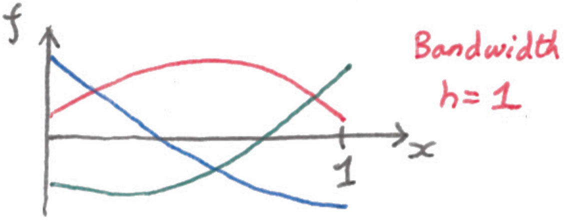
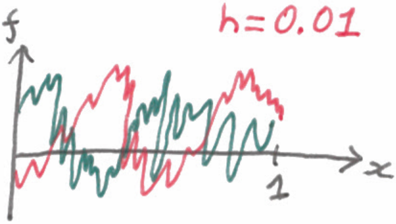
Bayesian Inference & Bandwidth

Assume $K=100$ RBF basis functions

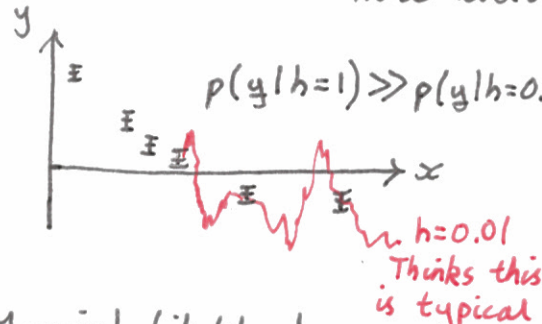
Centers evenly spaced between 0 and 1

Samples from prior $\underline{w} \sim N(\underline{0}, \mathbb{I}_K)$

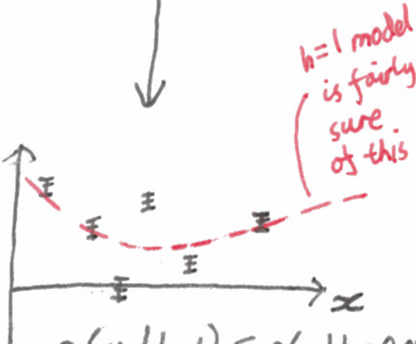




Typical observation
known noise level σ_y



$p(y|h=1) \gg p(y|h=0.01)$



$p(y|h=1) \leq p(y|h=0.01)$

Marginal likelihood

$$p(\text{data} \mid \text{model}) = \int p(\text{data} \mid \underline{w}, \text{model})\, p(\underline{w} \mid \text{model})\, d\underline{w}$$

Can compare models on training set.

Could also choose σ_y, σ_w in $\underline{w} \sim N(\underline{0}, \sigma_w^2 \mathbb{I}_K)$, ...

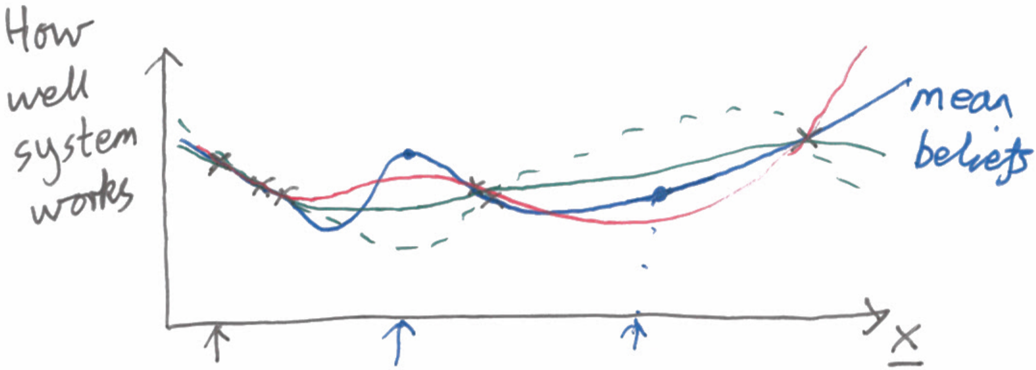
Likelihood of <u>all</u> your parameters

$$p(y \mid \underbrace{\underline{w}, h, \ldots,}_{} X)$$

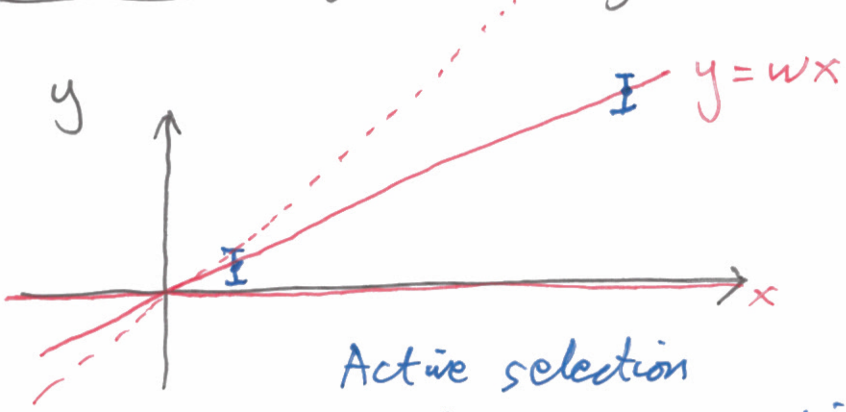
Marginal likelihood:

$$\int p(y \mid \underline{w}, h, \ldots, X)\, p(\underline{w} \mid h, \ldots, X)\, dw$$

Bayesian Optimization

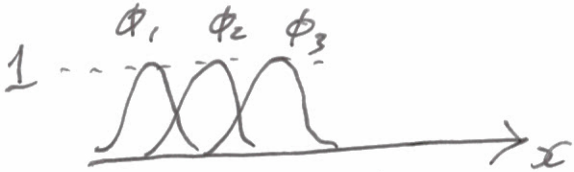


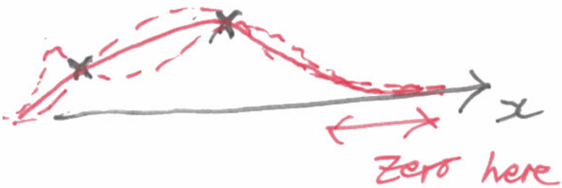
Limitations of linear regression



Active selection

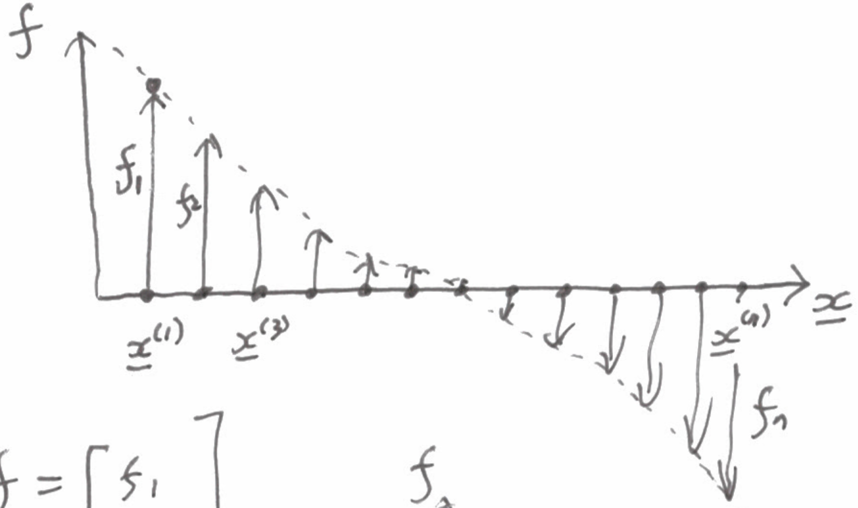
$\Rightarrow$ Make x as big as you can.



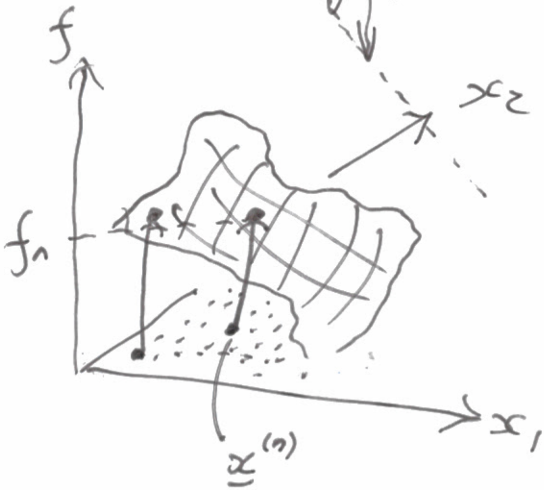


Gaussian Processes

Really <u>big</u> Gaussian distribution



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

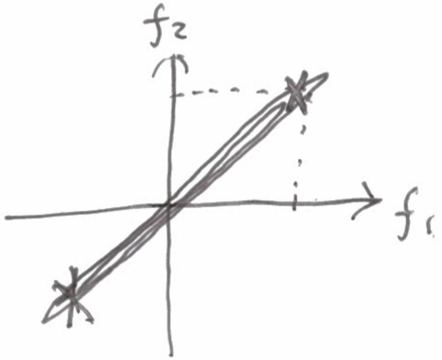


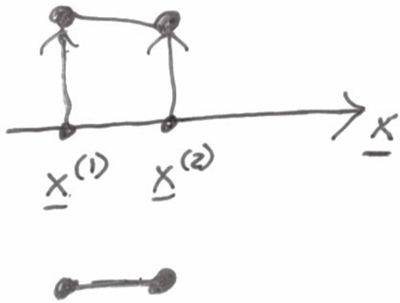
Gaussian process prior

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

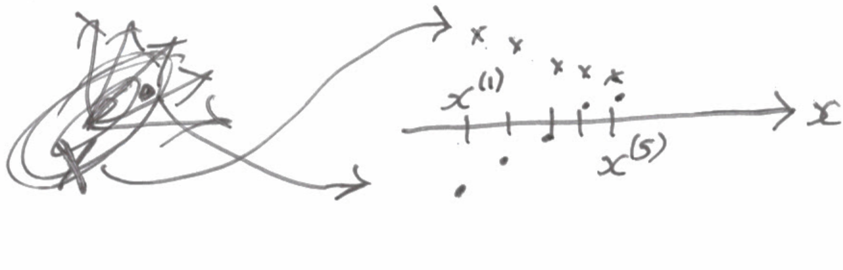
$$\Sigma_{ij} = \mathrm{cov}(f_i, f_j)$$

$$= \mathbb{E}[f_i f_j] - \cancelto{0}{\mathbb{E}[f_i]}\,\mathbb{E}[f_j]$$





5-Dim Gaussian



Covariance / kernel function

Function prior $f \sim GP$ with kernel k

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

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

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

kernel function
or covariance function

"Mercer kernels" / Positive definite functions

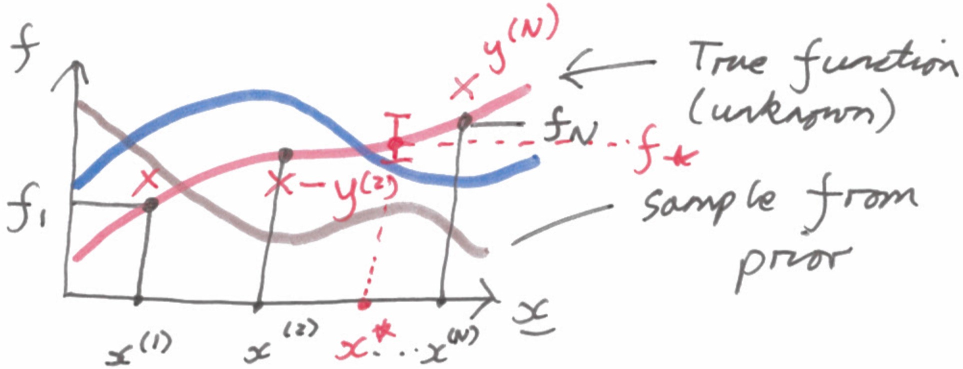
$\Rightarrow K$ +ve semi-definite

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

(an RBF)

Regression model

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



Observation model:

$$y_n \sim N(f_n, \sigma_y^2)$$

$\Rightarrow$ Likelihood $p(y_n | \underline{f}) = N(y_n; f_n, \sigma_y^2)$
product over $n = 1 \ldots N$

Posterior and Predictions

$$p(f_* | y, X)$$

$$p(y_* | y, X)$$