May 20, 2020

Warped Input Gaussian Processes for Time Series Forecasting

David Tolpin

Gaussian process is very useful for prediction in time series : it's able to work with almost any amount of data , it gives a prediction as distribution etc . It's more challenging to deal with nonstationarity and heteroscedasticity in time series using Gaussian Process.
In this paper a model for non-parametric warping of input space for GP is introduced to improve the prediction in non-stationary time series.

Gaussian process is fully specified by it's mean $m(x)$ and a covariance (kernel) $k(x, x')$ functions . And by providing a mean and a valid kernel , distribution over functions is defined $f \sim GP(m(x), k(x, x'))$
Posterior inference of GP can be performed analytically by closed formula.

Training a GP is done by choosing a hyper-parameters for kernel of the particular form ,And can be done using gradient based optimizations .
Warping the input space is a process of varying kernel hyper-parameters with respect to input . Warping $\theta(x, x')$ can be learned from the data (it imposes two main question - how to represent it and what is the optimization objective).

Model

To implement displacement of observation inputs , the prior on input is modelled as Gaussian with 3 diagonal stripped covariance matrix Σ ($\sigma_{ij} = 0\ \forall i, j : |i - j| > 1$), as a compromise between imposing GP and independent Gaussians on inputs .It can be viewed as a prior on distances on adjacent points .

Training:

Model training is executed via maximizing Log marginal likelihood :

$L + \sum log\ p_d \lambda_i + C($ C is normalization constant) . Where:

$$\lambda_i \sim D$$
$$\bar{x_i} = \bar{x_{i-1}} + \lambda(x_i - x_{i-1})$$
$$f \sim GP(\mu_{\bar{x}}, \Sigma_{\bar{x}})$$

To forecast some new location x_* we need to wrap it too before plugging in the formula of GP prediction . Several ways to do it are proposed .

Seasonality:

If seasonality is presented in the time series (it's modeled with GP with periodic kernel) , the input wrapping can distort the seasonality.

The proposed solution is to use both wrapped (for trend component) and unwrapped (seasonality component) input points in the Process. The kernel will receive two pairs of input points $x_i, x_j, \bar{x_i}, \bar{x_j}$ and pass the appropriate input to periodic and input components.

For empirical evaluation of the proposed model it was compared to deep GP, unwrapped data with the same kernel on natural and synthetic data .