13 Gaussian Process Regression

[See STAT 4430/5430]

DEF $\underline{x} = (x_1, \dots, x_n)^T$ is MVN$(\underline{\mu}, \Sigma)$ if it has pdf

$$\frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \exp\left(-\frac{1}{2}(\underline{x}-\underline{\mu})^T \Sigma^{-1} (\underline{x}-\underline{\mu})\right)$$

- $\underset{n\times 1}{\underline{\mu}} = \begin{pmatrix} Ex_1 \\ Ex_2 \\ \vdots \\ Ex_n \end{pmatrix}$

  $\Downarrow$

  Avg behavior of each variable

- $\underset{n\times n}{\Sigma} = \left(\text{Cov}(x_i, x_j)\right)_{i,j=1}^{n}$

  $\Downarrow$

  Covariances / correlations between variables

$\boxed{\text{Model}}$ For $\underline{x} = (x_1, \ldots, x_p)^T$ features & response Y,

the model is:

$$Y = f(\underline{x}) + \varepsilon$$

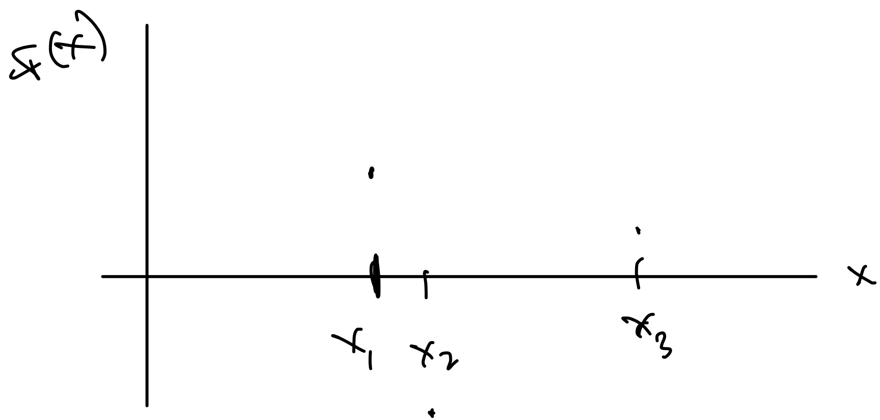
where ε is mean zero uncorrelated error and f is a random function in some class of functions

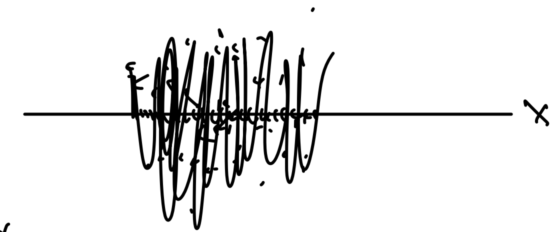
$\boxed{\text{DEF}}$ A Gaussian process (GP) $f(\underline{x})$ is a random function with

$$\begin{pmatrix} f(\underline{x}_1) \\ f(\underline{x}_2) \\ \vdots \\ f(\underline{x}_n) \end{pmatrix} \sim MVN \quad \text{for } \underline{\text{any}} \text{ choices of } \underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$$

$\boxed{\text{Ex}}$ $p=1$, $f(x) \sim N(0,1)$ and independent for any x_i, x_j.

$$\begin{pmatrix} f(x_1) \\ f(x_2) \\ f(x_3) \end{pmatrix} \sim N_3\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right)$$





Building GP framework

$f(\underline{x})$ being random $\Leftrightarrow$ putting a prior distribution on f

Simple model

$$f(\underline{x}) = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$

$$Y = f(\underline{x}) + \varepsilon$$

with

- ε iid $N(0, \sigma^2)$
- $\underline{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} \sim N_p(\underline{0}, \Sigma)$ with $\Sigma_{ij} = \mathrm{Cov}(\beta_i, \beta_j)$

here is where f gets random $\Rightarrow$

Thus,

$$f(\underline{x}) \sim N(0, \underline{x}^T \Sigma \underline{x})$$

and $\underline{f} = \begin{pmatrix} f(\underline{x}_1) \\ \vdots \\ f(\underline{x}_n) \end{pmatrix} \sim N(\underline{0}, X^T \Sigma X)$ with $X = [\underline{x}_1, \underline{x}_2 \cdots \underline{x}_n]^T$

$\Rightarrow f$ is a Gaussian process

Note: Given data $\underline{y}$ at $\underline{x}_1, \ldots, \underline{x}_n$, the observational model is $[\underline{y} \mid \underline{\beta}] \sim N(X\underline{\beta}, \sigma^2 I)$

[looks like A1!]. But goal: get posterior dist of $\underline{\beta}$:

$$[\underline{\beta} \mid \underline{y}] = \frac{[\underline{y} \mid \underline{\beta}][\underline{\beta}]}{[\underline{y}]}$$

Under this setup

$$[\underline{\beta} \mid \underline{y}] \sim N_p\left(\frac{1}{\sigma^2}\left(\frac{1}{\sigma^2} X^T X + \Sigma^{-1}\right)^{-1} X^T \underline{y}, \left(\frac{1}{\sigma^2} X^T X + \Sigma^{-1}\right)^{-1}\right) = N_p(\underline{\beta}_{post}, \Sigma_{post})$$

Prediction: Prediction at new $\underline{x}_*$ given $\underline{y}$ is:

$$[f(\underline{x}_*) \mid \underline{y}] \sim N(\underline{x}_*^T \beta_{post}, \underline{x}_*^T \Sigma_{post} \underline{x}_*)$$

[looks a lot like OLS prediction]

Problem: This model is still linear in each x_i.

A quick tour: Suppose $p=1$. Instead of $Y = \beta x + \varepsilon$, use

$$Y = \beta_1 \phi_1(x) + \beta_2 \phi_2(x) + \dots + \beta_N \phi_N(x) + \varepsilon$$

for some transformations $\phi_1, \dots, \phi_N$.

Under certain assumptions on $\phi_1, \dots$ & $\beta_1, \dots$

we get, as $N \to \infty$,

$$E f(x) = 0 \quad \text{and} \quad \mathrm{Cov}(f(x_1), f(x_2)) = k(x_1, x_2)$$

where $K(\cdot,\cdot)$ is a positive definite function.

Remark Most GP models specify a mean function $E f(x) = \mu(x)$ and a covariance function

$$k(x_1, x_2) = \operatorname{Cov}(f(x_1), f(x_2))$$

- μ + k tell us how to build $\underline{\mu}$ + Σ in $\underline{f} \sim N(\underline{\mu}, \Sigma)$.

Remark Different choices of k imply different behavior of the random function f.

A final note If $Y = f(x) + \varepsilon$ observed at $x_1, \ldots, x_n$, and $Ef(x) = 0, \forall x$ the fitted values of $f(x_1), \ldots, f(x_n)$ given data are

$$[\underline{f} \mid \underline{y}] \sim N\left(\Sigma(\Sigma + \sigma^2 I)^{-1}\underline{y},\ \Sigma - \Sigma(\Sigma + \sigma^2 I)^{-1}\Sigma\right)$$

where $\Sigma = \left(k(x_i, x_j)\right)_{i,j=1}^{n}$