5/14/21

Gaussian Processes as a Regression Tool

0. Intro remarks
1. Motivation
2. Formal definition and use in practice
3. Examples (Jupyter notebook)

---

0 Ref: Gaussian Processes for ML, Rasmussen & Williams.

Setting the stage: Supervised learning

$$\{(x_i, y_i)\}_{i=1}^{N}, \quad x_i \in \mathbb{R}^d, \quad y_i \in \mathbb{R}$$

1) $y_i = f(x_i)$

2) $y_i = f(x_i) + \eta$

1

Def. 0 $X \sim N(0,1)$ iff $f_X(x) =$

$$= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

$Y \sim N(m, \sigma^2)$ iff

1) $Y = \sigma X + m$

2) $f_Y(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y-m)^2}{2\sigma^2}\right)$

Gaussian vector: extension to $\mathbb{R}^d$.

Def 1 Let X be a $\mathbb{R}^d$-valued r.vector.

X is a Gaussian vector with mean $\mu \in \mathbb{R}^d$ and covariance matrix K iff

$$f_X(x) = \frac{1}{(2\pi)^{\frac{d}{2}}\sqrt{\det(K)}} \exp\left(-\frac{1}{2}(x-\mu)^T K^{-1}(x-\mu)\right)$$

$X \sim N(\mu, K)$

Def 2 $X : \Omega \to \mathbb{R}^d$ is a G. vec. iff $\forall u \in \mathbb{R}^d$

$\langle u, X \rangle$ is a G. r.v.

We have:

1) in 1-d, $X : \Omega \to \mathbb{R}$ $\qquad X(\omega) = r$
2) in d-dim. , $X : \Omega \to \mathbb{R}^d$ $\qquad X(\omega) = \begin{pmatrix} r_1 \\ \vdots \\ r_d \end{pmatrix}$

$X : \Omega \to \ell_d$, $\ell_d = \{ f : \{1, \dots, d\} \to \mathbb{R} \mid f(i) < \infty \ \forall i \in \{1, \dots, d\} \}$

space of functions

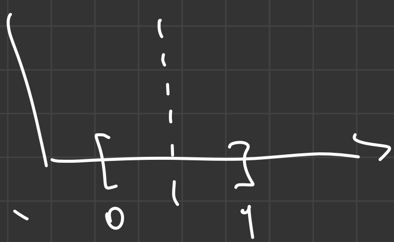
1d	d dim	∞ dim
$m \in \mathbb{R}$	$\mu \in \mathbb{R}^d \ (\mu \in \ell_d)$	$a \in \ell_T$
$\sigma^2 \in \mathbb{R}$	$K \in \mathbb{R}^{d \times d} \ (K \in \ell_{d \times d})$	$\Gamma \in \ell_{T \times T}$

$\ell_{d \times d} = \{ f : \{1, \dots, d\} \times \{1, \dots, d\} \to \mathbb{R} \}$

$\ell_T = \{ f : T \to \mathbb{R} \mid f(t) < \infty \ \forall t \in T \}$

Dual motivation:

1) Gaussian processes allow us to <u>sample</u> r. functions with certain properties (e.g. smoothness)



$\longrightarrow$ 2) Just like in the 1d and d-dim. <u>fit</u> the model (m, σ^2) or (μ, K) to the data, we can fit a GP to the data

<u>Def 3</u> A collection of random variables $\{X_t\}_{t \in T}$ is a <u>centered Gaussian process</u> iff

$$\forall n \in \mathbb{N}, \ \forall t_1, \dots, t_n \in T, \ \forall \lambda_1, \dots, \lambda_n \in \mathbb{R}$$

$$\lambda_1 X_{t_1} + \dots + \lambda_n X_{t_n} \text{ is a centered G.r.v.}$$

$$X \sim N(a, \Gamma)$$

mean $\quad a : T \to \mathbb{R}, \quad a(t) = \mathbb{E}[X_t]$

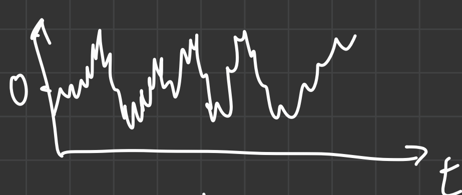
cov. func. $\quad \Gamma : T \times T \to \mathbb{R}, \quad \Gamma(t,s) = \mathrm{Cov}(X_t, X_s)$

$$= \mathbb{E}[X_t X_s]$$

if centered

Examples

1) $T = \mathbb{R}_+$, $\Gamma(s,t) = \min(s,t) \Rightarrow$ Brownian motion

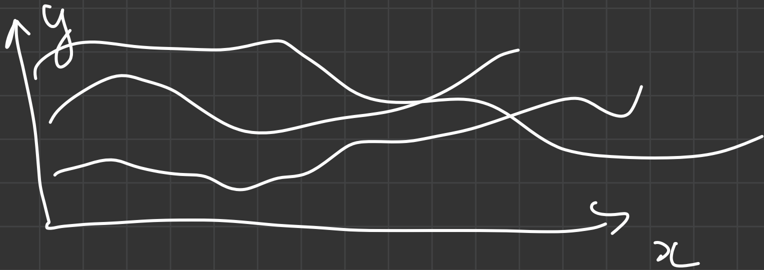


2) $T = \mathbb{R}^d$, Γ is symm., pos. $\Rightarrow$ G. process

3) $T = L^2$, $\Gamma(f,g) = \langle f, g \rangle_{L^2}$

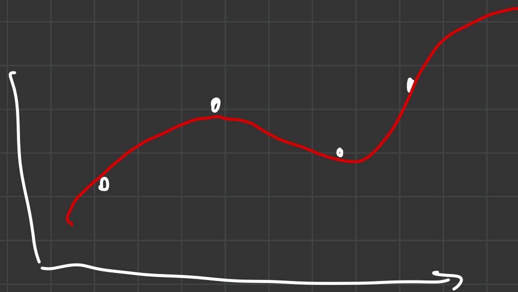
$\Rightarrow$ Gaussian field

4) $T = \{1, \dots, d\} \Rightarrow$ G. r. vec.



Regression

(a.k.a. kriging)



Theorem $(X_t)_{t\in T}$ — GP with 0 mean and cov. f. Γ

Let $(t_1, \dots, t_n) \in T^n$ with $t_i \neq t_j$ $\forall i \neq j$.

For $t \in T$ we have:

$$\mathbb{E}[X_t \mid X_{t_1}, \dots, X_{t_n}] = \sum_{ij=1}^{n} X_{t_i} \Theta^{-1}_{ij} \Gamma(t, t_j),$$

where $\Theta \in \mathbb{R}^{n\times n}$, $\Theta_{ij} = \Gamma(X_{t_i}, X_{t_j})$.

$\mathcal{O}(n^3)$

$y_i = f(x_i)$

$y_i \hat{=} X_{t_i}$; $x_i \hat{=} t_i$