1 Nonlinear regression/kernel regression exercises

1.1 Definitions

Let

$$\begin{aligned} X &= (x_1, \ldots, x_N) \\ T &= (t_1, \ldots, t_N) \\ \phi(x) &= (\phi_1(x), \ldots, \phi_M(x)) \\ \Phi(X) &= \begin{pmatrix} \phi_1(x_1) & \cdots & \phi_M(x_1) \\ \cdots & \cdots & \cdots \\ \phi_1(x_N) & \cdots & \phi_M(x_N) \end{pmatrix} \end{aligned} \tag{1}$$

1.2 Isotropic nonlinear regression

Let $w \sim \mathcal{N}\left(0, \alpha^{-1} I\right)$, $t|w, x \sim \mathcal{N}\left(\langle \phi(x), w \rangle, \beta^{-1}\right)$.

(a) Write out (log) of $P(T, w|X)$

(b) Using completing the square or other means, derive an expression for $P(w|T, X) = \mathcal{N}\left(\mu_{w|T}, \Sigma_{w|T}\right)$

(c) Continue completing the square with (b), and marginalize out w to come up with an expression for $P(T|X)$. You can ignore normalization factors for now (so just focus on the exponential term).

(d) In (c), the expression for Σ_T, the quadratic term, will be very complicated. Use Woodbury's identity to prove that $\Sigma_T = \Phi \alpha^{-1} I \Phi^{\intercal} + \beta^{-1} I$.

1.3 More general nonlinear regression

More generally, let $w \sim \mathcal{N}\left(\mu_w, \Sigma_w\right)$, $T|w, X \sim \mathcal{N}\left(\Phi w, \Sigma_{T|X}\right)$.
For now, assume $\mu_w = 0$ (problem 3 generalizes to other means).

(a) Write out (log) of $P(T, w|X)$

(b) Using completing the square or other means, derive an expression for $P(w|T, X) = \mathcal{N}\left(\mu_{w|T}, \Sigma_{w|T}\right)$

(c) Continue completing the square with (b), and marginalize out w to come up with an expression for $P(T|X)$. You can ignore normalization factors for now (so just focus on the exponential term).

(d) In (c), the expression for Σ_T, the quadratic term, will be very complicated. Use Woodbury's identity to prove that $\Sigma_T = \Phi \Sigma_w \Phi^{\intercal} + \Sigma_{T|X}$.

1.4 GP regression

Suppose $y \sim GP(\mu_y(\cdot), k(\cdot,\cdot))$
Let $Y = (y(x_1), \dots, y(x_N))$, $\mu_y = (\mu_y(x_1), \dots, \mu_y(x_N))$,

$$K = \begin{pmatrix} k(x_1, x_1) & \cdots & k(x_1, x_N) \\ \cdots & \cdots & \cdots \\ k(x_N, x_1) & \cdots & k(x_N, x_N) \end{pmatrix} \tag{2}$$

By definition, $Y \sim \mathcal{N}(\mu_y, K)$ and $T|Y, X \sim \mathcal{N}\left(Y, \Sigma_{T|Y}\right)$

(a) Write out (log) of $P(T, Y|X)$

(b) Using completing the square or other means, derive an expression for $P(Y|T, X)$

(c) Continue completing the square with (b), and marginalize out Y to come up with an expression for $P(T|X)$. You can ignore normalization factors for now (so just focus on the exponential term).

(d) In (c), the expression for Σ_T, the quadratic term, will be very complicated. Use Woodbury's identity to prove that $\Sigma_T = K + \Sigma_{T|Y}$.

1.5 Kernels

Suppose $w \sim \mathcal{N}(\mu_w, \Sigma_w)$ and define Φ, X, T as before.
Let $y(x) = \langle \phi(x), w \rangle$, $Y = (y(x_1), \dots, y(x_N))$.

(a) Derive $P(y(\hat{x}_1)) = \int P(y(\hat{x})|w)p(w|T)dw$, specifically by deriving the mean and covariance, $E_w[Y]$, $\mathrm{Cov}_w(Y)$.

(b) Suppose $E_w\{\Phi w\} = \mu_y$, $\mathrm{Cov}_w(Y) = K$, where both are determind by pre-defined functions. Then we can use the results of problem 3 to express all the results of Problems 1 and 2 in terms of μ_y and K.

(c) Suppose now that $\phi(x) = (K(x, x_1), \dots, K(x, x_N))$. Assume $w \sim \mathcal{N}(0, \Sigma_w)$. What is the mean and covariance of Y? What choice of Σ_w would make Y have a covariance of K?

1.6 Bayesian nonlinear regression

Define w, Φ, X, T as before.
Define the posterior distribution in terms of its covariance: $w|T \sim \mathcal{N}\left(\Sigma_{w|T}\Phi^{\intercal}T, \Sigma_{w|T}\right)$
Define $y(x) = \langle \phi(x), w \rangle$

(a) Derive $P(y(\hat{x}_1)|T) = \int P(y(\hat{x})|w)p(w|T)dw$ as a normal distribution (work out the mean and covariance), in terms of $\Sigma_{w|T}$.

(b) Using the GP regression problem, work out $P(\hat{t})$, in terms of Σ_w, Φ.

(c) Work out the above in terms of $K, K(\cdot,\cdot)$.

(d) If we plugged in our posterior, $\mathcal{N}\left(\mu_{w|T}, \Sigma_{w|T}\right)$ in place of our prior, and marginalized it out to solve for $P(Y)$, what is the result?

(e) Are these formulas the same as writing $t \sim GP\left(\mu_y, K\right)$?

1.7 Multivariate Gaussians

Note our random variables are unrelated to the previous problems.

$$
\begin{aligned}
X &= \begin{pmatrix} X_A \\ X_B \end{pmatrix} \\
\mu &= \begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix} \\
\Sigma &= \begin{pmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_{BB} \end{pmatrix} \\
\Lambda &:= \Sigma^{-1} \\
\Lambda &= \begin{pmatrix} \Lambda_{AA} & \Lambda_{AB} \\ \Lambda_{BA} & \Lambda_{BB} i \end{pmatrix} \\
X &\sim \mathcal{N}\left(\mu, \Sigma\right)
\end{aligned} \tag{3}
$$

(a) Write out the exponential term of $P(X)$ (ignore the normalizations constants).

(b) Complete the square for the x_A terms.

(c) Marginalize out x_A using (b) to get an expression for the exponential term of $P(x_B|x_A)$ in terms of $x_B - \mu_B$ and $(x_B - \mu_B)(x_b - \mu_B)^\intercal$.

(d) The quadratic coefficient for $x_B - \mu_B$ is written in terms of precision matrices. Use the Schur complement to convert it into covariance matrices.

(e) Use Woodbury's identity to simplify the expressions for $\mu_{x_B|x_A}, \Sigma_{x_b|x_A}$ ($x_B|x_A \sim \mathcal{N}\left(\mu_{x_B|x_A}, \Sigma_{x_b|x_A}\right)$).