GP Review and Primal and Dual Forms of BLR

Roni Khardon
Tufts University
Fall 2016

\[ f \sim GP(m(.), K(\cdot, \cdot)) \]
\[ f \sim N(m(x), K(x, x^T)) \]
\[ x = (x_1, \ldots, x_N)^T \]
\[ f = (f(x_1), \ldots, f(x_N))^T \]

Typically, do not observe \( f \) directly but observe \( t \) where \( t_i \) depends only on \( f_i \)

Regression
\[ p(t_i|f_i) = N(f_i, \frac{1}{\beta}) \]

Classification
\[ p(t_i|f_i) = \text{Bernoulli}(\sigma(f_i)) \]

In class we only discuss the regression case.

\[ \mathbf{x} = (x_1, \ldots, x_N)^T \]
\[ \mathbf{f} = (f(x_1), \ldots, f(x_N))^T \]
\[ \mathbf{t} = (t_1, \ldots, t_N)^T \]

Assuming \( m=0 \) and denoting
\[ C_N = K(x, x^T) + \frac{1}{\beta} I \]

So \( \mathbf{t} \) can be seen to be sampled from GP with a modified covariance/kernel.

Use \( C \) to denote covariance of GP we work with.

Model Selection in GP

- Evidence has closed form
  \[ t \sim N(0, C_N) \]

- So we can take derivatives of \( p(t) \) \text{wrt parameters in } C_N

- Typically optimize with gradient ascent.

BLR as GP

\[ w \sim N(0, \frac{1}{\alpha} I) \]
\[ y = \Phi w \]
\[ t|y \sim N(y, \frac{1}{\beta} I) \]

\[ E[y] = 0 \]
\[ E[y^T y] = \frac{1}{\alpha} \Phi \Phi^T = \frac{1}{\alpha} K = \hat{K} \]

Here \( K \) is the linear kernel and \( \hat{K} \) is a scaling by \( \{1/\alpha\} \)

\[ C_N = \hat{K} + \frac{1}{\beta} I = \frac{1}{\alpha} K + \frac{1}{\beta} I \]
\[ t \sim N(0, C_N) \]
The following slides show their equivalence.

**Prediction in BLR**

\[ w \sim N(0, \frac{1}{\alpha} I) \quad y = \Phi w \quad t|y \sim N(y, \frac{1}{\beta} I) \]

\[ Pr(t_{N+1}|t) \sim N(m_N^T \phi(x_{N+1}), \frac{1}{\beta} + \phi(x_{N+1})^T S_N \phi(x_{N+1})) \]

\[ m_N = \beta S_N \Phi^T t \]

\[ S_N = (\alpha I + \beta \Phi^T \Phi)^{-1} \]

The prediction should be identical to the one given by GP equations. But the expressions are different, one working in the m dimensional space and the other in the N dimensional space.

The following slides show their equivalence.

**Eq C.6**

\[ (I_k + AB)^{-1} A = A(I_{\ell} + BA)^{-1} \]

\[ A = [k \times \ell] \text{ and } B = [\ell \times k] \]

Why? Multiply right and left to get

\[ A(I_{\ell} + BA) = (I_k + AB)A \]

Which is the same as

\[ A + ABA = A + ABA \]

**Eq C.7 (Woodbury)**

\[ (A + BD^{-1}C)^{-1} = A^{-1} - A^{-1}B(D + CA^{-1}B)^{-1}CA^{-1} \]

\[ I = (A + BD^{-1}C)[A^{-1} - A^{-1}B(D + CA^{-1}B)^{-1}CA^{-1}] \]

\[ I + BD^{-1}CA^{-1} \]

\[ 0 = BD^{-1} - (A + BD^{-1}C)A^{-1}B(D + CA^{-1}B)^{-1} \]

\[ 0 = BD^{-1}(D + CA^{-1}B) - (A + BD^{-1}C)A^{-1}B \]

\[ 0 = B + BD^{-1}CA^{-1}B - B - BD^{-1}CA^{-1}B \]

**Prediction for BLR expressed via GP**

\[ t_{N+1}|t \sim N(v^T C_N^{-1} t, c - v^T C_N^{-1} v) \]

\[ v^T = (C(x_1, x_{N+1}), \ldots, C(x_1, x_{N+1})) \]

\[ c = C(x_{N+1}, x_{N+1}) \]

**BLR as GP**

\[ C_N = \frac{1}{\alpha} K + \frac{1}{\beta} I = \left[ \frac{1}{\alpha} \Phi^T + \frac{1}{\beta} I \right] \]

\[ v^T = \frac{1}{\alpha} \phi(x_{N+1})^T \Phi^T \]

\[ c = \frac{1}{\beta} + \frac{1}{\alpha} \phi(x_{N+1})^T \phi(x_{N+1}) \]

**Expressions for the mean**

\[ v^T C_N^{-1} t = \left[ \frac{1}{\alpha} \phi(x_{N+1})^T \Phi^T \right] \frac{1}{\beta} I_N + \frac{1}{\alpha} \Phi^T \Phi^{-1} t \]

\[ = \frac{1}{\alpha} \phi(x_{N+1})^T \beta I_N + \frac{1}{\alpha} \Phi^T \Phi^{-1} t \]

\[ = \frac{1}{\alpha} \phi(x_{N+1})^T I_N \frac{1}{\alpha} \beta \Phi^T \Phi^{-1} t \]

\[ = \phi(x_{N+1})^T \beta (I_N \frac{1}{\alpha} + \Phi^T \Phi^{-1} \Phi) t \]

\[ = \frac{1}{\alpha} \phi(x_{N+1})^T \beta (I_N \frac{1}{\alpha} + \Phi^T \Phi^{-1} \Phi) t \]

\[ = \phi(x_{N+1})^T \beta (I_N \frac{1}{\alpha} + \Phi^T \Phi^{-1} \Phi) t \]

**Expressions for the variance**

\[ c - v^T C_N^{-1} v \]

\[ = \left[ \frac{1}{\beta} + \frac{1}{\alpha} \phi(x_{N+1})^T \phi(x_{N+1}) \right] \left[ \frac{1}{\alpha} \phi(x_{N+1})^T \phi(x_{N+1}) \right] \frac{1}{\beta} I_N + \frac{1}{\alpha} \Phi^T \Phi^{-1} \Phi \phi(x_{N+1}) \]

\[ = \frac{1}{\beta} + \phi(x_{N+1})^T \left[ \frac{1}{\alpha} I + \Phi^T \Phi^{-1} \Phi \phi(x_{N+1}) \right] \]

\[ = \frac{1}{\beta} + \phi(x_{N+1})^T \left[ \frac{1}{\alpha} I + \Phi^T \Phi^{-1} \Phi \phi(x_{N+1}) \right] \]

\[ = \frac{1}{\beta} + \phi(x_{N+1})^T \left[ \frac{1}{\alpha} I + \Phi^T \Phi^{-1} \Phi \phi(x_{N+1}) \right] \]

\[ = \phi(x_{N+1})^T \Phi^T \Phi^{-1} \phi(x_{N+1}) \]

\[ = \phi(x_{N+1})^T S_N \phi(x_{N+1}) \]

\[ = \phi(x_{N+1})^T S_N \phi(x_{N+1}) \]