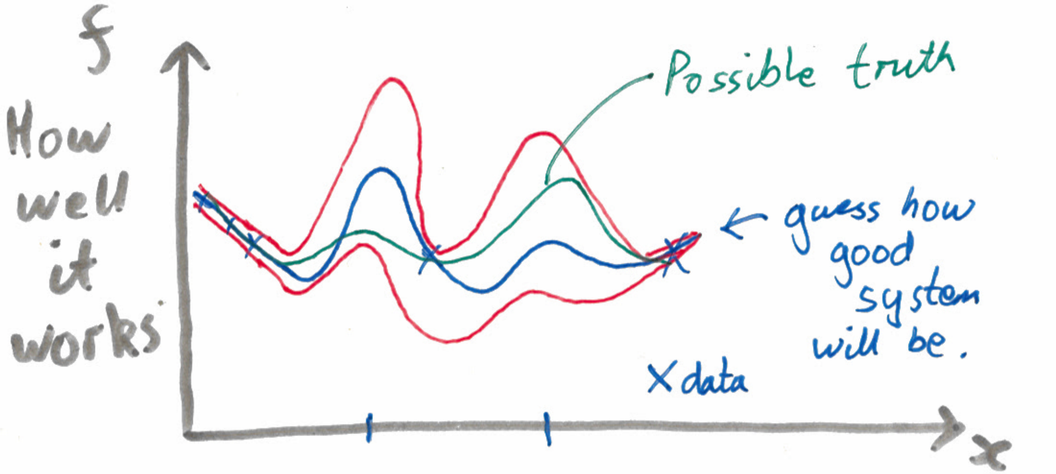
Bayesian Optimization



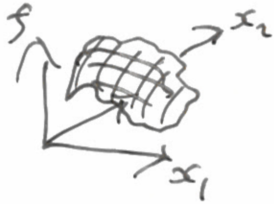
Surragote surface modelling of response surfaces.

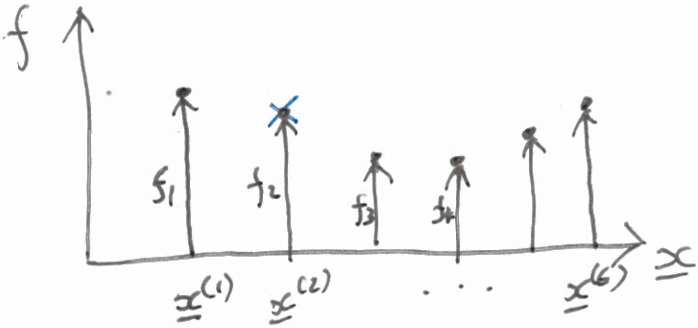
With simple linear regression:

Gaussian Processes

Really big Gaussian distribution

Functions are large vectors





$$\underline{f} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ \vdots \\ f_{lots} \end{bmatrix}$$

Gaussian process prior

$$p(\underline{f}) = \mathcal{N}(\underline{f}; \underline{0}, \Sigma)$$

$$\Sigma_{ij} = \mathrm{cov}[f_i f_j]$$
$$= \mathbb{E}[f_i f_j] - \mathbb{E}[f_i]\,\mathbb{E}[f_j]$$

Things we can do with Gaussians

For a joint Gaussian

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

Marginals:

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

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

Conditionals:

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

GP Regression

Function prior $f \sim GP$

For any subset of values $\underline{f}$

$$p(\underline{f}) = N(\underline{f}; \underline{0}, K)$$

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

kernel function

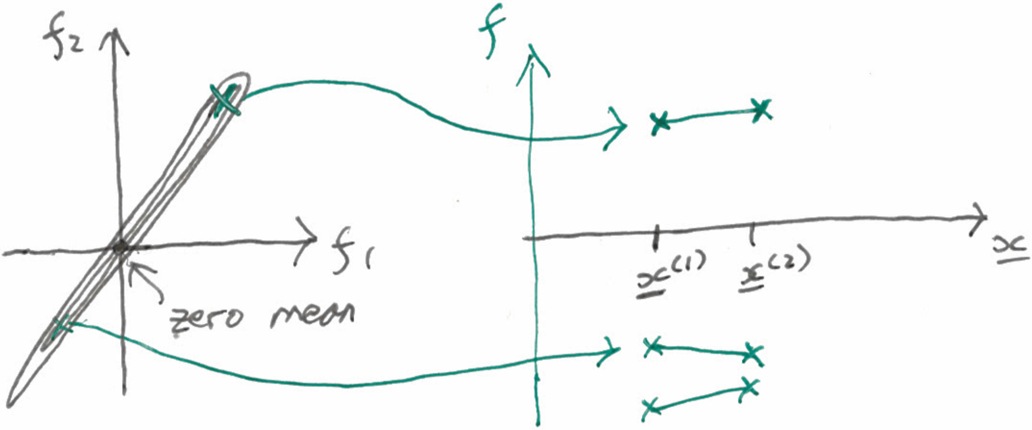
"Mercer kernels" / Positive kernels :

$\Rightarrow$ K will always be positive semi-definite

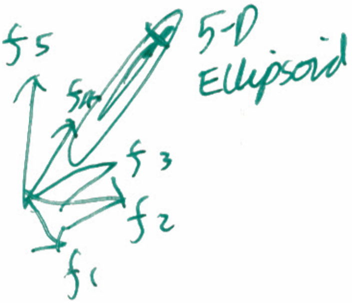
Example

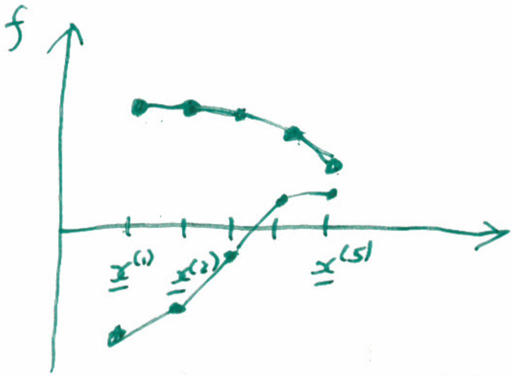
$$k(\underline{x}^{(i)}, \underline{x}^{(j)}) = \exp(-\|\underline{x}^{(i)} - \underline{x}^{(j)}\|^2)$$

$$\text{or} \quad = (1 + \|\underline{x}^{(i)} - \underline{x}^{(j)}\|)\, e^{-\|x^{(i)} - \underline{x}^{(j)}}$$



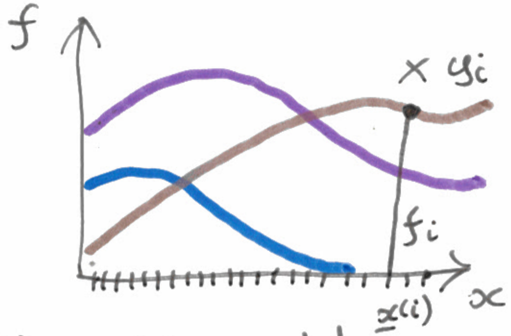
5-Dim. Gaussian





covariances fall with distance.

Sample from prior



for Gaussian kernel

Prior
$$p(\underline{f}) = N(\underline{f}; 0, K)$$

Observation model

$$y_i \sim N(f_i, \sigma_n^2)$$
noise

Likelihood

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

Posterior

$$p(\underline{f}_* \mid \underline{y}) = \text{Gaussian... need mean \& cov...}$$

Vector of values at test locations

Joint Distribution

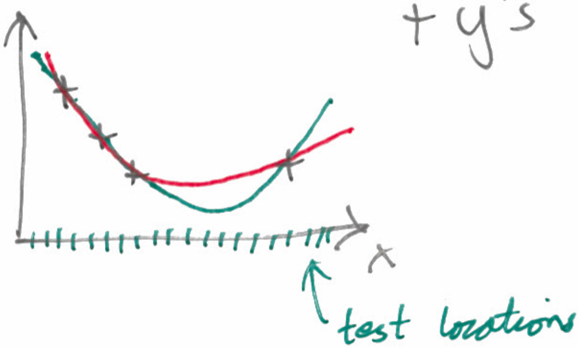
$$p(\underline{y}, \underline{f}_*) = \mathcal{N}\left(\begin{bmatrix} \underline{y} \\ \underline{f}_* \end{bmatrix}; \begin{bmatrix} \underline{0} \\ \underline{0} \end{bmatrix}, \begin{bmatrix} k(X,X)+\sigma_n^2 I & k(X,X_*) \\ k(X_*,X) & k(X_*,X_*) \end{bmatrix} \right)$$

Notation

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

Immediately get $p(\underline{f}_* | \underline{y})$

Sample from that



$+ y$'s

Error bars