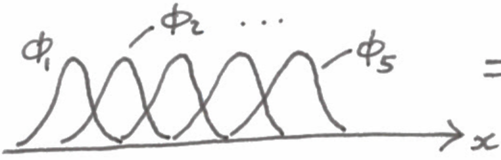
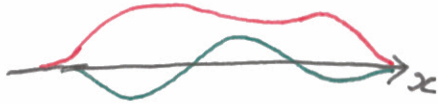
Bayesian update: prior → posterior

Prior: model choices + $p(\underline{w}) = \mathcal{N}(\underline{w};\, \underline{w}_0, V_0)$



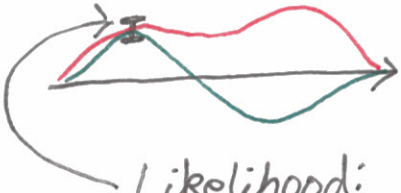
$\Rightarrow$

Samples from prior:



$\Downarrow$ Observe data

Posterior



Samples from posterior $p(\underline{w} \mid D)$

Likelihood:

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

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

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

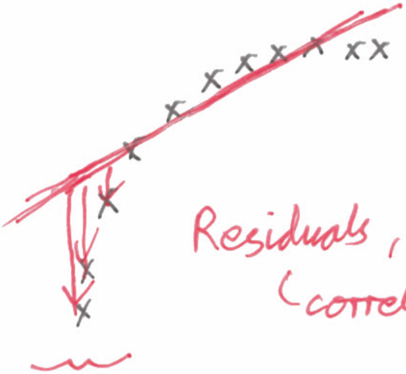
$$\underline{w}_N = V_N V_0^{-1} \underline{w}_0 + \frac{1}{\sigma_y^2} V_N \Phi^T y$$

"Underfitting"

No basis $f^{\underline{n}}$'s

Simple model

$\Rightarrow$ over-confident



Residuals, model checking

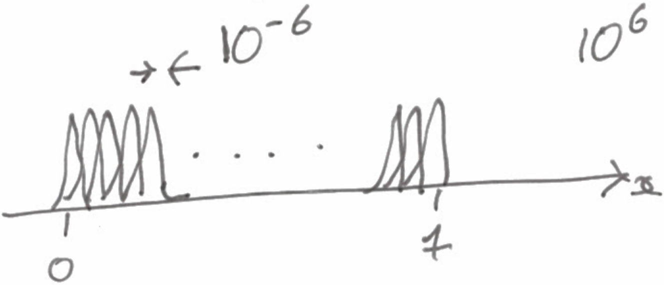
correlated

Overfitting

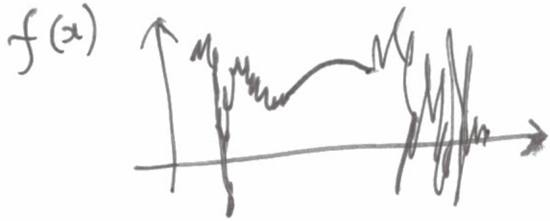
Bayesians don't fit.

Can't overfit.

Extremely flexible model:



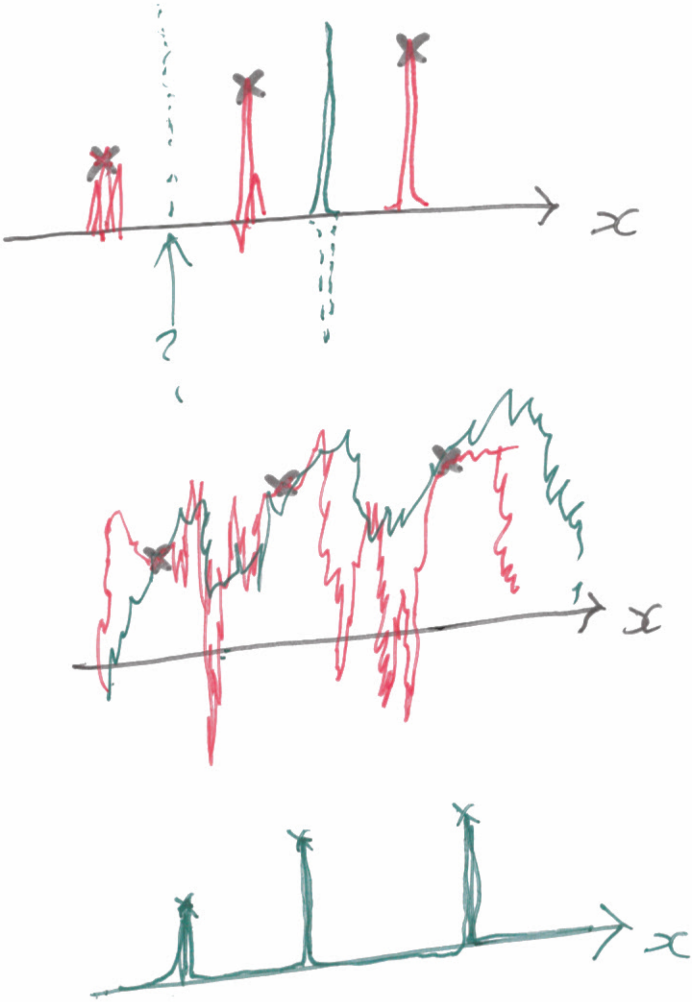
Can model



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

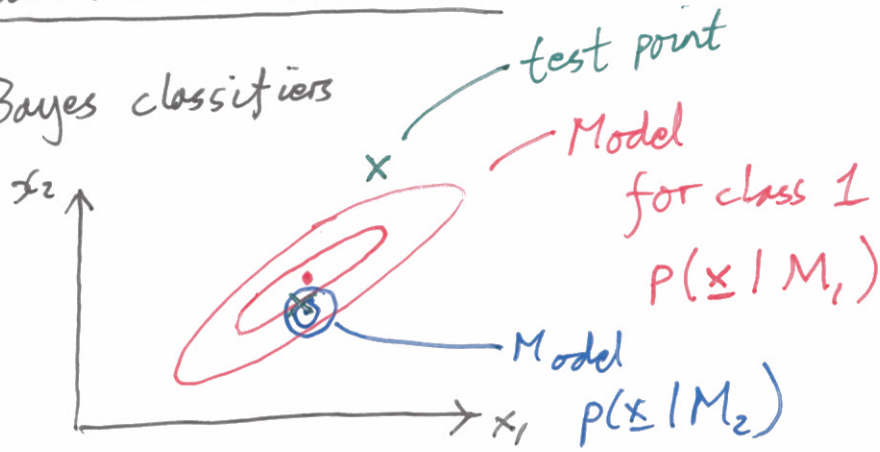
$$p(\underline{w}) = \mathcal{N}(\underline{w} ; \underline{0}, \sigma_w^2 \underline{I})$$

What's the posterior?



Probabilistic model choice

Bayes classifiers



Model choice regression

model choices

$$p(\underline{y} \mid X, M)$$

train inputs

all training labels

$\sigma_y^2, V_0 = \sigma_w^2 \mathbb{I}, \underline{\Phi}, \ldots$

(Marginal) Likelihood model

1. Bayes rule $p(M \mid \underline{y}, X)$
2. Maximize likelihood.

$$p(y \mid X, M) = \int p(y, \underline{w} \mid X, M)\, d\underline{w}$$

(sum rule)

Params of model

(product rule)

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

maybe.

Prior

Posterior weights

$$p(\underline{w} \mid y, X, M) = \frac{p(y \mid \underline{w}, X, M)\, p(\underline{w} \mid \cancel{X}, M)}{p(y \mid X, M)}$$

Gaussian we know.

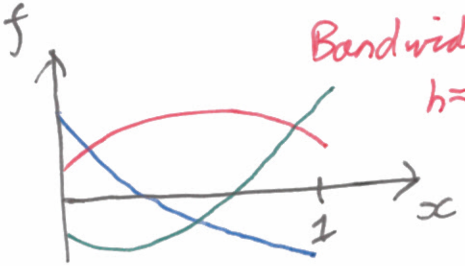
$\Longrightarrow$ Rearrange and evaluate

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

True for all $\underline{w}$ $\rightarrow$ so pick any setting you like. eg $\underline{w} = \underline{0}$

Example model choice, 100 RBFs spaced between 0 and 1

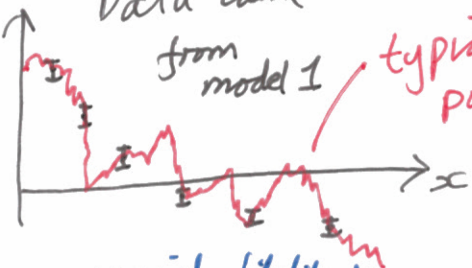
Bandwidth $h = 1$



Model 1

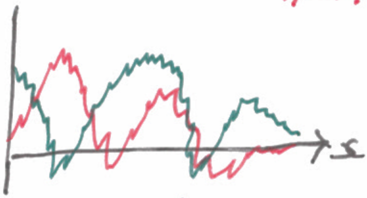
$\downarrow$

Data came from model 1



typical posterior draw for $h = 0.01$

$h = 0.01$



$\downarrow$

typical explanation for $h = 1$

Compare marginal likelihood:

$$p(\underline{y} \mid X, h = 1) \gg p(\underline{y} \mid X, h = 0.01)$$

Earlier in the course we compared likelihood:

$p(\underline{y} \mid X, h = 0.01, \hat{\underline{w}})$ big, can't use for model comparison