Gaussian Process Regression

Gaussian Processes are distributions over functions *f(x)* of which the distribution is defined by a mean function *m(x)* and positive definite covariance function *k(x,x')*:

$$f(x) \sim GP\left(m(x), k(x, x')\right)$$

$$where\ any\ finite\ subset\ X = \{x_1 \ldots\ldots x_n\}$$

The marginal distribution is a multivariate Gaussian distribution:

$$f(X) \sim GP\left(\mu(X), k(X, X)\right)$$

The general function of a GPR is:

$$y = f(x) + \epsilon_i \ \ where \ \ \epsilon_i \sim (0, \sigma^2 I)$$

A kernel is used to describe the covariance *k(x,x')* of the GP random variables. There are many possible kernels like the RBF (radial basis function), periodic or the exponentiated quadratic kernel. In this case the exponentiated quadratic Kernel is used:

$$k(x_a, x_b) = \sigma^2 \exp\left(-\frac{|x_a - x_b|^2}{2l^2}\right) where$$

$$where \ \ \sigma^2 = signal\ variance > 0$$

$$l = lengthscale > 0$$

Using the exponentiated quadratic will result in a smooth prior on functions sampled from the GP. The smoothness is described by the lengthscale, *l*. Therefore, the parameters σ^2 and l must be optimized for maximum accuracy. In this case using gradient descent, which means calculating the derivatives and using them in a cost function until it converges.

Most of the times, the mean function can be assumed to be zero. If that is not the case, a simple transformation will allow the use of mean =0.

The function:

$$f \sim GP\ (\mu, \sigma^2\ )$$

can become:

$$f' \sim GP\ (0, \sigma^2\ )$$

$$using\ the\ transformation\ f' = f - \mu$$

In order to achieve this, the linear regression model can be used. The predictions from the linear regression can be subtracted from the y-train values and form the *f'* for the GPR. Finally, μ and sigma can be used for the GP predictions:

$$P(y_* | X_*, X, y, \theta, \sigma^2\ ) \sim N(\mu, sigma)$$

$$where\ \mu = K_{x_*,x}\left(K_{x,x} + \sigma^2\ I\right)^{-1} y$$

$$sigma = K_{x_*,x_*} - \left(K_{x_*,x}\right)\left(K_{x,x} + \sigma^2\ I\right)^{-1}\left(K_{x,x_*}\right)$$