Longxiang Wang
12710172
wlongxiang1119@gmail.com

1 MAP solution with correlated responses

a) Write down the likelihood $p(\boldsymbol{D}|\boldsymbol{\theta})$ in matrix form

As stated in the assumption, we have:

$$p(\boldsymbol{D}|\boldsymbol{\theta}) = p(\boldsymbol{t}|\boldsymbol{\Psi}, \boldsymbol{w}, \boldsymbol{\Omega}) = \mathcal{N}(\boldsymbol{t}|\boldsymbol{\Psi w}, \boldsymbol{\Omega})$$

According to multivariate Gausssian distribution (Bishop 2.3), we can derive the final form:

$$p(\boldsymbol{D}|\boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi)^N det\boldsymbol{\Omega}}} e^{(-\frac{1}{2}(\boldsymbol{t}-\boldsymbol{\Psi w})^{\boldsymbol{T}}\boldsymbol{\Omega}^{-1}(\boldsymbol{t}-\boldsymbol{\Psi w}))}$$

b) Write down the likelihood $p(\boldsymbol{D}|\boldsymbol{\theta})$ in a Gaussian with a diagonal covariance matrix

let's first calculate the determinant of Ω. Since we have $\boldsymbol{\Omega = A^T D A}$, and $\boldsymbol{A^T = A^{-1}}$, and using the determinant rule in the hint, the determinant is:

$$\boldsymbol{det(\Omega) = det(A^T D A) = det(A)^{-1} det(D) det(A) = det(D)}$$

Using the orthogananity of $\boldsymbol{A}$, we can write the inverse of covariance matrix as :

$$\boldsymbol{\Omega^{-1} = A^T D^{-1} A}$$

Using the determinant of covariance and inverse of covariance, and the substitution, we have the likelihood as following :

$$\begin{aligned}
p(\boldsymbol{D}|\boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi)^N det\boldsymbol{\Omega}}} e^{(-\frac{1}{2}(\boldsymbol{t}-\boldsymbol{\Psi w})^{\boldsymbol{T}}\boldsymbol{\Omega}^{-1}(\boldsymbol{t}-\boldsymbol{\Psi w}))} &= \frac{1}{\sqrt{(2\pi)^N det\boldsymbol{D}}} e^{(-\frac{1}{2}(\boldsymbol{t}-\boldsymbol{\Psi w})^{\boldsymbol{T}}\boldsymbol{A^T D^{-1} A}(\boldsymbol{t}-\boldsymbol{\Psi w}))} \\
&= \frac{1}{\sqrt{(2\pi)^N det\boldsymbol{D}}} e^{(-\frac{1}{2}(\boldsymbol{At}-\boldsymbol{A\Psi w})^{\boldsymbol{T}}\boldsymbol{D}^{-1}(\boldsymbol{At}-\boldsymbol{A\Psi w}))} \\
&= \frac{1}{\sqrt{(2\pi)^N det\boldsymbol{D}}} e^{(-\frac{1}{2}(\boldsymbol{\tau}-\boldsymbol{\Phi w})^{\boldsymbol{T}}\boldsymbol{D}^{-1}(\boldsymbol{\tau}-\boldsymbol{\Phi w}))} \\
&= \mathcal{N}(\boldsymbol{\tau}|\boldsymbol{\Phi w}, \boldsymbol{D})
\end{aligned}$$

where $\boldsymbol{\tau = At}$ and $\boldsymbol{\Phi = A\Psi}$.

c) Factorize the likelihood $p(\boldsymbol{D}|\boldsymbol{\theta})$ into univariate Gaussian distribution

Since we have the determinant of a diagonal matrix $\boldsymbol{D}$ in the following form:

$$\boldsymbol{detD} = \prod_{i=1}^{N} d_i$$

Now, we can rewrite the likelihood into the following form (the simplification process is skipped because it is trivial):

$$p(\boldsymbol{D}|\boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi)^N \prod_{i=1}^{N} d_i}} e^{(-\frac{1}{2}\sum_{i=1}^{N}\frac{(\tau_i - \Phi_i^T w)^2}{d_i})}$$

Since we have $e^x \times e^y = e^{(x+y)}$, we can further simplify the likelihood into:

$$p(\boldsymbol{D}|\boldsymbol{\theta}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi d_i}} e^{(-\frac{1}{2}\frac{(\tau_i - \Phi_i^T w)^2}{d_i})} = \prod_{i=1}^{N} \mathcal{N}(\tau_i|\Phi_i^T w, d_i)$$

As we can see, now the likelihood is factored into the product of N normal distributions.

d) Write down the prior $p(w)$ in Gaussian form and compute its logarithm

Since we have $det(\boldsymbol{\alpha I}) = \alpha^N$, the prior can be simplified as following:

$$p(\boldsymbol{w}) = \mathcal{N}(\boldsymbol{w}|\boldsymbol{0}, \alpha\boldsymbol{I}) = \frac{1}{\sqrt{(2\pi)^N det(\boldsymbol{\alpha I})}} e^{(-\frac{1}{2}(\boldsymbol{w}-\boldsymbol{0})^{\boldsymbol{T}}(\alpha \boldsymbol{I})^{-1}(\boldsymbol{w}-\boldsymbol{0}))} = \frac{1}{\sqrt{(2\pi\alpha)^N}} e^{(-\frac{1}{2\alpha}\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{w})}$$

Since α is constant, with that we can derive the log of prior as following:

$$lnp(\boldsymbol{w}) = -\frac{1}{2\alpha}\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{w} - ln(\sqrt{(2\pi\alpha)^N}) = -\frac{1}{2\alpha}\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{w} + C$$

e) Write down the posterior $p(w|D)$ with evidence as probability

According to Bayes rule, the posterior is:

$$p(\boldsymbol{w}|\boldsymbol{D}) = \frac{p(\boldsymbol{D}|\boldsymbol{w})p(\boldsymbol{w})}{p(\boldsymbol{D})}$$

According to the product rule we can calculate the evidence as:

$$p(\boldsymbol{D}) = \int_w p(\boldsymbol{D}|\boldsymbol{w})p(\boldsymbol{w})dw$$

As we already calculated before, we can plug in the prior and likelihood, then we have:

$$p(\boldsymbol{w}|\boldsymbol{D}) = \frac{\mathcal{N}(\boldsymbol{\tau}|\boldsymbol{\Phi w}, \boldsymbol{D})\mathcal{N}(\boldsymbol{w}|\boldsymbol{0}, \alpha\boldsymbol{I})}{\int_w \mathcal{N}(\boldsymbol{\tau}|\boldsymbol{\Phi w}, \boldsymbol{D})\mathcal{N}(\boldsymbol{w}|\boldsymbol{0}, \alpha\boldsymbol{I})dw} = \frac{\mathcal{N}(\boldsymbol{\tau}|\boldsymbol{\Phi w}, \boldsymbol{D})\mathcal{N}(\boldsymbol{w}|\boldsymbol{0}, \boldsymbol{\alpha I})}{p(\boldsymbol{\tau}|\boldsymbol{\Phi}, \boldsymbol{D}, \alpha)}$$

f) Computer the log posterior $lnp(w|D)$ in matrix form and factorized form

From ***d)***, we already calculated the log of prior:

$$ln\mathcal{N}(\boldsymbol{w}|\boldsymbol{0}, \alpha\boldsymbol{I}) = -\frac{1}{2\alpha}\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{w} + C$$

From ***b)***, we already cacluated the matrix form of likelihood:

$$\mathcal{N}(\boldsymbol{\tau}|\boldsymbol{\Phi w}, \boldsymbol{D}) = \frac{1}{\sqrt{(2\pi)^N det\boldsymbol{D}}} e^{(-\frac{1}{2}(\boldsymbol{\tau}-\boldsymbol{\Phi w})^{\boldsymbol{T}}\boldsymbol{D}^{-1}(\boldsymbol{\tau}-\boldsymbol{\Phi w}))}$$

$$ln\mathcal{N}(\boldsymbol{\tau}|\boldsymbol{\Phi w}, \boldsymbol{D}) = -\frac{1}{2}(\boldsymbol{\tau} - \boldsymbol{\Phi w})^{\boldsymbol{T}}\boldsymbol{D}^{-1}(\boldsymbol{\tau} - \boldsymbol{\Phi w}) + C$$

Based on the solution from ***e)***, evidence is independent of $\boldsymbol{w}$, we can make $lnp(\boldsymbol{\tau}|\boldsymbol{\Phi}, \boldsymbol{D}, \alpha)$ part of a constant C, we have the matrix form of log posterior:

$$\begin{aligned} lnp(\boldsymbol{w}|\boldsymbol{D}) &= ln\mathcal{N}(\boldsymbol{\tau}|\boldsymbol{\Phi w}, \boldsymbol{D}) + ln\mathcal{N}(\boldsymbol{w}|\boldsymbol{0}, \alpha\boldsymbol{I}) - lnp(\boldsymbol{\tau}|\boldsymbol{\Phi}, \boldsymbol{D}, \alpha) \\ &= -\frac{1}{2\alpha}\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{w} - \frac{1}{2}(\boldsymbol{\tau} - \boldsymbol{\Phi w})^{\boldsymbol{T}}\boldsymbol{D}^{-1}(\boldsymbol{\tau} - \boldsymbol{\Phi w}) + C \end{aligned}$$

where $\boldsymbol{\tau} = \boldsymbol{At}$, $\boldsymbol{\Phi} = \boldsymbol{A\Psi}$ and $\boldsymbol{D} = diag([d_1, d_2, ..., d_N])$.
From ***c)***, we already calculated the likelihood in factorized form:

$$\mathcal{N}(\boldsymbol{\tau}|\boldsymbol{\Phi w}, \boldsymbol{D}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi d_i}} e^{(-\frac{1}{2}\frac{(\tau_i - \Phi_i^T w)^2}{d_i})}$$

$$ln\mathcal{N}(\boldsymbol{\tau}|\boldsymbol{\Phi w}, \boldsymbol{D}) = -\frac{1}{2}\sum_{i=1}^{N}\frac{(\tau_i - \Phi_i^T w)^2}{d_i} + C$$

Plugging all the components together we have the final form of log posterior in factorized form:

$$lnp(\boldsymbol{w}|\boldsymbol{D}) = -\frac{1}{2\alpha}\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{w} - \frac{1}{2}\sum_{i=1}^{N}\frac{(\tau_i - \Phi_i^T w)^2}{d_i} + C$$

Why MAP is easier than full posterior?
To compute the evidence, we need to integral over w, which is difficult in most cases. And since evidence is independent of w, thus MAP solution is easier than calculating the full posterior distribution because MAP could discard the constant. In other words, MAP solution doesn't need to estimate the normalization constant, but for full posterior distribution, expensive integral needs to be done for all weights.

g) Solving for $\boldsymbol{w}_{MAP}$ by taking derivative of log-posterior

Based on solution from ***f***), we know that

$$lnp(\boldsymbol{w}|\boldsymbol{D}) = -\frac{1}{2\alpha}\boldsymbol{w^T w} - \frac{1}{2}\sum_{i=1}^{N}\frac{(\tau_i - \Phi_i^T w)^2}{d_i} + C = -\frac{1}{2\alpha}\boldsymbol{w^T w} - \frac{1}{2}\boldsymbol{(\tau - \Phi w)^T D^{-1}(\tau - \Phi w)} + C$$

By setting the derivative to 0 we have:

$$\begin{aligned}
\frac{d}{d\mathbf{w}}lnp(\boldsymbol{w}|\boldsymbol{D}) &= -\frac{1}{2\alpha}\frac{d}{d\mathbf{w}}\boldsymbol{w^T w} - \frac{1}{2}\frac{d}{d\mathbf{w}}\boldsymbol{(\tau - \Phi w)^T D^{-1}(\tau - \Phi w)} \\
&= -\frac{1}{\alpha}\boldsymbol{w^T} - \frac{1}{2}\frac{d}{d\mathbf{w}}\boldsymbol{(\tau^T D^{-1}\tau - \tau^T D^{-1}\Phi w - w^T\Phi^T D^{-1}\tau + w^T\Phi^T D^{-1}\Phi w)} \\
&= -\frac{1}{\alpha}\boldsymbol{w^T} + \boldsymbol{(\tau^T D^{-1}\Phi - w^T\Phi^T D^{-1}\Phi)} \\
&= 0
\end{aligned}$$

$$\frac{1}{\alpha}\boldsymbol{w^T} = \boldsymbol{(\tau^T D^{-1}\Phi - w^T\Phi^T D^{-1}\Phi)} \xRightarrow{transpose\quad on\quad both\quad side} \frac{1}{\alpha}\boldsymbol{w} = \boldsymbol{(\Phi^T D^{-1}\tau - \Phi^T D^{-1}\Phi w)}$$

$$\boldsymbol{w}_{MAP} = \boldsymbol{(\alpha I + \Phi^T D^{-1}\Phi)^{-1}\Phi^T D^{-1}\tau}$$

h) Express the solution for $\boldsymbol{w}_{MAP}$ in terms of original quantities $\boldsymbol{t}, \boldsymbol{\Psi}$

Substitute $\boldsymbol{\tau = At}$, $\boldsymbol{\Phi = A\Psi}$ into the solution:

$$\begin{aligned}
\boldsymbol{w}_{MAP} &= \boldsymbol{(\alpha I + \Phi^T D^{-1}\Phi)^{-1}\Phi^T D^{-1}\tau} \\
&= \boldsymbol{(\alpha I + (A\Psi)^T D^{-1}(A\Psi))^{-1}(A\Psi)^T D^{-1}At} \\
&= \boldsymbol{(\alpha I + \Psi^T(A^T D^{-1}A)\Psi)^{-1}\Psi^T(A^T D^{-1}A)t} \\
&= \boldsymbol{(\alpha I + \Psi^T\Omega^{-1}\Psi)^{-1}\Psi^T\Omega^{-1}t}
\end{aligned}$$

2 ML estimate of angle measurements

The conditional probability of both c and s are given as a univariate Gaussian distribution:

$$p(c|\theta) = \mathcal{N}(c|cos\theta, \sigma^2)$$
$$p(s|\theta) = \mathcal{N}(s|sin\theta, \sigma^2)$$

Therefore, the maximal likelihood is:

$$\begin{aligned}
p(\boldsymbol{D}|\theta) &= p(c|\theta)p(s|\theta) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}}exp(-\frac{(c - cos(\theta))^2}{2\sigma^2}) \cdot \frac{1}{\sqrt{2\pi\sigma^2}}exp(-\frac{(s - sin(\theta))^2}{2\sigma^2}) \\
&= \frac{1}{2\pi\sigma^2}exp(-\frac{(c - cos(\theta))^2 + (s - sin(\theta))^2}{2\sigma^2}) \\
&= \frac{1}{2\pi\sigma^2}exp(-\frac{c^2 + s^2 + 1 - 2ccos(\theta) - 2ssin(\theta)}{2\sigma^2})
\end{aligned}$$

Now, we can find θ by finding the point where derivative of log-likelihood is zero:

$$\begin{aligned}
\frac{d}{d\theta}lnp(\boldsymbol{D}|\theta) &= \frac{d}{d\theta}(-\frac{c^2 + s^2 + 1 - 2ccos(\theta) - 2ssin(\theta)}{2\sigma^2}) \\
&= \frac{d}{d\theta}\frac{ccos(\theta) + ssin(\theta)}{\sigma^2} \\
&= \frac{-csin(\theta) + scos(\theta)}{\sigma^2} = 0 \longrightarrow tan(\theta) = s/c \longrightarrow \theta_{ML} = arctan(s/c)
\end{aligned}$$

3 ML and MAP solution for a Poisson distribution fit

a) Calculate Poisson distribution λ using MLE

Poisson distribution is given by:

$$p(x) = exp(-\lambda)\frac{\lambda^x}{x!}$$

Under i.i.d., the likelihood is given by:

$$\begin{aligned} p(\boldsymbol{D}|\lambda) &= \prod_{i=1}^{N} p(x_i|\lambda) \\ &= \prod_{i=1}^{N} exp(-\lambda)\frac{\lambda^{x_i}}{x_i!} \\ &= exp(-N\lambda)\frac{\lambda^{\sum_{i=1}^{N} x_i}}{\prod_{i=1}^{N} x_i!} \end{aligned}$$

The log-likelihood is given by:

$$\begin{aligned} lnp(\boldsymbol{D}|\lambda) &= -N\lambda + ln(\lambda^{\sum_{i=1}^{N} x_i}) + C \\ &= -N\lambda + \sum_{i=1}^{N} x_i \cdot ln\lambda + C \end{aligned}$$

Setting the derivative of log-likelihood to 0 gives:

$$-N + \frac{\sum_{i=1}^{N} x_i}{\lambda} = 0 \longrightarrow \lambda_{ML} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

b)Calculate Poisson distribution λ using MAP

According to Bayes rule:

$$\begin{aligned} lnp(\lambda|\boldsymbol{D}) &= lnp(\boldsymbol{D}|\lambda) + lnp(\lambda) + C \\ &= -N\lambda + (\sum_{i=1}^{N} x_i \cdot ln\lambda) - \frac{\lambda}{\alpha} + C \end{aligned}$$

Setting the derivative of log-likelihood to 0 gives:

$$-N + \frac{\sum_{i=1}^{N} x_i}{\lambda} - \frac{1}{\alpha} = 0 \longrightarrow \lambda_{MAP} = \frac{\alpha}{\alpha N + 1}\sum_{i=1}^{N} x_i$$

c) Effect of prior value α

Given $\alpha > 0$:

$$\frac{1}{N} > \frac{\alpha}{\alpha N + 1} = \frac{1}{N + \frac{1}{\alpha}}$$

Therefore, the estimated value of λ decreases with the prior.

1. when $\alpha \longrightarrow \infty$, MAP solution is the same as MLE, this is because $p(\lambda) \propto exp(-\lambda/\alpha)$ becomes a uniform distribution, thus the effect of prior vanishes.

2. when $\alpha \longrightarrow 0$, MAP solution $\lambda_{MAP} \longrightarrow 0$. As we know from the properties of Poisson distribution, it becomes a dealta peak around zero when $\lambda \longrightarrow 0$.