Gaussian likelihood with Gaussian prior on its mean

Emt
CS,UBC

1 Non-Conjugate Case

We have the following joint distribution of $\mathbf{y}, \mathbf{w}$,

$$p(\mathbf{y}, \mathbf{w}) = \mathcal{N}(\mathbf{y}|X\mathbf{w}, \Sigma_y) \cdot \mathcal{N}(\mathbf{w}; \mathbf{0}, \Gamma) \quad (1)$$

$$= (2\pi)^{-(n+d)/2}|\Sigma_y^{-1/2}||\Gamma^{-1/2}| \exp\left[-\frac{1}{2}(\mathbf{y} - X\mathbf{w})^T\Sigma_y^{-1}(\mathbf{y} - X\mathbf{w})\right] \exp\left[-\frac{1}{2}\mathbf{w}^T\Gamma^{-1}\mathbf{w}\right] \quad (2)$$

Simplify the terms in exponential:

$$(\mathbf{y} - X\mathbf{w})^T\Sigma_y^{-1}(\mathbf{y} - X\mathbf{w}) + \mathbf{w}^T\Gamma^{-1}\mathbf{w} \quad (3)$$

$$= \mathbf{y}^T\Sigma_y^{-1}\mathbf{y} - 2\mathbf{y}^T\Sigma_y^{-1}X\mathbf{w} + \mathbf{w}^TX^T\Sigma_y^{-1}X\mathbf{w} + \mathbf{w}^T\Gamma^{-1}\mathbf{w} \quad (4)$$

$$= \mathbf{w}^T(X^T\Sigma_y^{-1}X + \Gamma^{-1})\mathbf{w} - 2\mathbf{y}^T\Sigma_y^{-1}X\mathbf{w} + \mathbf{y}^T\Sigma_y^{-1}\mathbf{y} \quad (5)$$

Let $K^{-1} = X^T\Sigma_y^{-1}X + \Gamma^{-1}$ and $\mu = KX^T\Sigma_y^{-1}\mathbf{y}$, then the above expression is equal to

$$\mathbf{w}^TK^{-1}\mathbf{w} - 2\boldsymbol{\mu}^TK^{-1}\mathbf{w} + \mathbf{y}^T\Sigma_y^{-1}\mathbf{y} \quad (6)$$

$$= \mathbf{w}^TK^{-1}\mathbf{w} - 2\boldsymbol{\mu}^TK^{-1}\mathbf{w} + \boldsymbol{\mu}^TK^{-1}\boldsymbol{\mu} - \mathbf{y}^T\Sigma_y^{-1}XKX^T\Sigma_y^{-1}\mathbf{y} + \mathbf{y}^T\Sigma_y^{-1}\mathbf{y} \quad (7)$$

$$= (\mathbf{w} - \boldsymbol{\mu})^TK^{-1}(\mathbf{w} - \boldsymbol{\mu}) + \mathbf{y}^T(\Sigma_y^{-1} - \Sigma_y^{-1}XKX^T\Sigma_y^{-1})\mathbf{y} \quad (8)$$

$$= (\mathbf{w} - \boldsymbol{\mu})^TK^{-1}(\mathbf{w} - \boldsymbol{\mu}) + \mathbf{y}^T\left(\Sigma_y + X\Gamma X^T\right)^{-1}\mathbf{y} \quad (9)$$

where the last step is obtained using the following matrix inversion lemma,

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1} \quad (10)$$

Substituting back,

$$p(\mathbf{y}, \mathbf{w}) = (2\pi)^{-(n+d)/2}\left(|\Sigma_y||\Gamma|\right)^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{w} - \boldsymbol{\mu})^TK^{-1}(\mathbf{w} - \boldsymbol{\mu})\right] \exp\left[-\frac{1}{2}\mathbf{y}^T\left(\Sigma_y + X\Gamma X^T\right)^{-1}\mathbf{y}\right] \quad (11)$$

$$= (2\pi)^{-(n)/2}\left(|\Sigma_y||\Gamma||K^{-1}|\right)^{-1/2}\mathcal{N}(\mathbf{w}|\boldsymbol{\mu}, K) \exp\left[-\frac{1}{2}\mathbf{y}^T\left(\Sigma_y + X\Gamma X^T\right)^{-1}\mathbf{y}\right] \quad (12)$$

Using the following matrix-determinant lemma,

$$|A + UWV^T| = |W^{-1} + V^TA^{-1}U||W||A| \quad (13)$$

we get the following simplification for the determinant term,

$$|\Sigma_y||\Gamma||K^{-1}| = |\Sigma_y||\Gamma||X^T\Sigma_y^{-1}X + \Gamma^{-1}| \quad (14)$$

$$= |\Sigma_y||\Gamma||\Sigma_y^{-1}||\Gamma^{-1}||\Sigma_y + X\Gamma X^T| \quad (15)$$

$$= |\Sigma_y + X\Gamma X^T| \quad (16)$$

This gives us the final expression,

$$p(\mathbf{y},\mathbf{w}) \quad = \quad (2\pi)^{-(n)/2}|\Sigma_y + X\Gamma X^T|^{-1/2}\mathcal{N}(\mathbf{w}|\boldsymbol{\mu},K)\exp\left[-\frac{1}{2}\mathbf{y}^T\left(\Sigma_y + X\Gamma X^T\right)^{-1}\mathbf{y}\right] \tag{17}$$

We have the following posterior for weights and marginal likelihood,

$$p(\mathbf{w}|\mathbf{y}) \quad = \quad \mathcal{N}(\mathbf{w}|\boldsymbol{\mu},K) \tag{18}$$

$$p(\mathbf{y}) \quad = \quad (2\pi)^{-n/2}|\Sigma_y + X\Gamma X^T|^{-1/2}\exp\left[-\frac{1}{2}\mathbf{y}^T\left(\Sigma_y + X\Gamma X^T\right)^{-1}\mathbf{y}\right] \tag{19}$$

where $K^{-1} = X^T\Sigma_y^{-1}X + \Gamma^{-1}$ and $\mu = KX^T\Sigma_y^{-1}\mathbf{y}$.

1.1 Fast computation

If $n < d$, then we can compute the marginal likelihood by computing the matrix $\Sigma_y + X\Gamma X^T$ directly. In case Γ is diagonal, $X\Gamma X^T = \sum_i \Gamma_{ii}x_ix_i^T$. In any case, storing gram matrix should be sufficient for computing the inverse of the matrix.

If $n > d$, then we can resort to MIL and compute K instead which is $d \times d$. We can obtain the inverse of the desired matrix, using MIL as follows,

$$(\Sigma_y + X\Gamma X^T)^{-1} = \Sigma_y^{-1} - \Sigma_y^{-1}XKX^T\Sigma_y^{-1} \tag{20}$$