

Gaussian Processes - Principle Ideas

Adam Lee

I. INTRODUCTION

In simple and non-simple linear regression, we attempt to model a variable y as a function of independent variables \mathbf{x} assuming some error value ϵ e.g. $y = \mathbf{x}^T \beta + \epsilon$. It is common to assume the error terms ϵ are Gaussian distributed with constant variance σ^2 . A Bayesian approach to linear regression assumes a probabilistic space on the parameters β which get updated as more data is observed. Gaussian processes (or GPs) provide a *non-parametric* approach to regression problems by finding a distribution over the possible functions $f(\mathbf{x})$ that are consistent with our observed data. This is still a Bayesian approach, we begin with a prior distribution (in the forms of a mean function m and a kernel function $\kappa(\mathbf{x}, \mathbf{x}')$) which we update using Bayes' rule as more and more data is observed. In this report, we shall discuss the principle ideas behind Gaussian processes, and begin to build the foundations for constructing such a process. To finish, we shall consider why we would want to use a GP instead of linear regression and some of the problems we face with implementing GPs.

II. MULTIVARIATE GAUSSIAN DISTRIBUTIONS

We start by considering multivariate gaussian distribution. We say a k dimension random vector \mathbf{x} is Gaussian distributed if

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma), \quad (1)$$

where $\boldsymbol{\mu} = [\mu_1, \dots, \mu_k]$ is the expected value of \mathbf{x} and Σ is a covariance matrix, determining the similarity of two random variables $x_i, x_j \in \mathbf{x}$. If we consider a bivariate case centered about a mean $\boldsymbol{\mu} = [0, 0]$ with covariance matrix

$$\Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \quad (2)$$