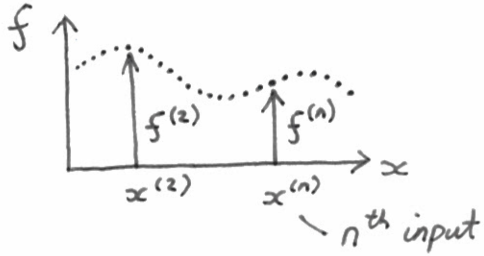
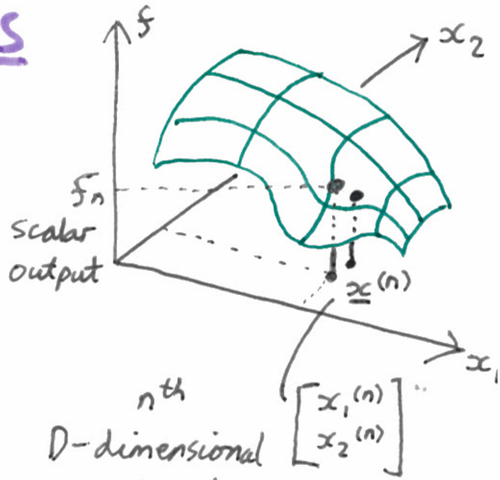
Gaussian Processes





$$\underline{f} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{bmatrix}$$ N outputs for $\{\underline{x}^{(n)}\}$ stored in X.

covariance/kernel function

If function $f \sim GP(m, k)$ $\Rightarrow$ $\boxed{p(\underline{f}) = \mathcal{N}(\underline{f}; \underline{m}, K)}$

mean function

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

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

Example:

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

We need k to always give positive semi-definite K

Things we can do with Gaussians

For a joint Gaussian

$$p(\underline{f}, \underline{g}) = \mathcal{N}\left(\begin{bmatrix}\underline{f}\\ \underline{g}\end{bmatrix}; \begin{bmatrix}\underline{a}\\ \underline{b}\end{bmatrix}, \begin{bmatrix}A & C\\ C^T & B\end{bmatrix}\right)$$

Marginals

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

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

Conditionals

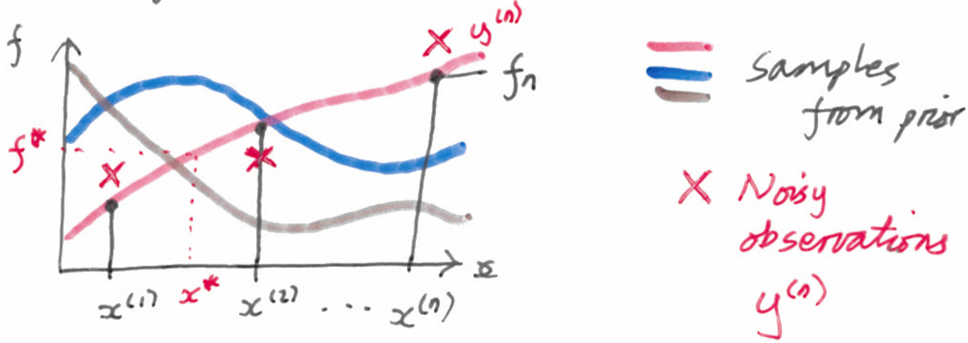
$$p(\underline{f} \mid \underline{g}) = \mathcal{N}(\underline{f}; \underline{a} + CB^{-1}(\underline{g} - \underline{b}), A - CB^{-1}C^T)$$

also

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

Regression model

Prior on functions $f \sim GP(0, k)$



Observation model:

$$y_i \sim \mathcal{N}(f_i, \sigma_y^2)$$

observation noise

Likelihood:

$$p(y_i \mid \underline{f}) = p(y_i \mid f_i) = \mathcal{N}(y_i; f_i, \sigma_y^2)$$

Posterior

$$p(\underline{f}^* \mid \underline{y}) = \ldots \text{Gaussian} \ldots \text{ need mean and cov.}$$

vector of values at test locations

The mechanical way to get

$$\underbrace{p(\underline{f}^*|\underline{y})}_{\text{Gaussian}} = \int p(\underline{f}^*, \underline{f}|\underline{y})\, d\underline{f}$$

$$= \int \underbrace{\underbrace{p(\underline{f}^*|\underline{f})}_{\text{Gaussian}}\ \underbrace{p(\underline{f}|\underline{y})}_{\propto \underbrace{p(\underline{y}|\underline{f})\,p(\underline{f})}_{\text{Gaussian}}}}_{\text{Gaussian}}\, d\underline{f}$$

Joint Distribution

$$p(\underline{y}, \underline{f}_*) = \mathcal{N}\left(\begin{bmatrix} \underline{y} \\ \underline{f}_* \end{bmatrix}; \begin{bmatrix} \underline{0} \\ \underline{0} \end{bmatrix}, \begin{bmatrix} k(X,X) + \overbrace{\sigma_y^2 I}^{\text{obs. noise}} & k(X, X_*) \\ k(X_*, X) & k(X_*, X_*) \end{bmatrix} \right)$$

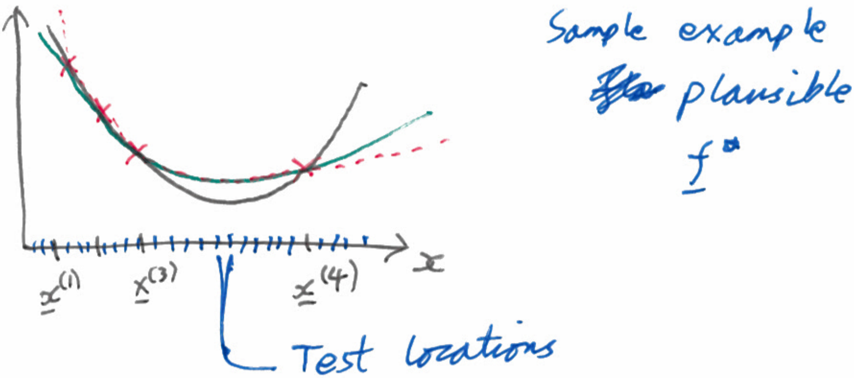
Notation:

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

$\underline{f}_*$ are f^n values at locations X_*

$\Rightarrow$ Immediately get

$$p(\underline{f}_* \mid \underline{y}) = \mathcal{N}(f_*; \; \ldots, \ldots)$$



Bayesian Linear Regression is a GP

Model

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

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

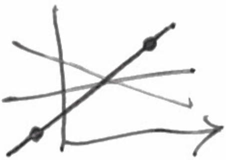
$$\text{cov}(f_i, f_j) = \mathbb{E}[f_i f_j] - \cancelto{0}{\mathbb{E}[f_i]\mathbb{E}[f_j]}$$

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

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

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

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



Prior on functions

Basis Functions

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

eg ϕ is RBFs

"Kernel trick"

- Rewrite algorithm so it only needs inner products between features.

- We'll use very large / infinite numbers of basis functions.

  → Transform features into really high dimensions

- Replace the inner products with analytic expression we can compute.

It can be shown that...

If we put RBFs <u>everywhere</u> we can find

$$k(\underline{x}^{(i)}, \underline{x}^{(j)}) = \sigma_f^2 \exp\left(-\frac{1}{2}\sum_d (x_d^{(i)} - x_d^{(j)})^2 / \ell_d^2\right)$$