Surrogate models and Gaussian Process regression – lecture 5/5

Advanced GPR

Mines St-Étienne – Majeure Data Science – 2016/2017

Nicolas Durrande (durrande@emse.fr)

Multioutputs GPR / GPR with categorical inputs

Example

We observe the temperature in two cities A and B for a few time points X_A and X_B. We assume a Gaussian process prior for these $T_A(t)$ and $T_B(t)$. What would be your prediction for the temperature in A at a new time point t?

Example

We observe the temperature in two cities A and B for a few time points X_A and X_B. We assume a Gaussian process prior for these $T_A(t)$ and $T_B(t)$. What would be your prediction for the temperature in A at a new time point t?

Ideally, we are interested in $T_A(t)|T_A(X_A), T_B(X_B)$. If $(T_A(t),\ T_A(X_A),\ T_B(X_B))$ is a Gaussian vector, we know how to compute the conditional distribution. However, it requires the cross covariance $k_{AB}(t, t') = \text{cov}[T_A(t), T_B(t')]$.

Exercise

Compute the distribution of $T_A(t)|T_A(X_A), T_B(X_B)$.

Exercise

Compute the distribution of $T_A(t)|T_A(X_A), T_B(X_B)$.

Solution

The conditional mean is:

$$m_A(t) = \mathrm{E}[T_A(t)|T_A(X_A){=}F_A, T_B(X_B){=}F_B]$$

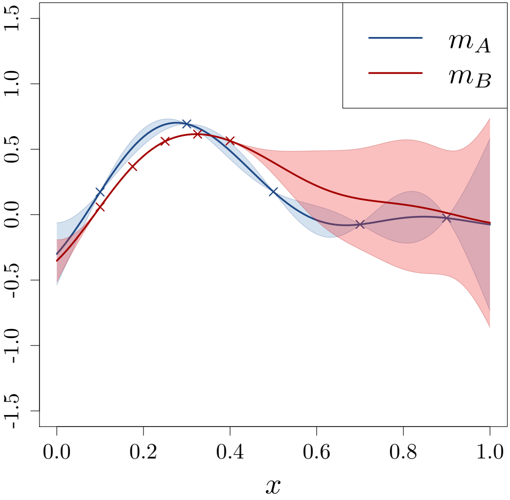
$$\begin{pmatrix} k_A(t, X_A) & k_{AB}(t, X_B) \end{pmatrix} \begin{pmatrix} k_A(X_A, X_A) & k_{AB}(X_A, X_B) \\ k_{AB}(X_A, X_B)^t & k_B(X_B, X_B) \end{pmatrix}^{-1} \begin{pmatrix} F_A \\ F_B \end{pmatrix}$$

The conditional covariance is:

$$\begin{aligned} c_A(t, t') &= \mathrm{cov}[T_A(t), T_A(t')|T_A(X_A){=}F_A, T_B(X_B){=}F_B] \\ &= k_A(t, t') - \begin{pmatrix} k_A(t, X_A) & k_{AB}(t, X_B) \end{pmatrix} \\ &\qquad \times \begin{pmatrix} k_A(X_A, X_A) & k_{AB}(X_A, X_B) \\ k_{AB}(X_A, X_B)^t & k_B(X_B, X_B) \end{pmatrix}^{-1} \begin{pmatrix} k_A(t', X_A)^t \\ k_{AB}(t', X_B)^t \end{pmatrix} \end{aligned}$$

Example

If we do the same thing fot T_B we obtain:



Instead of considering the GP to be multioutput, it is possible to see the GP as having one input but one extra categorical variable:

$$Z(t,c) = \begin{cases} Z_A(t) & \text{if } c = A \\ Z_B(t) & \text{if } c = B. \end{cases}$$

Exercise:

Compute the kernel of Z.

With this settings, the conditional mean

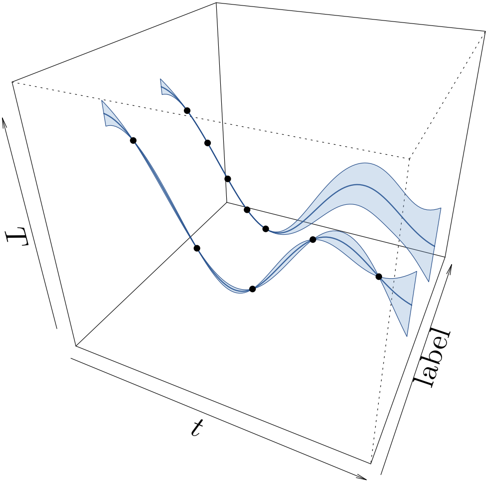
$$m_A(t) = \begin{pmatrix} k_A(t, X_A) & k_{AB}(t, X_B) \end{pmatrix} \begin{pmatrix} k_A(X_A, X_A) & k_{AB}(X_A, X_B) \\ k_{AB}(X_A, X_B)^t & k_B(X_B, X_B) \end{pmatrix}^{-1} \begin{pmatrix} F_A \\ F_B \end{pmatrix}$$

writes as an usual conditional mean

$$m_A(t) = m(t, A) = \mathrm{E}[T(t,A) | T(X){=}F] = k(\left(\begin{smallmatrix} t \\ A \end{smallmatrix}\right), X) k(X,X)^{-1} F$$

Example

We obtain this representation for the model



In the end, multioutputs GPs can be seen as GPs with one extra categorical variable indicating the output label.

All the math stay the same, we just need to specify a covariance function that takes into account this extra variable. A common approach is to consider a product covariance structure

$$k\left(\left(\begin{smallmatrix} t \\ c \end{smallmatrix}\right), \left(\begin{smallmatrix} t' \\ c' \end{smallmatrix}\right)\right) = k_{cont}(t, t')k_{disc}(c, c')$$

where $k_{disc}(c, c')$ can be described by a covariance matrix. In practice, this covariance matrix has to be estimated. If there are to outputs (or two levels for the categorical variable), it is a 2 *times*2 covariance matrix. It can be parameterised by

$$k_{disc} = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}$$

with $\sigma_1,\ \sigma_1 \geq 0$ and $\rho \in [-1, 1]$.

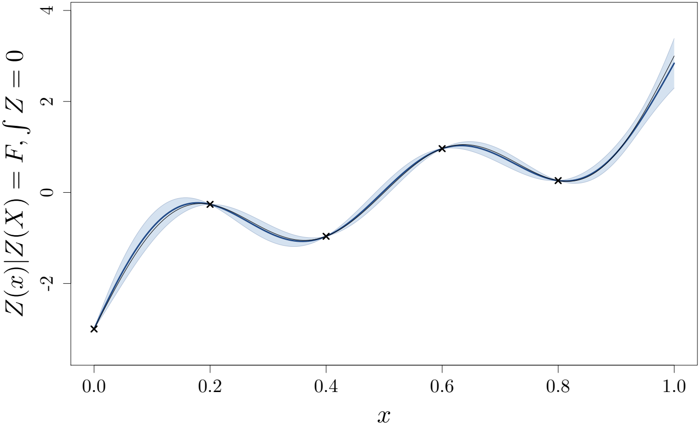
With this framework, the conditioning can also include observations more sophisticated than $Z(X) = F$...

For instance, if we know the integral of the function to approximate and it's derivative in a few points, we want to consider

$$Z \;\Big|\; Z(X) = F, \int Z = a, \frac{\mathrm{d}Z}{\mathrm{d}x}(X') = F'$$

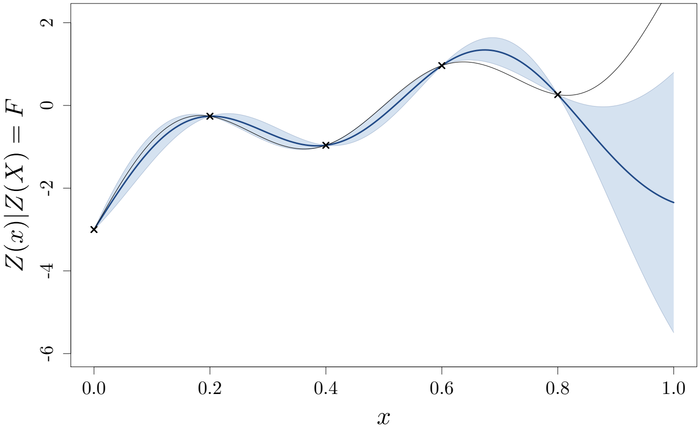
Example

If we take into account that the function is centred, we obtain:

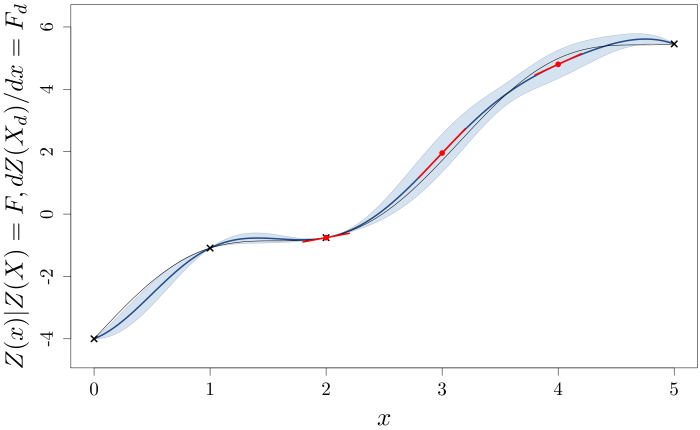


Example

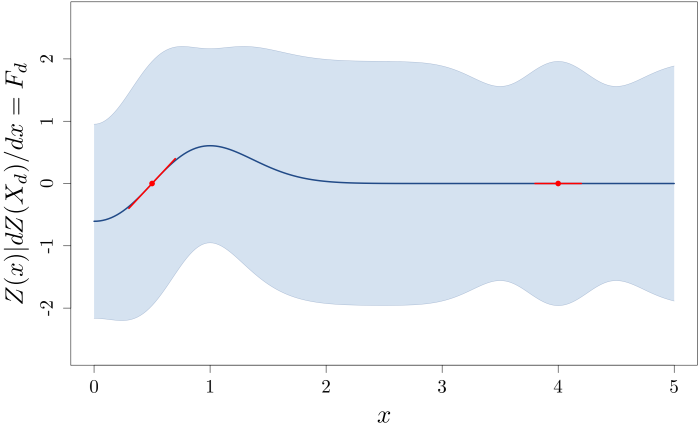
Whereas if we ignore it:



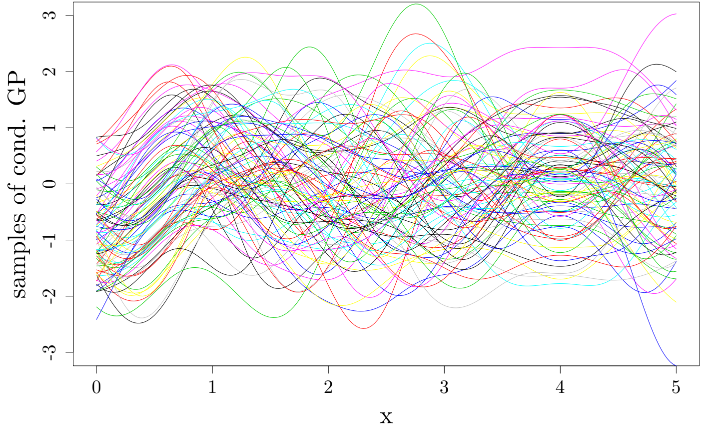
Similarly, we can include in the model some derivative observations:



We can see interesting behaviour if we look at a model with only derivatives.



As always, we can simulate conditional paths:



Conclusion

Small Recap

- GPR models do not necessarily interpolate.
- Multi-output GPR is straightforward
- Various way to create new kernels:
  - Finite dim kernels
  - Bochner theorem
  - Making New from Old
  - Linear operators

Making new from old

- sum
- product
- kernel rescaling