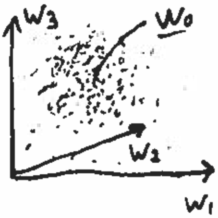
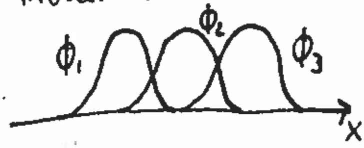
Bayesian update with basis functions

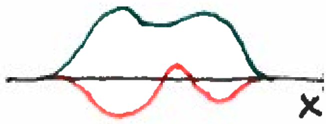
Prior: $p(\underline{w}) = \mathcal{N}(\underline{w}; \underline{w}_0, V_0)$ like before



Prior also includes model choices!



Samples from prior



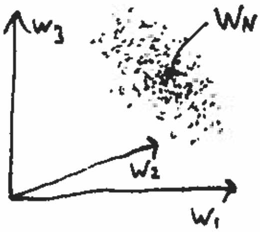
Bayes' Rule + data

$\rightsquigarrow$ Posterior $p(\underline{w}|D) = \mathcal{N}(\underline{w}; \underline{w}_N, V_N)$

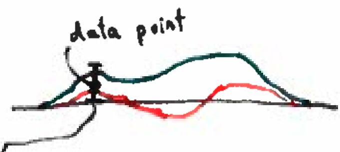
don't memorize.

$$V_N = \sigma_y^2 \left(\sigma_y^2 V_0^{-1} + \Phi^T \Phi\right)^{-1}$$

$$\underline{w}_N = V_N V_0^{-1} \underline{w}_0 + \frac{1}{\sigma_y^2} V_N \Phi^T \underline{y}$$



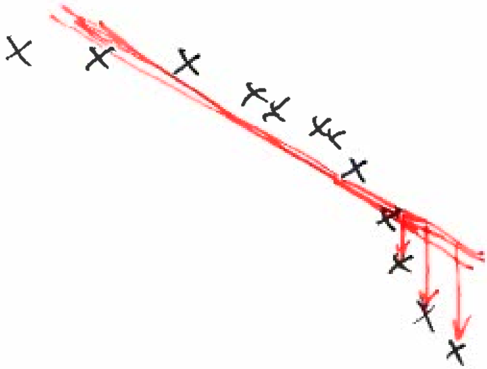
Samples from posterior $p(w|D)$



Likelihood

$$p(y|x,\underline{w}) = \mathcal{N}(y; \Phi \underline{w}, \sigma_y^2)$$

"Underfitting"



No basis functions

Simple model

⇒ over-confident

Residuals

Correlations of Residuals

⇝ Model checking

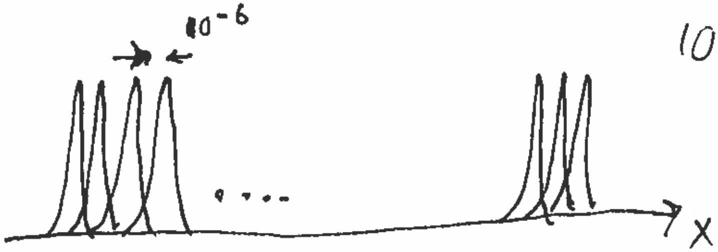
Overfitting

The Bayesian Method does not involve fitting.

⇒ Don't overfit

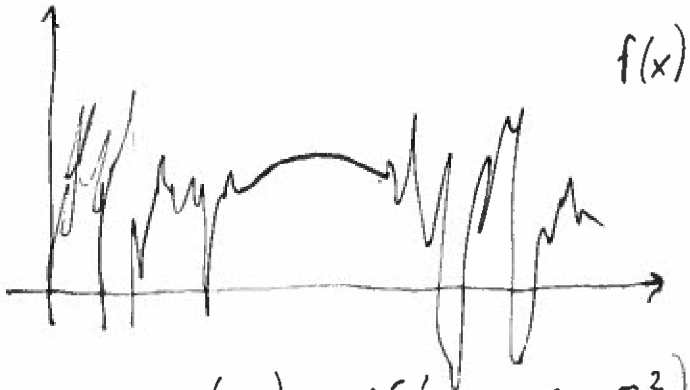
But...

Extremely flexible model



10^6 basis functions (RBFs)

Can represent the function



prior: $p(w_k) = \mathcal{N}(w_k; 0, \sigma_w^2)$

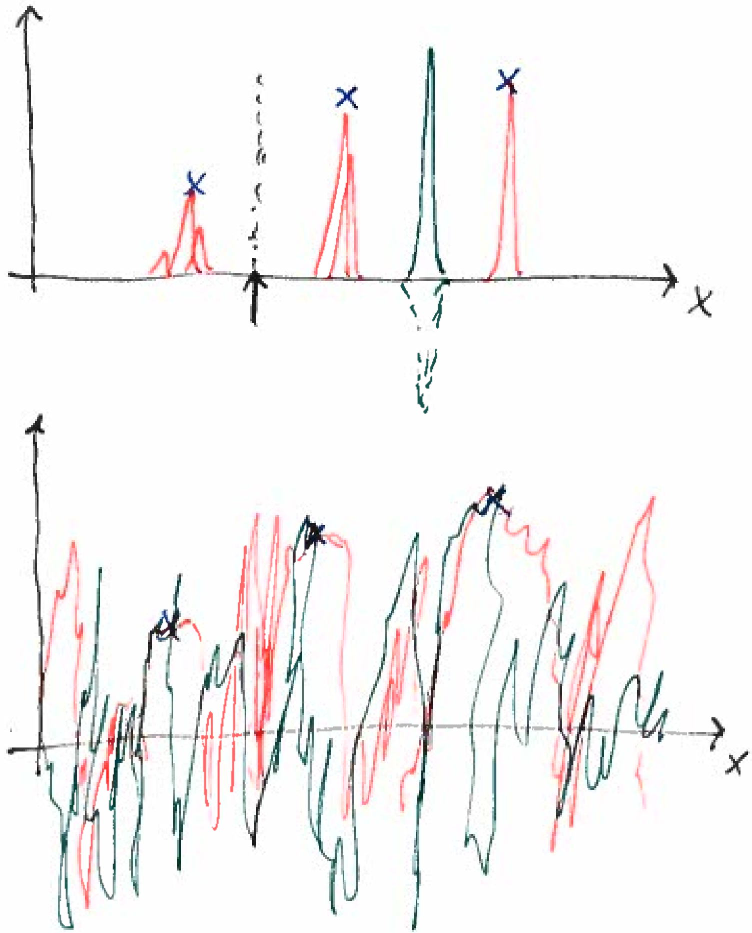
for each function k

$k = 1, \ldots, 10^6$

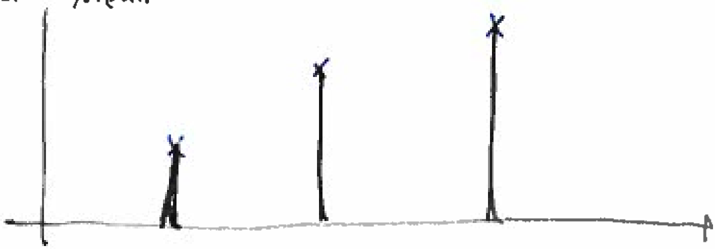
$$p(\underline{w}) = \mathcal{N}(\underline{w}; \underline{0}, \sigma_w^2 \, \mathbb{I})$$

big matrix

Posterior for the flexible model

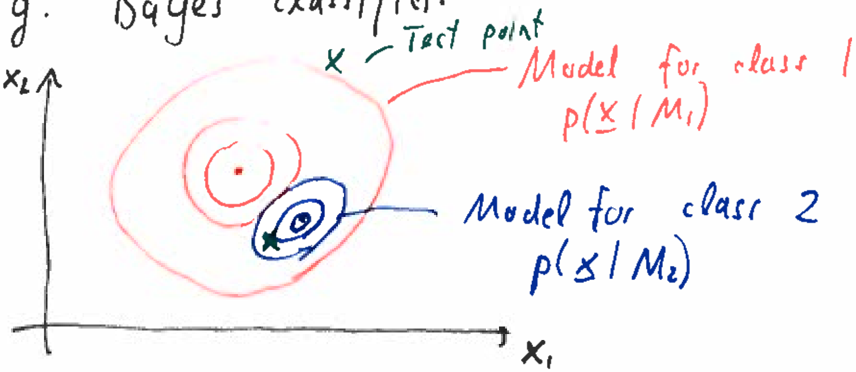


Posterior Mean



Probabilistic model choice

E.g. Bayes classifiers



Used likelihood of model fitted

Model choice for regression

$p(\underline{y} \mid X, M)$ — Model choices: $\sigma_y^2, V_0 = \sigma_w^2 \mathbb{I}, w_0 = 0, \Phi, \dots$

Training inputs (X)

All training labels ($\underline{y}$)

"Marginal likelihood"

or just "likelihood" of the model

Could apply:

① Bayes' rule $\rightsquigarrow p(M \mid \underline{y}, X)$

② Maximise likelihood

$$p(y \mid X, M) = \underbrace{\int p(y, \underline{w} \mid X, M)\, d\underline{w}}_{\text{Sum rule}}$$

$$= \int \underbrace{p(y \mid X, w, M)}_{\text{"Standard" likelihood}} \underbrace{p(w \mid X, M)}_{\text{Prior}}\, d\underline{w}$$

Trick to solve integral:

Posterior weights:

known

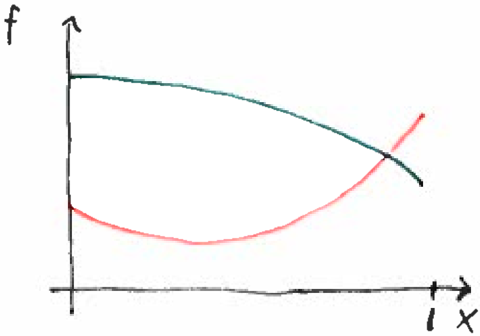
$$\underbrace{p(\underline{w} \mid y, X, M)}_{p(\underline{w} \mid D)} = \frac{p(y \mid \underline{w}, X, M) \cdot p(\underline{w} \mid X, M)}{p(y \mid X, M)}$$

Rearrange and evaluate

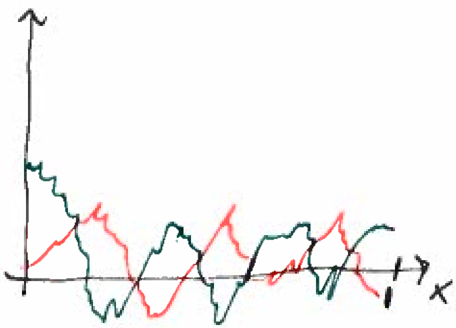
Example model choice

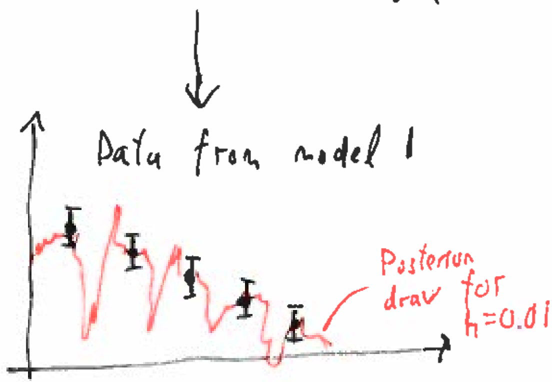
100 basis functions, spaced between 0, 1

Model 1: Bandwidth $h=1$



Model 2: $h=0.01$





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

$$p(y|X, h=0.01, \hat{w}) \text{ big}$$