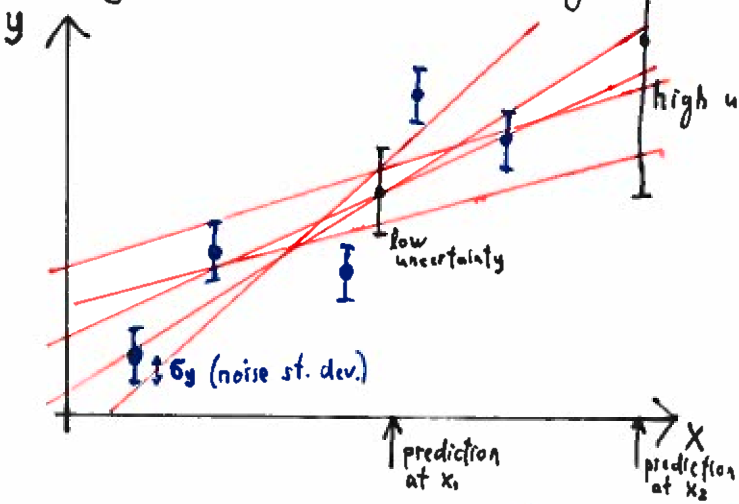
Bayesian Linear Regression

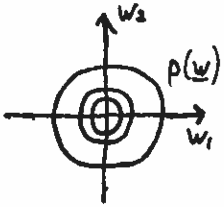


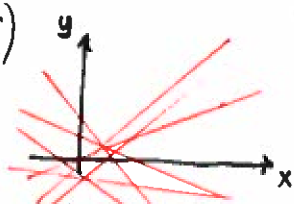
$f(\underline{x}; \underline{w}) = \underline{w}^T \underline{x}$, here $x_2 = 1$; different plausible lines

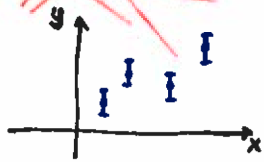
Observation model : $p(y \mid \underline{w}, \underline{x}) = \mathcal{N}(y ; f(\underline{x}; \underline{w}), \sigma_y^2)$

Prior over models : Expressing our belief

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





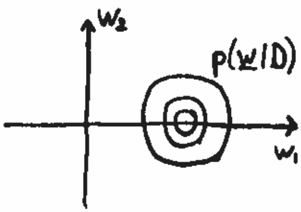


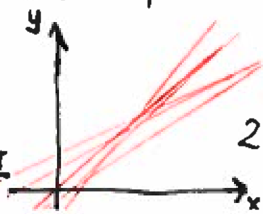
Bayes' rule + data D

→ Posterior $p(\underline{w} \mid D) \propto p(\underline{w})\, p(\underline{y} \mid X, \underline{w})$

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

for Gaussian prior and noise



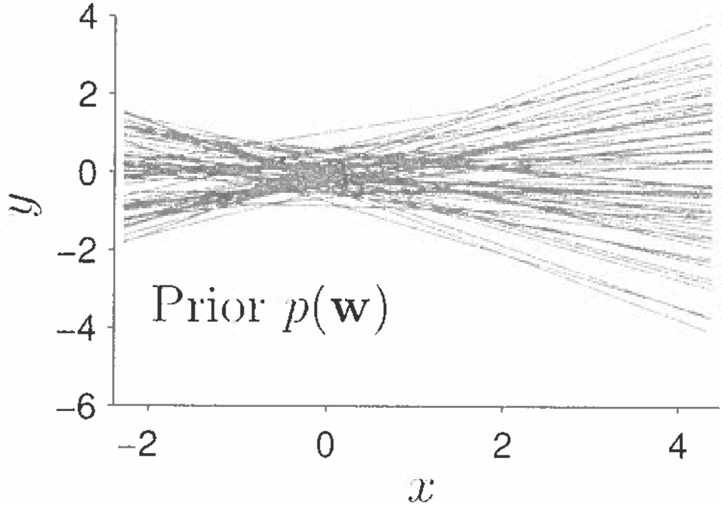


Posterior is a distribution over models

2019 L10 1

Linear regression

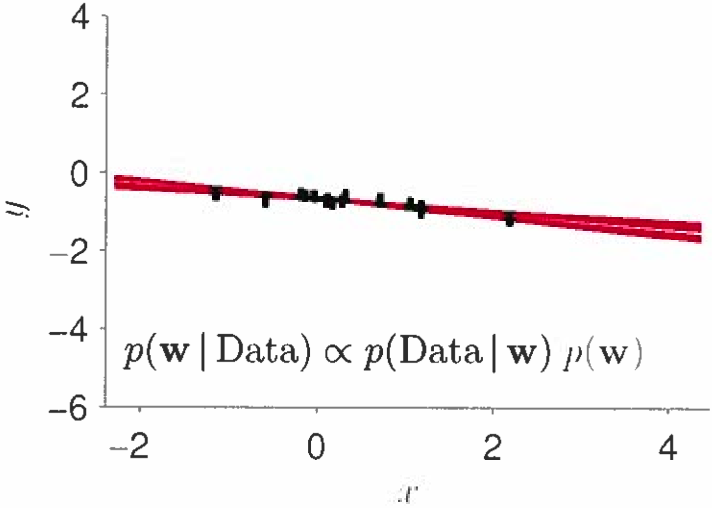
$y = w_1 x + w_2, \qquad p(\mathbf{w}) = \mathcal{N}(\mathbf{w}; 0, 0.4^2 I)$



2019 L10 ②

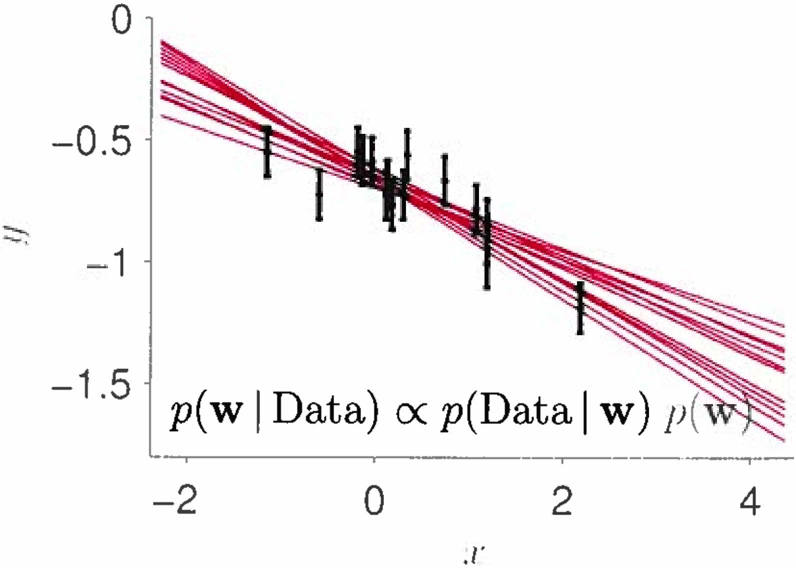
Linear regression

$y^{(n)} = w_1 x^{(n)} + w_2 + \epsilon^{(n)}, \qquad \epsilon^{(n)} \sim \mathcal{N}(0, 0.1^2)$



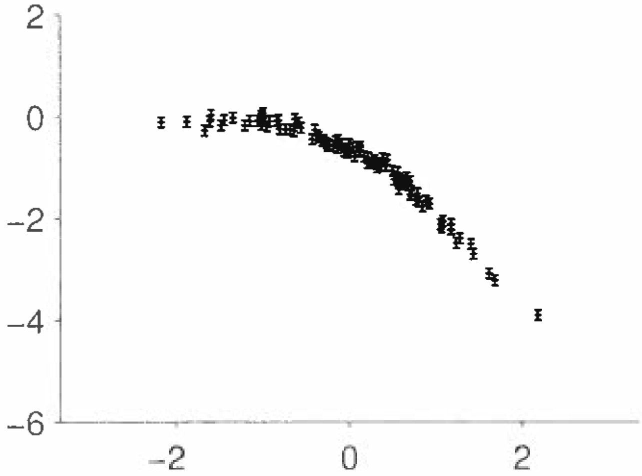
2019 L10 3

Linear regression (zoomed in)



2019 L10 4

Model mismatch

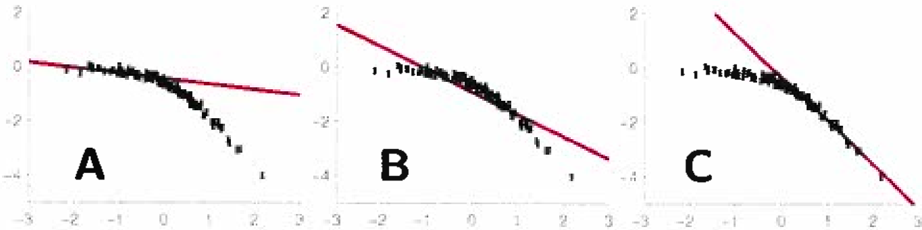


What will Bayesian linear regression do?

2019 L10 5

Quiz

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



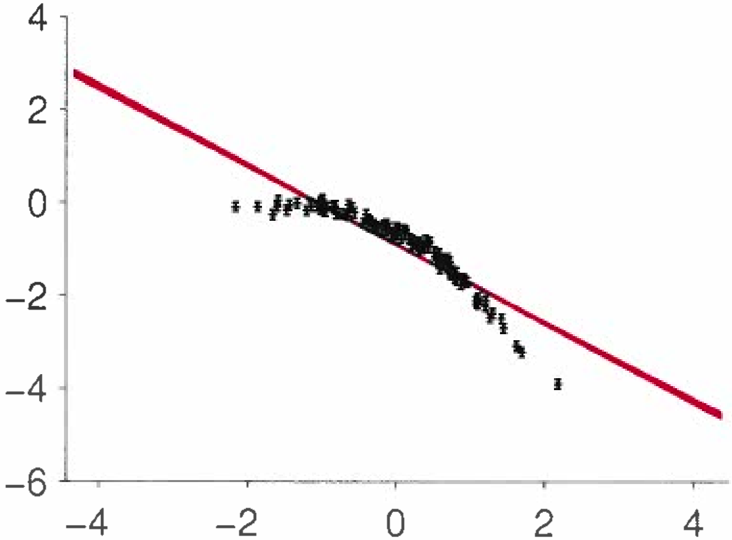
D All of the above

E None of the above

Z Not sure

2019 L10 6

'Underfitting'



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

2019 L10 7

3 cards

W | B   B | B   W | W

1   2   3

Picked one card at random

Way up is also random

Observation $x_1 = B$ (Black face)

Q) $P(X_2 = W \mid x_1 = B)$?

other side of the card

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

Z) Don't know.

Applying Bayes' Rule does not give you the answer right away

$$P(x_2 = W \mid x_1 = B) = \frac{\boxed{P(x_1 = B \mid x_2 = W)} \cdot P(x_2 = W)}{P(x_1 = B)}$$

? same problem

First step: Formalise the model

Picked a card, c

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

Observed face x_1:

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

Second step: inference

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

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

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

A more extreme example with dice

6-sided dice

10-sided dice

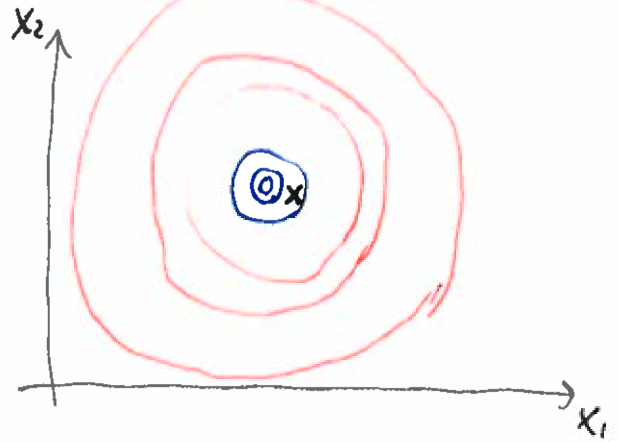
20-sided dice

Pick a dice at random

Let's say we got an 8

→ belief more towards 10-sided die

Another example



Prediction for the card example

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

(Sum Rule)

$$= \sum_{c} P(x_2 = W \mid c, x_1 = B) \cdot P(c \mid x_1 = B)$$

(Product Rule)

$$= 1/3$$

Prediction for linear regression

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

(Sum Rule)

$$= \int \underbrace{p(y \mid \underline{w}, \underline{x}, \cancel{D})}_{\substack{\mathcal{N}(y;\, \underline{w}^T \underline{x},\, \sigma_y^2) \\ \text{observation model}}}\ \underbrace{p(\underline{w} \mid \cancel{\underline{x}}, D)}_{\substack{\mathcal{N}(\underline{w};\, \underline{w}_N, V_N) \\ \text{Posterior}}}\, d\underline{w}$$

(Product Rule)

for standard linear regression