0.1 Gaussian Process Regression

A Gaussian process (GP) is a collection of random variables such that any collection obey a multivariate normal distribution (MVN). Generally,

$$\text{MVN} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{1}$$

is defined by its mean ($\boldsymbol{\mu} \in \mathbb{R}^d$) and covariance ($\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$, positive semi-definite) for a d-dimensional distribution.

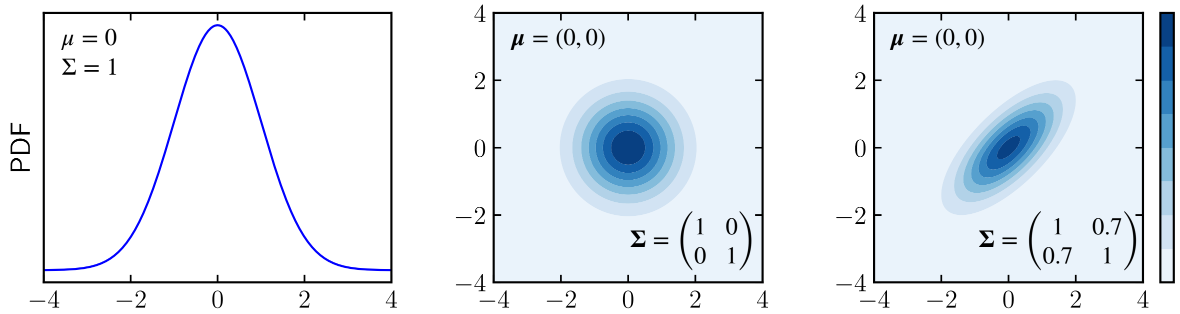

Figure 1: Various multivariate normal distributions in one and two dimensions.

To use a GP for regression the output (y) is defined in a linear way,

$$y = \sum_i w_i\, \phi_i(\boldsymbol{x}) \equiv \boldsymbol{w}^T \tilde{\boldsymbol{x}} \tag{2}$$

where $\boldsymbol{w}$ are the weights to be found and ϕ_i is a non-linear basis function of the input values ($\boldsymbol{x}$). The distribution of outputs and weights are set to be Gaussians,

$$\begin{aligned} y^* &= \boldsymbol{w}^T\tilde{\boldsymbol{x}} + \varepsilon \qquad &:&\quad \varepsilon \sim \mathcal{N}(0, \sigma^2) \\ w &\sim \mathcal{N}(0, c\mathbb{I}) \qquad &:&\quad c \in \mathbb{R} \end{aligned} \tag{3}$$

where the asterisk denotes a predicted value, $\boldsymbol{w}^T\tilde{\boldsymbol{x}} := z$ form the GP and ϵ is a Gaussian error.

To do inference requires the posterior (predictive) distribution of y, given a set of known values. For the full set of y then,[a]

$$\begin{aligned} \boldsymbol{y}^* &= \boldsymbol{z} + \boldsymbol{\varepsilon} \\ &\sim \mathcal{N}(\boldsymbol{\mu}, \mathsf{K} + \sigma^2\mathbb{I}) \end{aligned} \tag{4}$$

defining $\mathsf{C} = \mathsf{K} + \sigma\mathbb{I}$ and splitting the vectors and matrices into known (train) and an unknown ($*$) components,

$$\begin{pmatrix} y^* \\ \boldsymbol{y}_{\text{train}} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu* \\ \boldsymbol{\mu}_{\text{train}} \end{pmatrix}, \begin{pmatrix} \mathsf{C}^* & \mathsf{C}^*_{\text{train}} \\ \mathsf{C}^{*T}_{\text{train}} & \mathsf{C}_{\text{train}} \end{pmatrix} \right) \tag{5}$$

then the conditional distribution of y^* is,

[a] As the sum of independent MVNs just sum their means and covariances.

$$(y^*|\boldsymbol{y}_{\text{train}}) \sim \mathcal{N}(\mu^* + \mathsf{C}^{*T}_{\text{train}}\mathsf{C}^{-1}_{\text{train}}(\boldsymbol{y}_{\text{train}} - \boldsymbol{\mu}_{\text{train}}), \mathsf{D}) \tag{6}$$

where D is the covariance the exact form of which is omitted.

References

[1] C. E. Rasmussen and C. K. I. Williams, Gaussian Processes for Machine Learning, the MIT Press, 2006, ISBN 026218253X.