Computational Statistics
Lab V - Gaussian Processes

Luca Bortolussi

Department of Mathematics and Geosciences
University of Trieste

Office 238, third floor, H2bis
luca@dmi.units.it

Trieste, Winter Semester 2015/2016

Outline

1. GP

Task 1

- Download data from moodle2 (use the datasets for regression).
- Implement GP regression:
  - implement a kernel object, with methods to evaluate the kernel, get and set hyperparameters, get the number of hyperparameters.
  - implement methods initialising a Gaussian/ Matern kernel/ ARD Gaussian Kernel.
  - implement GP regression parametric w.r.t. a kernel object. Suggestion: initialise a GP regression object with data, compute all data dependent quantities needed to do prediction once and for all, and provide the object with methods to do prediction in a set of points.
  - implement a method to set hyperparameters optimising the marginal likelihood. Use built-in matlab optimisation functions.

How to implement GP regression

input: X (inputs), $\mathbf{y}$ (targets), k (covariance function), σ_n^2 (noise level), $\mathbf{x}_*$ (test input)

2: $L := \text{cholesky}(K + \sigma_n^2 I)$

$\boldsymbol{\alpha} := L^\top \backslash (L \backslash \mathbf{y})$ } predictive mean eq. (2.25)

4: $\bar{f}_* := \mathbf{k}_*^\top \boldsymbol{\alpha}$

$\mathbf{v} := L \backslash \mathbf{k}_*$ } predictive variance eq. (2.26)

6: $\mathbb{V}[f_*] := k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{v}^\top \mathbf{v}$

$\log p(\mathbf{y}|X) := -\frac{1}{2}\mathbf{y}^\top \boldsymbol{\alpha} - \sum_i \log L_{ii} - \frac{n}{2} \log 2\pi$ eq. (2.30)

8: **return**: $\bar{f}_*$ (mean), $\mathbb{V}[f_*]$ (variance), $\log p(\mathbf{y}|X)$ (log marginal likelihood)

Algorithm 2.1: Predictions and log marginal likelihood for Gaussian process regression. The implementation addresses the matrix inversion required by eq. (2.25) and (2.26) using Cholesky factorization, see section A.4. For multiple test cases lines 4-6 are repeated. The log determinant required in eq. (2.30) is computed from the Cholesky factor (for large n it may not be possible to represent the determinant itself). The computational complexity is $n^3/6$ for the Cholesky decomposition in line 2, and $n^2/2$ for solving triangular systems in line 3 and (for each test case) in line 5.

Cholesky factorisation

- Given a positive definite $n \times n$ matrix A, a lower triangular matrix L is the Cholesky decomposition of A iff

$$A = LL^T$$

- The computation of the Cholesky decomposition is very stable numerically, and takes $O(n^3)$ time.
- The Cholesky decomposition of A can be used to solve the linear system $Ax = b$ in two steps:
    1. solve by forward substitution the system $Ly = b$
    2. solve by backward substitution the system $L^T x = y$.
- The determinant of A is

$$|A| = \prod_i L_{ii}^2; \quad \log |A| = 2 \sum_i \log L_{ii}$$