(\boldsymbol{\mu}-\boldsymbol{\mu}_0)^\top\Sigma^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_0) - \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})^\top\Sigma^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu})\right) \\
&= -\beta\Sigma^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_0) - \frac{1}{2}\sum_{i=1}^{n}\frac{\partial}{\partial \boldsymbol{\mu}}(\boldsymbol{x}_i-\boldsymbol{\mu})^\top\Sigma^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}) \\
&= -\beta\Sigma^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_0) - \Sigma^{-1}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})\frac{\partial}{\partial \boldsymbol{\mu}}(\boldsymbol{x}_i-\boldsymbol{\mu}) \\
&= -\beta\Sigma^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_0) + \Sigma^{-1}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})
\end{aligned}$$

$$\frac{\partial \boldsymbol{x}^\top \mathrm{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = (\mathrm{A}+\mathrm{A}^\top)\boldsymbol{x}$$

Gaussian & Joint Prior

$$\boldsymbol{\mu}, \Sigma \sim \mathcal{N}(\boldsymbol{\mu}|\boldsymbol{\mu}_0, \beta^{-1}\Sigma)\mathcal{W}^{-1}(\Sigma|\Psi, \nu)$$
$$\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \overset{iid}{\sim} \mathcal{N}(\boldsymbol{\mu}, \Sigma)$$

Derive $\nabla_\Sigma \mathcal{L}(\boldsymbol{\mu}, \Sigma)$.

$$\begin{aligned}\nabla_\Sigma \mathcal{L}(\boldsymbol{\mu}, \Sigma) &= -\frac{\beta}{2}\frac{\partial}{\partial\Sigma}(\boldsymbol{\mu}-\boldsymbol{\mu}_0)^\top\Sigma^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_0) - \frac{\nu+d+n+2}{2}\frac{\partial}{\partial\Sigma}\log|\Sigma| - \frac{1}{2}\frac{\partial}{\partial\Sigma}\mathrm{Trace}(\Psi\Sigma^{-1}) \\ &\quad -\frac{1}{2}\frac{\partial}{\partial\Sigma}\sum_{i=1}^n(\boldsymbol{x}_i-\boldsymbol{\mu})^\top\Sigma^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}) \\ &= -\frac{\nu+d+n+2}{2}\Sigma^{-1} + \frac{\beta}{2}\Sigma^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_0)(\boldsymbol{\mu}-\boldsymbol{\mu}_0)^\top\Sigma^{-1} + \frac{1}{2}\Sigma^{-1}\Psi\Sigma^{-1} \\ &\quad + \frac{1}{2}\Sigma^{-1}\left[\sum_{i=1}^n(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})^\top\right]\Sigma^{-1}\end{aligned}$$

$$\frac{\partial \log|\mathrm{A}|}{\partial \mathrm{A}} = \mathrm{A}^{-\top}$$
$$\frac{\partial \boldsymbol{x}^\top \mathrm{A}^{-1}\boldsymbol{y}}{\partial \mathrm{A}} = -\mathrm{A}^{-\top}\boldsymbol{x}\boldsymbol{y}^\top\mathrm{A}^{-\top}$$
$$\frac{\partial \mathrm{Trace}(\mathrm{B}\mathrm{A}^{-1}\mathrm{C})}{\partial \mathrm{A}} = -(\mathrm{A}^{-1}\mathrm{C}\mathrm{B}\mathrm{A}^{-1})^\top$$

Gaussian & Joint Prior

$$\boldsymbol{\mu}, \Sigma \sim \mathcal{N}(\boldsymbol{\mu}|\boldsymbol{\mu}_0, \beta^{-1}\Sigma)\mathcal{W}^{-1}(\Sigma|\Psi, \nu)$$
$$\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \overset{iid}{\sim} \mathcal{N}(\boldsymbol{\mu}, \Sigma)$$

Derive the MAP solution.

$$\nabla_{\boldsymbol{\mu}} \mathcal{L}(\boldsymbol{\mu}, \Sigma) = 0 \Rightarrow$$

$$-\beta\Sigma^{-1}(\boldsymbol{\mu}^* - \boldsymbol{\mu}_0) + \Sigma^{-1}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu}^*) = 0 \Rightarrow$$

$$(\beta + n)\,\boldsymbol{\mu}^* = \sum_{i=1}^{n}\boldsymbol{x}_i + \beta\boldsymbol{\mu}_0 \Rightarrow$$

$$\boldsymbol{\mu}^* = \frac{\sum_{i=1}^{n}\boldsymbol{x}_i + \beta\boldsymbol{\mu}_0}{\beta + n}$$

$$\nabla_{\Sigma} \mathcal{L}(\boldsymbol{\mu}, \Sigma) = 0 \Rightarrow$$

$$\beta{\Sigma^*}^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)(\boldsymbol{\mu} - \boldsymbol{\mu}_0)^\top + {\Sigma^*}^{-1}\Psi + {\Sigma^*}^{-1}\left[\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^\top\right] = \nu + d + n + 2 \Rightarrow$$

$$\Sigma^* = \frac{\beta(\boldsymbol{\mu} - \boldsymbol{\mu}_0)(\boldsymbol{\mu} - \boldsymbol{\mu}_0)^\top + \Psi + \sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^\top}{\nu + d + n + 2}$$

Done! :)