

## ISD parameter updates

**Natural parameter form:**

$$\begin{aligned} \mathcal{N}(\mathbf{x}|\mu, \Sigma) &= H(\mathbf{x}) \exp(\theta^T T(\mathbf{x}) - A(\theta)) \\ H(\mathbf{x}) &= (2\pi)^{-D/2}, \quad T(\mathbf{x}) = \begin{bmatrix} \mathbf{x} \\ \mathbf{x}\mathbf{x}^T \end{bmatrix}, \quad \theta = \begin{bmatrix} \Sigma^{-1}\mu \\ -\frac{1}{2}\Sigma^{-1} \end{bmatrix}, \quad A(\theta) = \frac{1}{2}(\mu^T \Sigma^{-1} \mu + \log |\Sigma|) \end{aligned}$$

**In natural parameter form:**

$$\theta = \begin{bmatrix} \Sigma^{-1}\mu \\ -\frac{1}{2}\Sigma^{-1} \end{bmatrix}, \quad \begin{aligned} \phi &= \theta_1 = \Sigma^{-1}\mu \\ \psi &= \theta_2 = -\frac{1}{2}\Sigma^{-1} \end{aligned}$$

$$\begin{aligned} A(\theta) &= -\frac{1}{4}\phi^T \psi^{-1} \phi + \frac{1}{2} \log \left| -\frac{1}{2}\psi^{-1} \right| \\ &= -\frac{1}{4}\mu^T (\Sigma^{-1})^T \left( -\frac{1}{2}\Sigma^{-1} \right)^{-1} \Sigma^{-1}\mu + \frac{1}{2} \log \left| -\frac{1}{2} \left( -\frac{1}{2}\Sigma^{-1} \right)^{-1} \right| \\ &= -\frac{1}{4}\mu^T \Sigma^{-1} (-2\Sigma) \Sigma^{-1}\mu + \frac{1}{2} \log \left| -\frac{1}{2} (-2\Sigma) \right| \\ &= \frac{1}{2}\mu^T \Sigma^{-1}\mu + \frac{1}{2} \log |\Sigma| \quad \text{as above} \end{aligned}$$

**Partial derivatives:**

$$\begin{aligned} \frac{\partial A(\theta)}{\partial \phi} &= -\frac{1}{4} \frac{\partial}{\partial \phi} (\phi^T \psi^{-1} \phi) + \frac{1}{2} \frac{\partial}{\partial \phi} \log \left| -\frac{1}{2}\psi^{-1} \right| \\ &= -\frac{1}{2} (\psi^{-1} \phi) \\ \frac{\partial A(\theta)}{\partial \psi} &= -\frac{1}{4} \frac{\partial}{\partial \psi} (\phi^T \psi^{-1} \phi) + \frac{1}{2} \frac{\partial}{\partial \psi} \log \left| -\frac{1}{2}\psi^{-1} \right| \\ &= -\frac{1}{4} (-\psi^{-1} (\phi \phi^T) \psi^{-1}) + \frac{1}{2} \frac{\partial}{\partial \psi} \log \left( \left( -\frac{1}{2} \right)^N |\psi^{-1}| \right) \\ &= \frac{1}{4} (\psi^{-1} (\phi \phi^T) \psi^{-1}) + \frac{1}{2} \frac{\partial}{\partial \psi} \left( N \log \left( -\frac{1}{2} \right) + \log |\psi^{-1}| \right) \\ &= \frac{1}{4} (\psi^{-1} (\phi \phi^T) \psi^{-1}) + \frac{1}{2} \frac{\partial}{\partial \psi} \left( N \log \left( -\frac{1}{2} \right) - \log |\psi| \right) \\ &= \frac{1}{4} (\psi^{-1} (\phi \phi^T) \psi^{-1}) - \frac{1}{2} \frac{\partial}{\partial \psi} \log |\psi| \\ &= \frac{1}{4} (\psi^{-1} (\phi \phi^T) \psi^{-1}) - \frac{1}{2} \psi^{-1} \end{aligned}$$

**ISD log-posterior:**

$$\begin{aligned} \log p_\lambda(\mathcal{X}, \Theta) &= \text{const} + \sum_n \log p(\mathbf{x}_n | \theta_n) + \log p(\theta_n) + \frac{\lambda}{N} \sum_n \sum_{m \neq n} \log \mathcal{B}(p(\mathbf{x} | \theta_m), p(\mathbf{x} | \theta_n)) \\ \log p_\lambda(\mathcal{X}, \theta_n, \tilde{\Theta}_{/n}) &= \text{const} + \theta_n^T T(\mathbf{x}_n) - A(\theta_n) + \frac{2\lambda}{N} \sum_{m \neq n} \left( A(\tilde{\theta}_m/2 + \theta_n/2) - \frac{A(\tilde{\theta}_m)}{2} - \frac{A(\theta_n)}{2} \right) \\ \log p_\lambda(\mathcal{X}, \theta_n, \tilde{\Theta}_{/n}) &= \text{const} + \theta_n^T T(\mathbf{x}_n) - A(\theta_n) - \frac{\lambda(N-1)}{N} A(\theta_n) + \frac{2\lambda}{N} \sum_{m \neq n} A(\tilde{\theta}_m/2 + \theta_n/2) \\ \log p_\lambda(\mathcal{X}, \theta_n, \tilde{\Theta}_{/n}) &= \text{const} + \theta_n^T T(\mathbf{x}_n) - \frac{N + \lambda(N-1)}{N} A(\theta_n) + \frac{2\lambda}{N} \sum_{m \neq n} A(\tilde{\theta}_m/2 + \theta_n/2) \end{aligned}$$

**Maximizing:**

$$0 = T(\mathbf{x}_n) - \frac{N + \lambda(N-1)}{N} A'(\theta_n) + \frac{2\lambda}{N} \sum_{m \neq n} \frac{1}{2} A'(\tilde{\theta}_m/2 + \theta_n/2)$$

$$A'(\theta_n) = \frac{N}{N + \lambda(N-1)} \left( T(\mathbf{x}_n) + \frac{\lambda}{N} \sum_{m \neq n} A'(\tilde{\theta}_m/2 + \theta_n/2) \right)$$

$\theta_n$  is replaced by  $\tilde{\theta}_n$  on the right-hand side of the equation to obtain the iterative update rules.

**For  $\phi$ :**

$$\begin{aligned} A'(\theta_n) &= -\frac{1}{2} (\psi_n^{-1} \phi_n) \\ &= -\frac{1}{2} (-2\Sigma_n) \Sigma_n^{-1} \mu_n \\ &= \mu_n \\ A'(\tilde{\theta}_m/2 + \tilde{\theta}_n/2) &= -\frac{1}{2} \left( (\tilde{\psi}_m/2 + \tilde{\psi}_n/2)^{-1} (\tilde{\phi}_m/2 + \tilde{\phi}_n/2) \right) \\ &= -\frac{1}{2} \left( 2 (\tilde{\psi}_m + \tilde{\psi}_n)^{-1} \frac{1}{2} (\tilde{\phi}_m + \tilde{\phi}_n) \right) \\ &= -\frac{1}{2} (\tilde{\psi}_m + \tilde{\psi}_n)^{-1} (\tilde{\phi}_m + \tilde{\phi}_n) \end{aligned}$$

This yields the update rule for  $\mu$  as follows

$$\begin{aligned} A'(\theta_n) &= \frac{N}{N + \lambda(N-1)} \left( T(\mathbf{x}_n) + \frac{\lambda}{N} \sum_{m \neq n} A'(\tilde{\theta}_m/2 + \tilde{\theta}_n/2) \right) \\ \mu_n &= \frac{N}{N + \lambda(N-1)} \left( \mathbf{x}_n + \frac{\lambda}{N} \sum_{m \neq n} \left( -\frac{1}{2} (\tilde{\psi}_m + \tilde{\psi}_n)^{-1} (\tilde{\phi}_m + \tilde{\phi}_n) \right) \right) \end{aligned}$$

**Special case:** For Gaussians with fixed covariance, i.e.  $\Sigma = I$ , we have

$$\begin{aligned} \phi &= I^{-1} \mu = \mu \\ \psi &= -\frac{1}{2} I^{-1} = -\frac{1}{2} \end{aligned}$$

Replacing values for  $\phi$  and  $\psi$ ,

$$\begin{aligned} \mu_n &= \frac{N}{N + \lambda(N-1)} \left( \mathbf{x}_n + \frac{\lambda}{N} \sum_{m \neq n} \left( -\frac{1}{2} (\tilde{\psi}_m + \tilde{\psi}_n)^{-1} (\tilde{\phi}_m + \tilde{\phi}_n) \right) \right) \\ &= \frac{N}{N + \lambda(N-1)} \left( \mathbf{x}_n + \frac{\lambda}{N} \sum_{m \neq n} \left( -\frac{1}{2} \left( -\frac{1}{2} - \frac{1}{2} \right)^{-1} (\tilde{\mu}_m + \tilde{\mu}_n) \right) \right) \\ &= \frac{N}{N + \lambda(N-1)} \left( \mathbf{x}_n + \frac{\lambda}{N} \sum_{m \neq n} \left( \frac{1}{2} (\tilde{\mu}_m + \tilde{\mu}_n) \right) \right) \\ &= \frac{N}{N + \lambda(N-1)} \left( \mathbf{x}_n + \frac{\lambda}{2N} \sum_{m \neq n} (\tilde{\mu}_m + \tilde{\mu}_n) \right) \end{aligned}$$

For  $\psi$ :

$$\begin{aligned}
A'(\theta_n) &= \frac{1}{4} (\psi_n^{-1} (\phi_n \phi_n^T) \psi_n^{-1}) - \frac{1}{2} \psi_n^{-1} \\
&= \frac{1}{4} \left( \left( -\frac{1}{2} \Sigma_n^{-1} \right)^{-1} (\Sigma_n^{-1} \mu_n (\Sigma_n^{-1} \mu_n)^T) \left( -\frac{1}{2} \Sigma_n^{-1} \right)^{-1} \right) - \frac{1}{2} \left( -\frac{1}{2} \Sigma_n^{-1} \right)^{-1} \\
&= \frac{1}{4} \left( (-2 \Sigma_n) (\Sigma_n^{-1} \mu_n \mu_n^T (\Sigma_n^{-1})^T) (-2 \Sigma_n) \right) - \frac{1}{2} (-2 \Sigma_n) \\
&= (\Sigma_n (\Sigma_n^{-1} \mu_n \mu_n^T \Sigma_n^{-1}) \Sigma_n) + \Sigma_n \\
&= \mu_n \mu_n^T + \Sigma_n \\
A'(\tilde{\theta}_m/2 + \tilde{\theta}_n/2) &= \frac{1}{4} (\tilde{\psi}_m/2 + \tilde{\psi}_n/2)^{-1} (\tilde{\phi}_m/2 + \tilde{\phi}_n/2) (\tilde{\phi}_m/2 + \tilde{\phi}_n/2)^T (\tilde{\psi}_m/2 + \tilde{\psi}_n/2)^{-1} \\
&\quad - \frac{1}{2} (\tilde{\psi}_m/2 + \tilde{\psi}_n/2)^{-1} \\
&= \frac{1}{4} \left( 2 (\tilde{\psi}_m + \tilde{\psi}_n)^{-1} \frac{1}{2} (\tilde{\phi}_m + \tilde{\phi}_n) \frac{1}{2} (\tilde{\phi}_m + \tilde{\phi}_n)^T 2 (\tilde{\psi}_m + \tilde{\psi}_n)^{-1} \right) \\
&\quad - \frac{1}{2} \left( 2 (\tilde{\psi}_m + \tilde{\psi}_n)^{-1} \right) \\
&= \frac{1}{4} (\tilde{\psi}_m + \tilde{\psi}_n)^{-1} (\tilde{\phi}_m + \tilde{\phi}_n) (\tilde{\phi}_m + \tilde{\phi}_n)^T (\tilde{\psi}_m + \tilde{\psi}_n)^{-1} - (\tilde{\psi}_m + \tilde{\psi}_n)^{-1}
\end{aligned}$$

This yields the update rule for  $\Sigma$  as follows

$$\begin{aligned}
A'(\theta_n) &= \frac{N}{N + \lambda(N - 1)} \left( T(\mathbf{x}_n) + \frac{\lambda}{N} \sum_{m \neq n} A'(\tilde{\theta}_m/2 + \tilde{\theta}_n/2) \right) \\
\mu_n \mu_n^T + \Sigma_n &= \frac{N}{N + \lambda(N - 1)} \left( \mathbf{x}_n \mathbf{x}_n^T + \frac{\lambda}{N} \sum_{m \neq n} \left( \frac{1}{4} (\tilde{\psi}_m + \tilde{\psi}_n)^{-1} (\tilde{\phi}_m + \tilde{\phi}_n) (\tilde{\phi}_m + \tilde{\phi}_n)^T (\tilde{\psi}_m + \tilde{\psi}_n)^{-1} - (\tilde{\psi}_m + \tilde{\psi}_n)^{-1} \right) \right) \\
\Sigma_n &= \frac{N}{N + \lambda(N - 1)} \left( \mathbf{x}_n \mathbf{x}_n^T + \frac{\lambda}{N} \sum_{m \neq n} \left( \frac{1}{4} (\tilde{\psi}_m + \tilde{\psi}_n)^{-1} (\tilde{\phi}_m + \tilde{\phi}_n) (\tilde{\phi}_m + \tilde{\phi}_n)^T (\tilde{\psi}_m + \tilde{\psi}_n)^{-1} - (\tilde{\psi}_m + \tilde{\psi}_n)^{-1} \right) \right) \\
&\hspace{20em} - \mu_n \mu_n^T
\end{aligned}$$

**Final parameter updates:**

$$\begin{aligned}
\Psi &= (\tilde{\psi}_m + \tilde{\psi}_n), \quad \Phi = (\tilde{\phi}_m + \tilde{\phi}_n) \\
\mu_n &= \frac{N}{N + \lambda(N - 1)} \left( \mathbf{x}_n + \frac{\lambda}{N} \sum_{m \neq n} \left( -\frac{1}{2} \Psi^{-1} \Phi \right) \right) \\
\Sigma_n &= \frac{N}{N + \lambda(N - 1)} \left( \mathbf{x}_n \mathbf{x}_n^T + \frac{\lambda}{N} \sum_{m \neq n} \left( \frac{1}{4} \Psi^{-1} \Phi \Phi^T \Psi^{-1} - \Psi^{-1} \right) \right) - \mu_n \mu_n^T
\end{aligned}$$