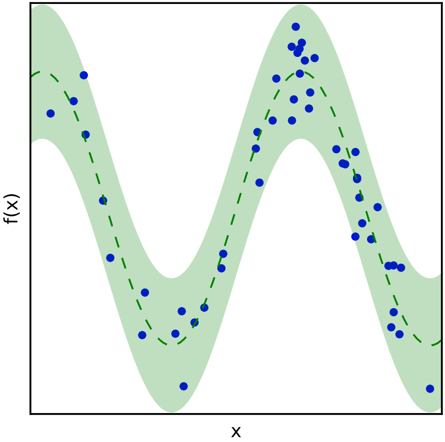
Supervised Learning

Gaussian processes for regression

Gilles Louppe (@glouppe)

October 17, 2016

Problem statement



Regression, with error bars

Problem statement

- Training data $\mathcal{L} = \{\mathbf{x}_i, f_i\}_{i=1}^{N}$
- $\mathbf{x}_i \in \mathbb{R}^d$
- $y_i = f(\mathbf{x}_i)$ for some unknown f
- We wish to recover the underlying process f from the observed data, i.e infer f^* for some unseen $\mathbf{x}^*$, using $p(f^*|\mathbf{x}^*, \mathcal{L})$.

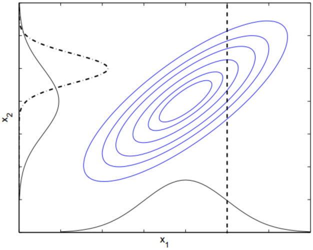
The multivariate gaussian distribution

Let assume a random variable $\mathbf{f}$ following a multivariate Gaussian distribution, and let partition its dimensions into two sets A and B:

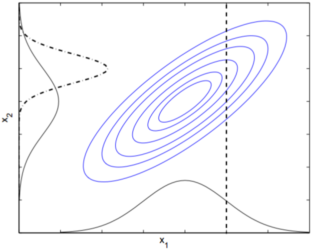
$$\underbrace{f_1, \ldots, f_i}_{\mathbf{f}_A}, \underbrace{f_{i+1}, \ldots, f_N}_{\mathbf{f}_B} \sim \mathcal{N}(\mu, K)$$

$$\mu = \begin{bmatrix} \mu_A \\ \mu_B \end{bmatrix} \in \mathbb{R}^N$$

$$K = \begin{bmatrix} K_{AA} & K_{AB} \\ K_{BA} & K_{BB} \end{bmatrix} \in \mathbb{R}^{N \times N}$$



Marginal and conditional distributions



- The *marginal* distribution of a multivariate gaussian is a multivariate gaussian:

$$\mathbf{f}_A \sim \mathcal{N}(\mu_A, K_{AA})$$

- The *conditional* distribution of a multivariate gaussian is a multivariate gaussian:

$$\mathbf{f}_A | \mathbf{f}_B \sim \mathcal{N}(\mu_A + K_{AB} K_{BB}^{-1} (\mathbf{f}_B - \mu_B), K_{AA} - K_{AB} K_{BB}^{-1} K_{BA})$$

Gaussian Processes

Definition. A *Gaussian process* is a (potentially infinite) collection of random variables such that the joint distribution of any finite number them is multivariate Gaussian.

Gaussian Processes: the simpler explanation

A *Gaussian process* is a

HUGE[1]

multivariate gaussian distribution.

[1] The dimension is the number of data points.

Gaussian distributions vs. Gaussian processes

Gaussian distribution

$$x \sim \mathcal{N}(\mu, K)$$

- Distribution over vectors.
- Fully specified by a mean and covariance.
- The position of the random variable in the vector plays the role of the index.

Gaussian process

$$f \sim \mathcal{GP}(\mu(\cdot), K(\cdot, \cdot))$$

- Distribution over functions.
- Fully specified by a mean *function* and a covariance *function*.
- The argument of the random function plays the role of the index.

Specifying a gaussian process

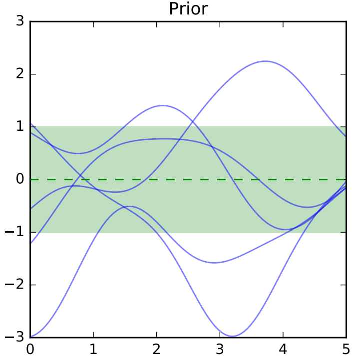
- **Mean function** $\mu(\cdot)$: Usually defined to be zero.

- **Covariance function** $K(\cdot, \cdot)$: This defines the prior properties of the functions considered for inference:
  - Stationarity
  - Smoothness
  - Length scales
  - ...

Gaussian process prior

Assuming the zero mean function, for any set of inputs $\mathbf{x}_1, \ldots, \mathbf{x}_N$, we may compute the covariance matrix K such that $K_{ij} = K(\mathbf{x}_i, \mathbf{x}_j)$, which defines a joint distribution over function values at those points:

$$\begin{bmatrix} f(\mathbf{x}_1) \\ \ldots \\ f(\mathbf{x}_N) \end{bmatrix} \sim \mathcal{N}(\mathbf{0}, K)$$



References