6.439 Problem Set 3

Xudong Sun

4.1

(a)

In a prediction process using Gaussian process with N training input and output pairs (X, Y), and T test inputs X_T, the joint training and test marginal likelihood:

$$P(Y, Y_T) = N(0, K_{N+T}),\ K_{N+T} = \begin{bmatrix} K_N & K_{NT} \\ K_{TN} & K_T \end{bmatrix}$$

$$(Suppose\ the\ data\ are\ centered\ to\ \mu = 0)$$

Condition on training outputs:

$$P(Y_T|Y) = \frac{P(Y, Y_T)}{P(Y)} = N(\mu_T, \Sigma_T)$$

$$\mu_T = \mu + K_{TN}K_N^{-1}Y$$

$$\Sigma_T = K_T - K_{TN}K_N^{-1}K_{NT}$$

When N is large, the inverse of the covariance matrix K_N will be computationally expensive . $The\ complexity\ of\ n\ by\ n\ matrix\ inversion\ is\ O(n^3\ )$

(b)

We can reduce the rank of the covariance using following approximation:

Given a positive definite matrix K of order n × n and a randomly generated Johnson -Lindenstrauss matrix Ω of order r × n, we find the projection matrix Φ of order m × n which approximates the range and compute the approximate SVD decomposition via Nystrom approximation with Φ.[1]

$$1. \Phi^T = left\ factor\ of\ the\ rank\ m\ spectral\ projection\ of\ K\Omega$$

$$2. K_1 = \Phi K \Phi^T$$

$$3. Choleski\ Factorization, K_1 = BB^T$$

$$4. Nystrom\ factor\ C = K\Phi^T(B^T)^{-1}, spectual\ decomposition\ C = UDV^T$$

$$K \approx\ UD^2U^T = UDV^TVDU^T = K\Phi^T(B^T)^{-1}B^{-1}\Phi K$$

$$= K\Phi^T K_1^{-1}\Phi K = (\Phi K)^{-1}(\Phi K \Phi^T)^{-1}\Phi K$$

Consider the Gaussian Process with (Gaussian) noisy observations:

$$\mu_T = \mu + K_{TN}(K_N + \sigma^2 I)^{-1}Y$$

$$\Sigma_T = K_T - K_{TN}(K_N + \sigma^2 I)^{-1}K_{NT} + \sigma^2 I$$

[1] Banerjee, Anjishnu, David B. Dunson, and Surya T. Tokdar. "Efficient Gaussian process regression for large datasets." Biometrika 100.1 (2012): 75-89.

6.439 Problem Set 3

Xudong Sun

Using the Sherman-Morrison-Woodbury matrix inversion lemma,

$$K_N \approx UD^2U^T$$

$$(K_N + \sigma^2 I)^{-1} = \sigma^{-2} I - \sigma^{-2} U (D^{-2} + \sigma^{-2} U^T U)^{-1} U^T \sigma^{-2}$$

$$= \sigma^{-2} I - \sigma^{-4} U (D^{-2} + \sigma^{-2} I)^{-1} U^T$$

In the above $(D^{-2} + \sigma^{-2} I)^{-1}$ is a diagonal matrix and the matrix inversion is avoided.