Presentation Summary: Gaussian Processes

Mathias Engel

March 7, 2018

The Article

This is a summary write-up for the presentation of the paper *Gaussian Processes, A Replacement for Supervised Neural Networks?* from 1997 by David J.C. Mackay. The paper is a published set of excellent lecture notes and covers a lot of different material regarding Gaussian Processes.

It contains

- theoretical definitions of Gaussian Processes
- discussion of parametric versus non-parametric modelling
- relationship with multi-layer neural networks
- co-variance functions and ways of combining them
- exact and approximate implementations in software
- real world data examples and performance
- classification with Gaussian Processes

The article mainly draws from work by Radford Neal, Carl Edward Rasmussen and Chris Williams, all of whom greatly contributed to the study and development of Gaussian Processes. This presentation will focus on introducing Gaussian Processes and their non-parametric nature, which can seem strange at first when coming from normal nonlinear regression. It will present the papers handling of the section "From parametric models to Gaussian Processes", equation 16-31. Finally application on real data will be demonstrated.
Similar to the application in "Figure 5" the author will show data from his own research on cell movement.

The Problem of nonlinear regression

Often we are give N sets of measured input variables $\mathbf{x}$ and target (dependent) variables t

$\mathbf{X}_N, \mathbf{t}_N = \{\mathbf{x}^{(n)}, t_n\}_{n=1}^N$.

Considering typical scientific use cases, $\mathbf{x}$ is a data vector confined to I dimensions and t is a scalar measurement from an experiment.
Imagine the oxygen level ($\mathbf{t}_N$) of neuronal cells as a function of their location in the human brain and the time of measurement ($\mathbf{X}$), e.g. done with fMRI[1]. A model is then hypothesized to explain the observed phenomenon, in this case brain activity. It is sometimes forgotten that the explicit represent of a model is actually irrelevant, because we mainly want two properties that are not dependent on the representation

- predictive, preferable with estimates of uncertainty, $P(t_{N+1}|\mathbf{x}_N, \mathbf{X}_{N+1})$
- comparability to other models by comparing the joint probabilities of the observed data $P(t_N|\mathbf{x}_N, \mathbf{X}_N)$ between given models

Turning the parametric approach on its head

The most common approach to the regression problem is to form an explicit function $y(\mathbf{x}, \mathbf{w})$ to serve as our model hypothesis.

If we want to inject prior belief about a hypothesis and have a natural way of dealing with uncertainties, we would need a Bayesian framework. In that case, in addition to defining the basis function y, we will also have to decide on prior functions of the parameters w. The goal then becomes to find the optimal values of the parameters $\hat{w}$ given the observed data and our priors collectively.

The Gaussian Process reverses this thinking. In this new framework the hypothesis is instead stated only by a choice of covariance function that relates the observed data in the $\mathbf{x}$ domain.

$C_{nn'} = C(\mathbf{x}^{(n)}, \mathbf{x}^{(n')}) + \sigma_\nu^2 \delta_{nn'}$.

The Gaussian Process is therefore a natural choice for use cases with uncertain measurements from continuous time or space domains where the target values $\mathbf{t}$ are expected to have a certain covariance structure.

[1] Functional magnetic resonance imaging. Low oxygen concentration is assumed correlated with local high brain activity.