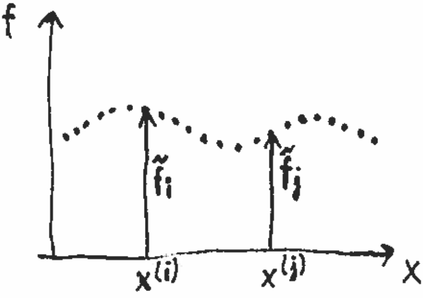
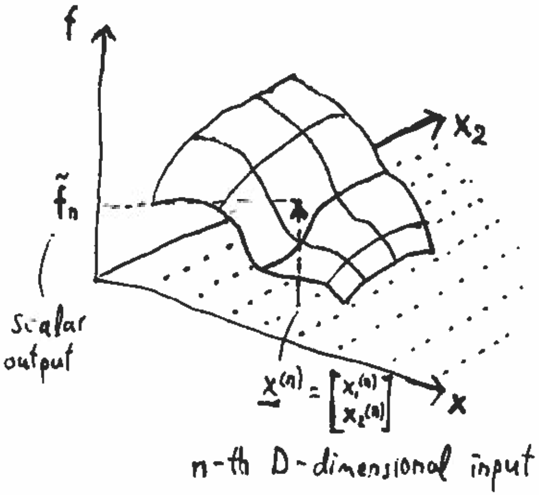
Gaussian Processes





$$\underline{\tilde{f}} = \begin{bmatrix} \tilde{f}_1 \\ \tilde{f}_2 \\ \vdots \\ \tilde{f}_N \end{bmatrix}$$ N outputs for $\{x^{(n)}\}_{n=1,\dots,N}$ stored in X

covariance / kernel function

function $f \sim GP(m, k)$

mean function

$$\Rightarrow p(\underline{\tilde{f}}) = \mathcal{N}(\underline{\tilde{f}}; m, K)$$

$m_i = m(\underline{x}^{(i)})$ usually 0

$K_{ij} = k(\underline{x}^{(i)}, \underline{x}^{(j)})$

Mercer kernels

Example kernel function:

$$k(\underline{x}^{(i)}, \underline{x}^{(j)}) = \exp\left(-\|\underline{x}^{(i)} - \underline{x}^{(j)}\|^2\right)$$

$\boxed{\text{k has to always give positive semi-definite K}}$

Properties of the Gaussian

For a joint Gaussian

$$p(\underbrace{\underline{f}}_{D_1\text{-dim}}, \underbrace{\underline{g}}_{D_2\text{-dim}}) = \mathcal{N}\left(\underbrace{\begin{bmatrix} \underline{f} \\ \underline{g} \end{bmatrix}}_{D_1+D_2\text{-dim}}, \begin{bmatrix} \underline{a} \\ \underline{b} \end{bmatrix}, \begin{bmatrix} \overset{D_1 \times D_1}{A} & \overset{D_1 \times D_2}{C} \\ C^T & \underset{D_2 \times D_2}{B} \end{bmatrix}\right)$$

Marginals: (sum rule)

$$p(\underline{f}) = \int p(\underline{f}, \underline{g})\, dg$$

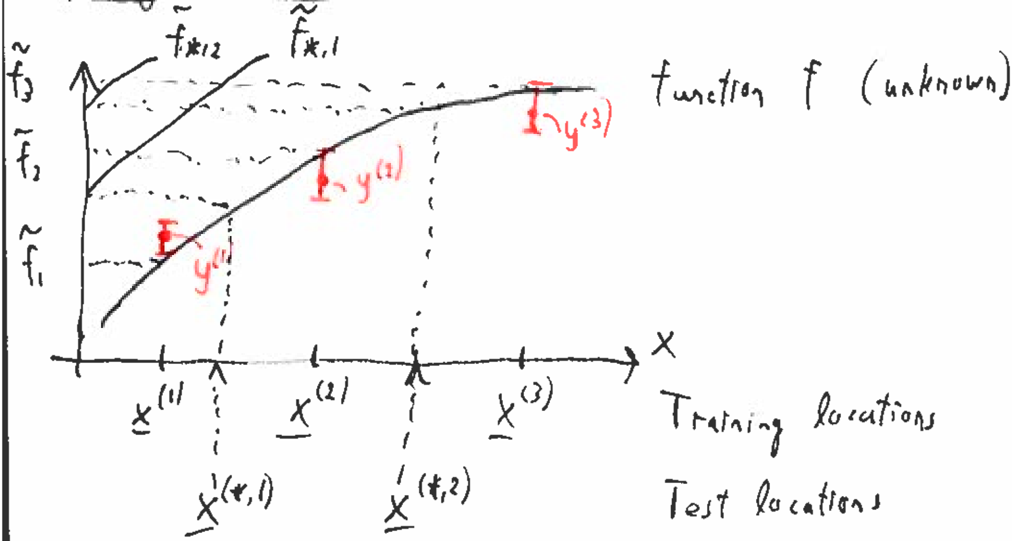
$$= \mathcal{N}(\underline{f}; \underline{a}, A)$$

Conditionals:

$$p(\underline{f} \mid \underline{g}) = \mathcal{N}(\underline{f}; \underline{a} + C B^{-1}(\underline{g} - \underline{b}), A - C B^{-1} C^T)$$

$$p(\underline{g} \mid \underline{f}) = \mathcal{N}(\underline{g}; \underline{b} + C^T A^{-1}(\underline{f} - \underline{a}), B - C^T A^{-1} C)$$

Regression with GPs



$$p(\underline{\tilde{f}}, \underline{\tilde{f}}_*) = \mathcal{N}\left(\begin{bmatrix}\underline{\tilde{f}}\\ \underline{\tilde{f}}_*\end{bmatrix}; \begin{bmatrix}\underline{0}\\ \underline{0}\end{bmatrix}, \begin{bmatrix}K(X,X) & K(X,X_*)\\ K(X_*,X) & K(X_*,X_*)\end{bmatrix}\right)$$

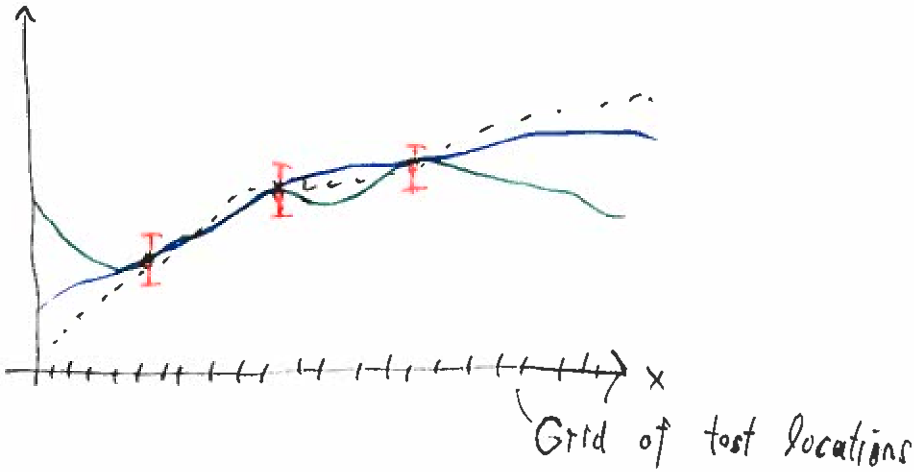
$$\underbrace{K}_{N\times M}(\underbrace{X}_{N\times D}, \underbrace{Z}_{M\times D})_{ij} = k(\underline{x}^{(i)}, \underline{z}^{(j)})$$

$$p(\underline{y}, \underline{\tilde{f}}_*)$$
$$= \mathcal{N}\left(\begin{bmatrix}\underline{y}\\ \underline{\tilde{f}}_*\end{bmatrix}; \begin{bmatrix}\underline{0}\\ \underline{0}\end{bmatrix}, \begin{bmatrix}K(X,X)+\sigma_y^2\mathbb{I} & K(X,X_*)\\ K(X_*,X) & K(X_*,X_*)\end{bmatrix}\right)$$

__Inference:__

$$p(\underline{\tilde{f}}_* | \underline{y}) = \mathcal{N}(\underline{\tilde{f}}_*; \underline{\quad\quad}, \underline{\quad\quad})$$

from standard results for Gaussians



Bayesian linear regression is a GP

Model (linear regression)

$$\tilde{f}_i = f(\underline{x}^{(i)}) = \underline{w}^T \underline{x}^{(i)} + b$$

Prior

$$\underline{w} \sim \mathcal{N}(\underline{0}, \sigma_w^2 \mathbb{I}), \quad b \sim \mathcal{N}(0, \sigma_b^2)$$

$$\text{cov}(\tilde{f}_i, \tilde{f}_j) = \mathbb{E}[\tilde{f}_i \tilde{f}_j] - \cancel{\mathbb{E}[\tilde{f}_i]\mathbb{E}[\tilde{f}_j]}$$

$$= \mathbb{E}[(\underline{w}^T \underline{x}^{(i)} + b)^T (\underline{w}^T \underline{x}^{(j)} + b)]$$

$$= \mathbb{E}[\underline{x}^{(i)T} \underline{w}\,\underline{w}^T \underline{x}^{(j)} + b^2 + \ldots]$$

$$= \underline{x}^{(i)T} \underbrace{\mathbb{E}[\underline{w}\,\underline{w}^T]}_{\sigma_w^2 \mathbb{I}} \underline{x}^{(j)} + \underbrace{\mathbb{E}[b^2]}_{\sigma_b^2} + \underbrace{\ldots}_{0}$$

$$\Rightarrow k(\underline{x}^{(i)}, \underline{x}^{(j)}) = \sigma_w^2 \underbrace{\underline{x}^{(i)T} \underline{x}^{(j)}}_{\phi(\underline{x}^{(i)})^T \phi(\underline{x}^{(j)})} + \sigma_b^2$$