Gaussian Process Regression

Prof. S.R Khare
Poorvi Agrawal

25 August 2015

Implementation

Gaussian Process regression is a flexible regression method that gives uncertainity estimate too. Gaussian process is specified by its mean function and covariance function. Free parameters in these functions are called hyperparameter. Most common covariance function is Radial basis function (RBF) which is nothing but squared exponential function

$$k(x, x') = \sigma_f^2 exp[\frac{-(x - x')^2}{2l^2}]$$

where l is the distance between x and x' and σ_f is large for functions which covers high range on y axis. When x $\approx$ x' then k(x,x') approaches to the maximum σ_f^2 and f(x) and f(x') are highly correlated which makes our function smooth. When x is far away from x' then k(x,x') $\approx$ 0, no correlation. Above kernel equation we have considered is for clean data (without noise). But in general we have noise in our data, so our model typically looks like

$$y = f(x) + N(0, \sigma_n^2)$$

and kernal equation can be modified as

$$k(x, x') = \sigma_f^2 exp[\frac{-(x - x')^2}{2l^2}] + \sigma_n^2 \delta(x, x')$$

where $\delta(x, x')$ is Kronecker delta function which takes value 1 when x equals to x' and 0 otherwise.

Our aim is to find $p(y_*|\mathbf{y})$. It can be shown that this probability follows gaussian distribution.

$$y_*|\mathbf{y} \sim N(\overline{y}_*, var(y_*))$$

where $\overline{y}_* = K_* K^{-1}\mathbf{y}$ and $var(y_*) = K_{**} - K_* K^{-1} K_*^T$ and K, K_*, K_{**} are the covariance matrix.