

# Bayesian Gaussian Processes

Casey Davis

## This Github Repository

This particular Github repo, *bayesianGPs*, is a rough draft of putting together some software that will do the stuff in my dissertation, *A Bayesian Approach to Prediction and Variable Selection Using Non-Stationary Gaussian Processes*. This one will use Stan, a Bayesian programming language/software. I might also duplicate in R using Rcpp what I did when I originally wrote this software in MATLAB, but that'll probably only be to practice certain things and random experimentation.

I'll probably build up from the most basic stationary model and work towards the stuff in my dissertation. Eventually, I'll make an actual R package (Python, too?) in a different repo.

## Setup

We have observed data,  $\mathbf{Y}$ . Conditional on model parameters,  $\Lambda$ , the data are treated as a realization of a Gaussian process,  $Y(\mathbf{x})$ .

$$Y(\mathbf{x})|\Lambda \sim GP(\mu, C(\cdot, \cdot))$$

where  $\mu$  is the process mean, and  $C(\cdot, \cdot)$  is the process covariance.

$$C(\mathbf{x}, \mathbf{x}') = Cov(Y(\mathbf{x}), Y(\mathbf{x}'))$$

A common covariance function is the stationary *squared exponential* or *Gaussian* covariance function. For  $\mathbf{x} \in \mathbb{R}^D$ , it commonly is seen in some version of the form

$$C(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\sum_{d=1}^D \theta_d (x_d - x'_d)^2\right) + \sigma_\epsilon^2 \delta$$

where  $\sigma^2$  is the process variance,  $\theta_d$  is a scale-parameter for each dimension  $d$ ,  $\sigma_\epsilon^2$  is the scale of random noise, and  $\delta$  is the Kronecker delta function with  $\delta = 1$  if  $\mathbf{x} = \mathbf{x}'$  and 0 otherwise. I prefer a slightly different parameterization that eases interpretability:

$$C(\mathbf{x}, \mathbf{x}') = \sigma^2 \prod_{d=1}^D \rho_d^{16(x_d - x'_d)^2}$$

In this parameterization,  $\rho_d$  is the correlation between the response at two points that differ only in the  $d^{th}$  dimension by  $\frac{1}{4}$  of a unit. So if the inputs are all standardized to  $[0, 1]^D$ , then  $\rho_d$  is the correlation between the response at two points that differ only in the  $d^{th}$  dimension by  $\frac{1}{4}$  the range of the data in the  $d^{th}$  dimension.

Let  $\mathbf{R}$  be an  $n \times n$  correlation matrix for the observed data:

$$R_{ij} = \prod_{h=1}^d \rho_h^{16(x_{i_h} - x_{j_h})^2}$$