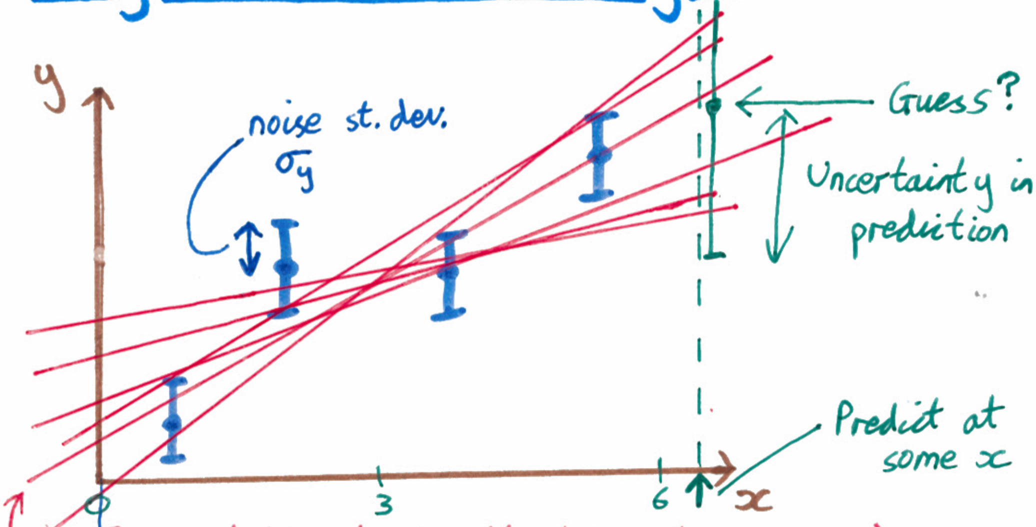
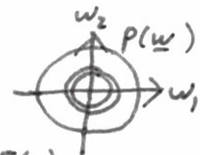
Bayesian Linear Regression

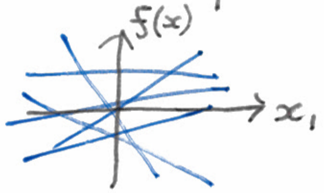


Some different plausible lines $\underbrace{(w_1 x + w_2 x_2)}_{\mathbf{w}^\top \mathbf{x},\ \text{with } x_2 = 1}$

Prior (Example) $p(\mathbf{w}) = N(\mathbf{w}; \mathbf{0}, \sigma_w^2 \mathbb{I})$



$\Rightarrow$ Broad range functions plausible before see data

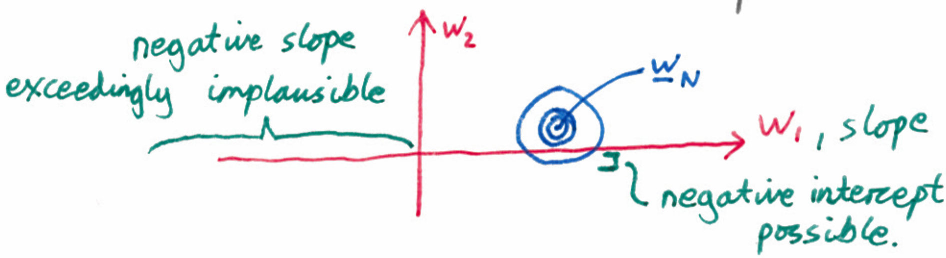


Posterior

$$p(\mathbf{w} \mid \overset{\text{"data"}}{\mathcal{D}}) \propto p(\mathbf{w}) \overbrace{p(\mathbf{y} \mid X, \mathbf{w})}^{\text{likelihood}}$$

$$= N(\mathbf{w}; \mathbf{w}_N, V_N)$$

(for Gaussian prior and noise)



Probabilistic Prediction

$$f(\underline{x}) = \underline{w}^\top \underline{x} = \underline{x}^\top \underline{w}$$

$$p(\,f(\underline{x}) \mid \text{Data}\,) = N(\,f;\ \underline{w}_N^\top \underline{x},\ \ \underline{x}^\top V_N \underline{x}\,)$$

$$p(\,y \mid \text{Data}\,) = N(\,y;\ \ ''\ ,\ \ ''\ + \sigma_y^2\,)$$

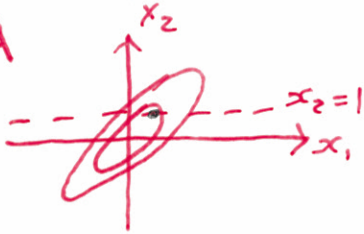
Questions

Uncertainty $\underline{x}^\top V_N \underline{x}$ grows with $\underline{x}$

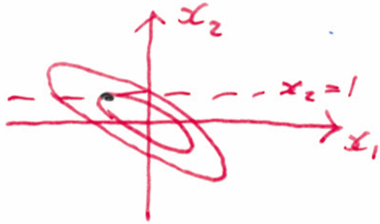
① Why in figure is most certain region at $x > 0$ (around $x = 3$)?

② What do contours of $\underline{x}^\top V_N \underline{x}$ look like?

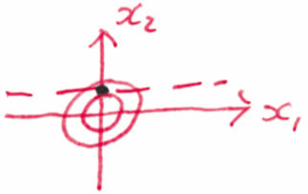
A



B



C

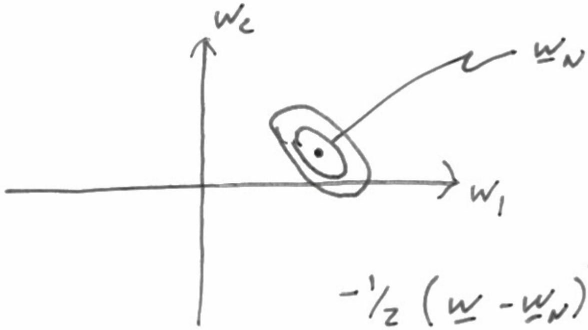


D Other

Z ???

V_N was posterior covariance of $\underline{w}$

Contours $N(\underline{w}\,;\,\underline{w}_N,\,V_N)$



$$-\tfrac{1}{2}(\underline{w}-\underline{w}_N)^T V_N^{-1}(\underline{w}-\underline{w}_N)$$

Decisions — Loss function

$$L(y, \hat{y})$$

Loss   what happens   "point estimate"/ "guess"

Minimize expected loss

$$c = \int L(y,\hat{y})\, p(y \mid \text{Data})\, dy$$

$$= \mathbb{E}_{p(y \mid \text{Data})}\left[L(y, \hat{y} \right]$$

Find $\hat{y}$ that minimizes c

Eg square loss $L(y, \hat{y}) = (y - \hat{y})^2$

$$\frac{\partial c}{\partial \hat{y}} = \mathbb{E}\left[2(y - \hat{y}) \right]$$

$$= 0 \text{ if } \mathbb{E}[y] = \hat{y}$$

Estimate $\hat{y}$ = mean belief.

Overfitting

Bayesian don't fit

We don't do $\cancel{\hat{\underline{w}} = \underset{\underline{w}}{\arg\min}\ \text{cost}(\underline{w})}$

Compute beliefs $p(\underline{w} \mid D)$ . . . decision .

"Underfitting"

Over simple models

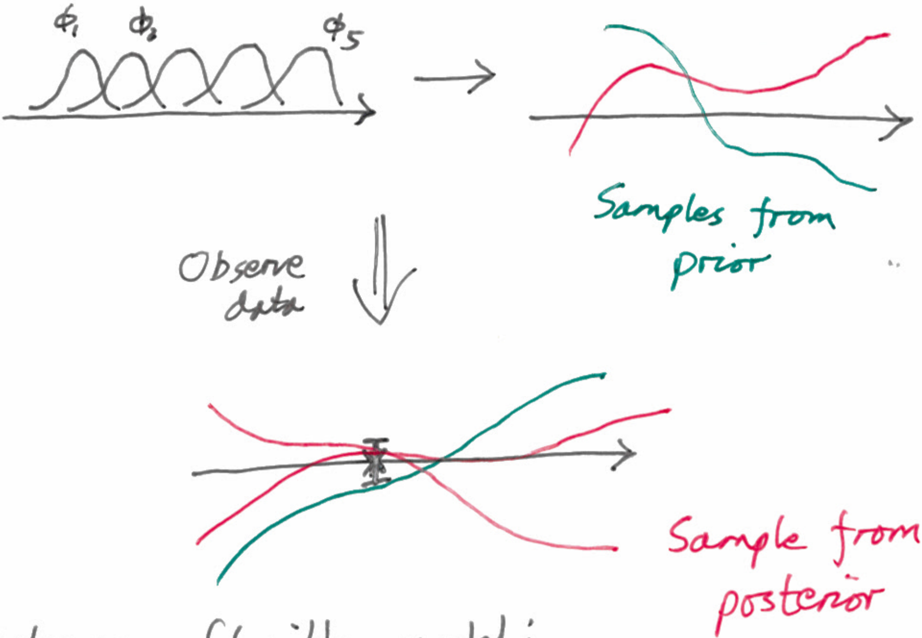
$\Rightarrow$ Over-confident .

Residuals

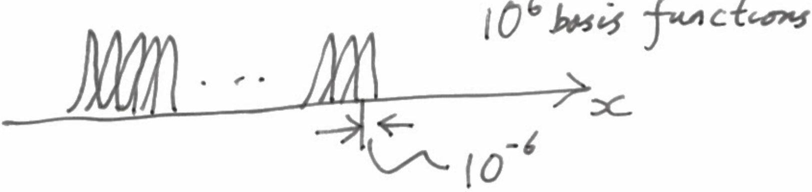
→ Tell us things are wrong

"Model checking"

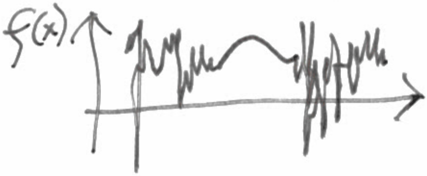
Bayesian methods with lots parameters



Extreme flexible model:



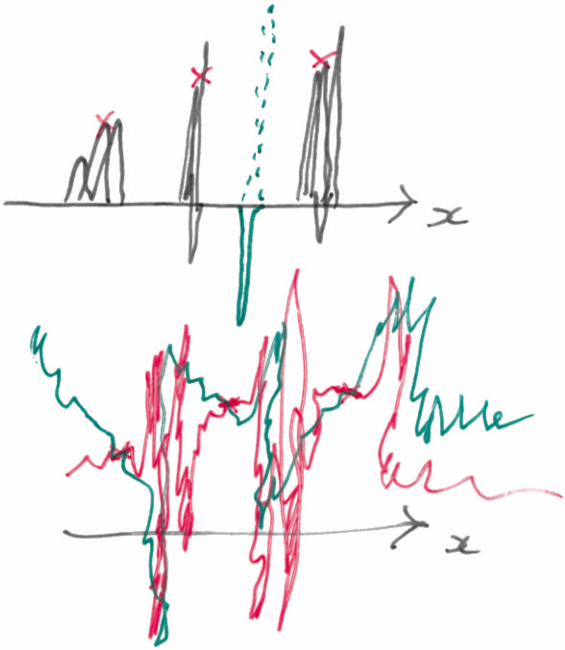
Can model:



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

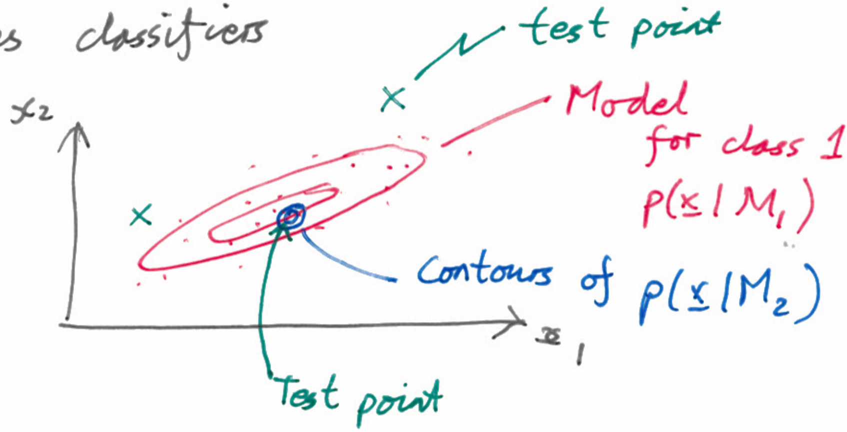
Independent $p(\underline{w}) = \mathcal{N}(\underline{w}; \underline{0}, \sigma_w^2 \mathbb{I})$

What's the posterior?



Probabilistic model choice

Bayes classifiers



Regression

a regression model class.

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

all Training labels

All training inputs

parameters of model

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

(usually)

Marginal Likelihood of Model, scores how good model is.