TAM 598 Lecture 21:

Gaussian Process Regression

Announcements:

- HW 5 covers lectures 17-20; due on Fri Apr 18
- HW 6 covers lectures 21-23; due on Fri May 2

Gaussian Process Regression - fully non-parametric, Bayesian regression

goal: given some data, learn a function $f(\cdot): \mathcal{X} \to \mathbb{R}$

approach:

(1) before gathering any data, use your beliefs to generate a probability measure $p(f(\cdot))$ so that we can sample possible f's. This is your prior

(2) gather data D and model the likelihood of the data $p(D \mid f(\cdot))$

(3) Use Bayes' rule to come up with your posterior probability measure over the space of functions

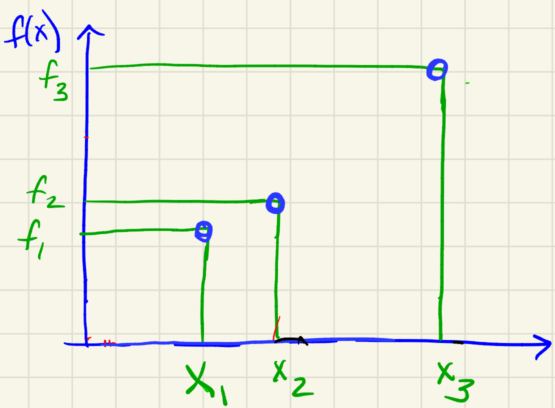
stochastic process - collection of random variables $\{\underset{\sim}{x}_i\}$ for some i in set I.

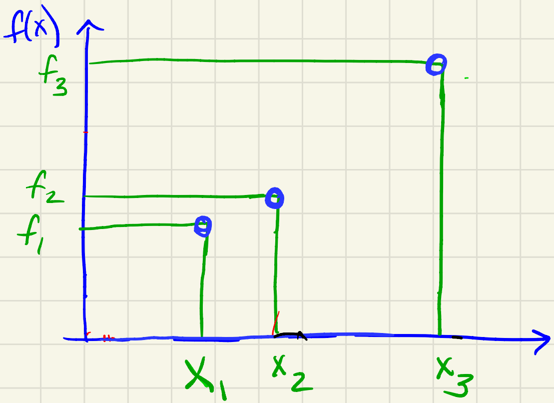
eg $\underset{\sim}{x}_t = \underset{\sim}{x}(t)$ is a stochastic process, parametrized by time usually discrete

gaussian process - a generalization of a multivariate gaussian to infinite dimensions, so a continuous stochastic process difference w/ multivariate Gaussian

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

example:



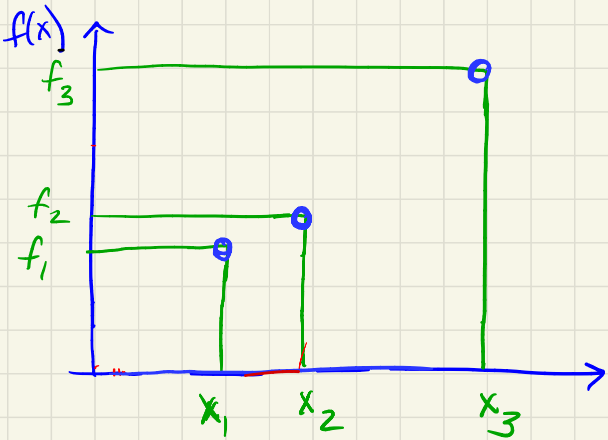


Approach :

1

2

Regression



→ x's are given

→ model f's as if pulled from a multivariate gaussian

$$\begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} \sim \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} K_{11} & K_{12} & K_{13} \\ K_{21} & K_{22} & K_{23} \\ K_{31} & K_{32} & K_{33} \end{bmatrix} \right)$$

How to interpret the mean $m(\cdot)$?

- For any point $\underset{\sim}{x}$, $m(\underset{\sim}{x})$ should give us the expected value $\mathbb{E}[f(\underset{\sim}{x})]$

- before we measure any data, options for $m(\underset{\sim}{x})$ are

⊙ $m(\underset{\sim}{x}) = 0$ or $m(\underset{\sim}{x}) =$ constant

⊙ $m(\underset{\sim}{x}) = c_0 + \sum_{i=1}^{d} c_i \underset{\sim}{x}_i$ linear

⊙ use basis functions

$$m(\underset{\sim}{x}) = \sum_{i=1}^{m} c_i \phi_i(\underset{\sim}{x})$$

⊙ generalized polynomial chaos (gPC): use d polynomial basis functions up to degree p: $m(\underset{\sim}{x}) = \sum_{i=1}^{d} c_i \phi_i(\underset{\sim}{x})$ with $\underbrace{\int \phi_i(x)\,\phi_j(x)\,d\mu(\underset{\sim}{x})}_{=\delta_{ij}}$

How to interpret the covariance?

- diagonal terms $k(\underset{\sim}{x}, \underset{\sim}{x})$

- off-diagonal terms $k(\underset{\sim}{x}, \underset{\sim}{x}')$

properties of covariance function

1) $k(\underset{\sim}{x}, \underset{\sim}{x}) > 0$ since it is a variance

2) $k(\underset{\sim}{x}, \underset{\sim}{x}')$ becomes smaller as the distance between $\underset{\sim}{x}, \underset{\sim}{x}'$ grows

3) for any choice of specific points $\underset{\sim}{x}_{1:n}$, the covariance matrix K needs to be positive definite

- covariance functions govern smoothness
- they can be designed to model invariances, eg if $k(T\underset{\sim}{x}, \underset{\sim}{x}') = k(\underset{\sim}{x}, T\underset{\sim}{x}') = k(\underset{\sim}{x}, \underset{\sim}{x}')$ then the GP is also invariant w/r/t T

Gaussian Process : a distribution over functions

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

$$m(x) = \mathbb{E}\left[f(x)\right]$$

$$K(x,x') = \mathbb{E}\left[\left(f(x) - m(x)\right)\left(f(x') - m(x')\right)^T\right]$$

$$K(x,x') = \exp\left(-\frac{1}{2}(x-x')^2\right)$$