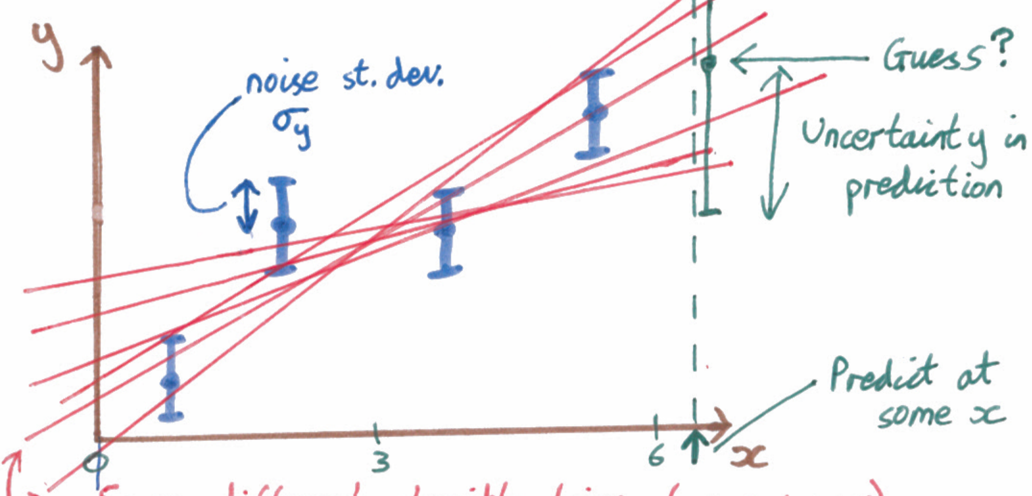
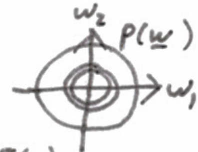
Bayesian Linear Regression



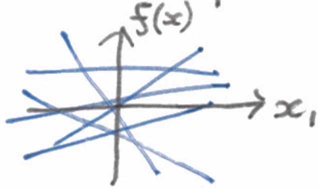
Some different plausible lines $(w_1 x + w_2 x_2)$

$\underbrace{w_1 x + w_2 x_2}_{\mathbf{w}^\top \mathbf{x},\ \text{with } x_2 = 1}$

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



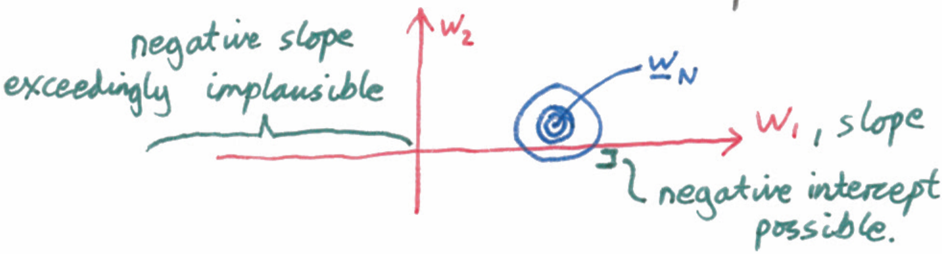
$\Rightarrow$ Broad range functions plausible before see data



Posterior

$p(\mathbf{w} \mid \underbrace{D}_{\text{"data"}}) \propto p(\mathbf{w}) \underbrace{p(\mathbf{y} \mid X, \mathbf{w})}_{\text{likelihood}}$

$= \mathcal{N}(\mathbf{w};\, \mathbf{w}_N,\, V_N)$ (for Gaussian prior and noise)



Mid-Semester Survey

Thanks for the useful responses!

tinyurl.com/

mlpr2018mss

(I will reply next week.)

Assignment 2

Due 20 Nov
(3 days before MLP)

Shouldn't take that long.

→ Do it next week?

Compute posterior

$$p(\underline{w} \mid D) \propto p(\underline{w})\, \underbrace{p(D \mid \underline{w})}_{p(y \mid \underline{w}, X)}$$

Data, y for regression

Assume X known.

$$\propto \mathcal{N}(\underline{w}; \underline{0}, \sigma_w^2 \mathbb{I}) \cdot \mathcal{N}(y; \underline{f}, \sigma_y^2 \mathbb{I})$$

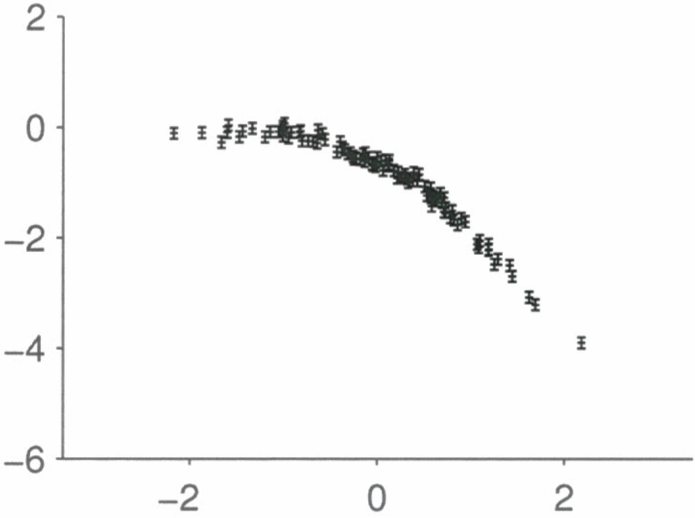
$\underline{f}$ = $\Phi \underline{w}$

$$\propto e^{-\frac{1}{2}(\text{some quadratic } \underline{w})}$$

$$\propto \mathcal{N}(\underline{w}; \sqcup, \sqcup)$$

Linear Algebra Exercise
(Cf Tut 2 Q2)

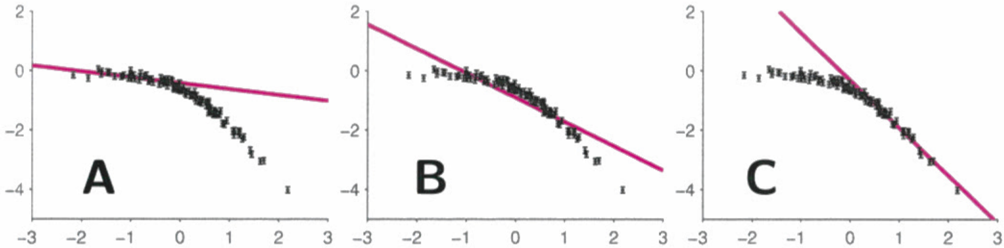
Model mismatch



What will Bayesian linear regression do?

Quiz

Given a (wrong) linear assumption, which explanations are typical of the posterior distribution?

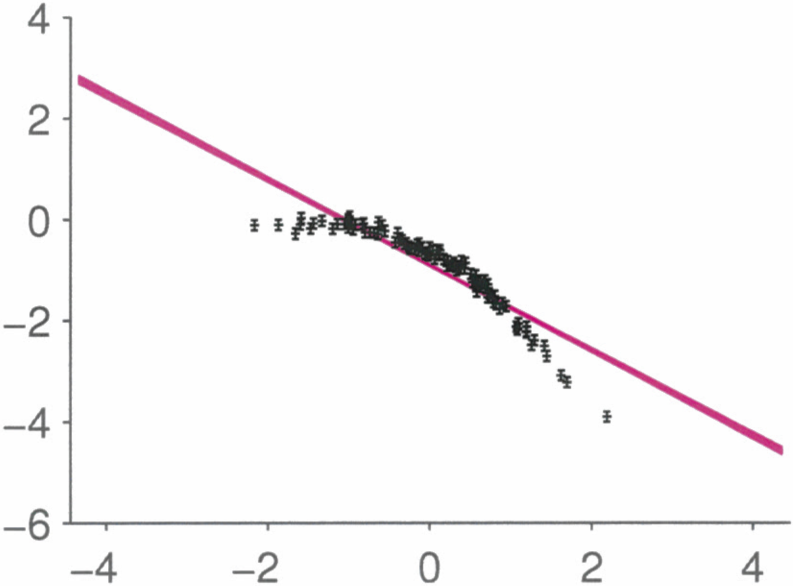


D All of the above

E None of the above

Z Not sure

'Underfitting'



Posterior *very* certain despite blatant misfit. Peaked around least bad option.

3 cards

W | B   B / B   W / W

1   2   3

Picked 1 at random

Way up is random

Observation $x_1 = B$ (Black face)

Q) $P(x_2 = W \mid x_1 = B)$?

other side of same card

A) $1/3$ B) $1/2$ C) $2/3$ D) Other

Z) Don't know.

What not to do

$$P(x_2 = w \mid x_1 = B) = \frac{\boxed{P(x_1 = B \mid x_2 = w)}\, P(x_2 = w)}{P(x_1 = B)}$$

Don't know

First step: write down model

Picked a card, c:

$$P(c) = \begin{cases} 1/3 & c=1 \quad W/B \\ 1/3 & c=2 \quad B/B \\ 1/3 & c=3 \quad W/W \end{cases}$$

Observed face 1:

$$P(x_1 = B \mid c) = \begin{cases} 1/2 & c=1 \\ 1 & c=2 \\ 0 & c=3 \end{cases}$$

Inference

$$P(c \mid x_1 = B) \propto p(x_1 = B \mid c)\, P(c)$$

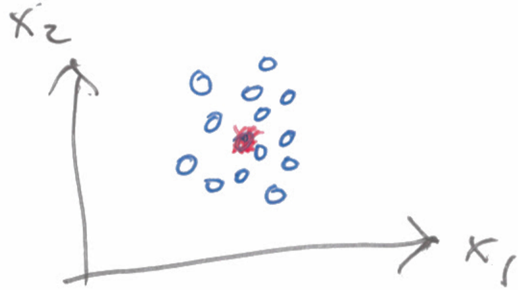
$$\propto \begin{cases} 1/2 & c=1 \\ 1 & c=2 \\ 0 & c=3 \end{cases} = \begin{cases} 1/3 & c=1 \\ 2/3 & c=2 \end{cases}$$

Another example

6-sided Dice　　10-sided Dice　　100-sided Dice

Pick random die

Roll it → got an 8



Making a Prediction

$$P(x_2 = W \mid x_1 = B) = \sum_{c \in \{1,2,3\}} P(x_2 = W, c \mid x_1 = B)$$

(Sum Rule)

$$= \sum_c P(x_2 = W \mid c, x_1 = B)\, P(c \mid x_1 = B)$$

(Product Rule)

$$= \underline{\underline{1/3}}$$

Prediction for Linear Regression

$$p(y \mid \underline{x}, D) = \int p(y, \underline{w} \mid \underline{x}, D)\, d\underline{w}$$

(Sum rule)

$$= \int \underbrace{p(y \mid \underline{w}, \underline{x}, \cancel{D})}_{N(y; \underline{w}^T \underline{x}, \sigma_y^2)}\, p(\underline{w} \mid D, \cancel{\underline{x}})\, d\underline{w}$$

(Product Rule)

For standard regression.