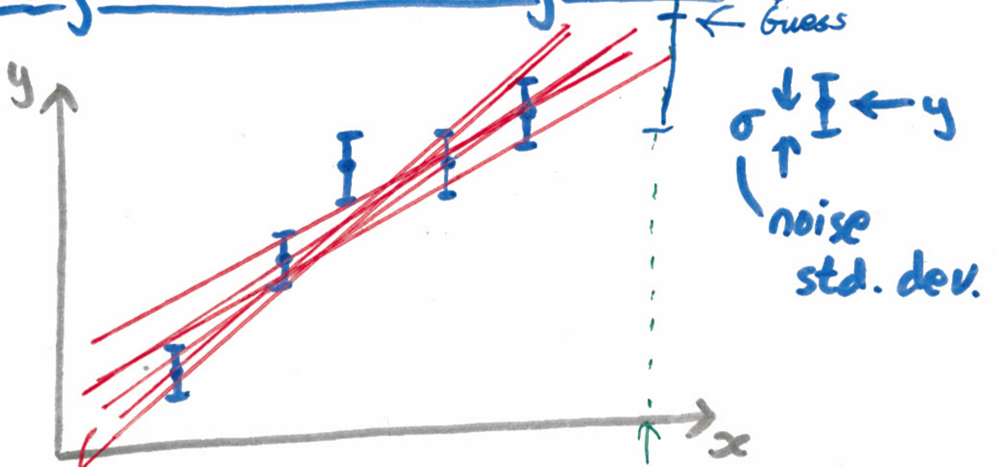
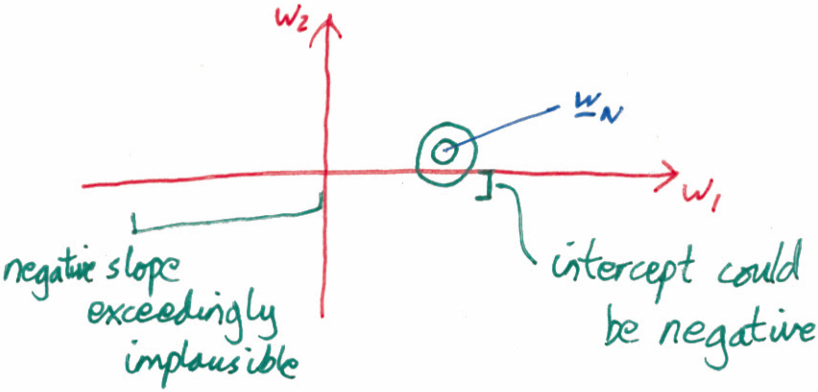
Bayesian Linear Regression



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

Posterior distribution gives plausibility of fits:



B | B B | W W | W 3 cards

2 1 3

Pick card at random

Pick a random side

Oberve $\underbrace{x_1}_{\text{observation of side 1}} = B$

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

Prob. other side of card is white

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

Z) ~~to~~ Don't know

What not to do

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

Don't know!

First step: write down model

Picked a card:

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

Observed face 1:

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

Inference

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

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

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

Aside: another example

6 sided Dice

D10
Ten sided Dice

D100
100-sided Dice.

Pick random dice:

Roll it → get 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)

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

(Product Rule)

$$= 1/3$$

Prediction for Linear Regression

What's

$$p(y \mid \underline{x}, D) \quad D = \underbrace{\{X, y\}}_{\text{Train. Data}}$$

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

(sum rule)

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

(Product rule)

$$N(y;\ \underline{w}^T \underline{x},\ \sigma^2) \qquad N(\underline{w};\ \underline{w}_N,\ V_N)$$

σ^2: noise var.

$$p(y \mid \underline{x}, D) = \int \underbrace{p(y, \underline{w} \mid \underline{x}, D)}_{\text{Joint Gaussian on } y, \underline{w}} d\underline{w}$$

$$= \mathcal{N}\left(\begin{bmatrix} \underline{w} \\ y \end{bmatrix} ; \begin{bmatrix} \underline{w}_N \\ \mu_y \end{bmatrix}, \begin{bmatrix} V_N & \Sigma_{\underline{w},y} \\ \Sigma_{y,\underline{w}} & \sigma_y^2 \end{bmatrix} \right)$$

$$= \mathcal{N}(y ; \underbrace{\mu_y, \sigma_y^2}_{\text{Lots of work to identify these.}})$$

...

YUCK!

Instead

$$y = f(\underline{x}) + v, \qquad v \sim N(0, \sigma^2)$$

what do we believe about f^{n} value

$$f = \underline{w}^T \underline{x} = \underline{x}^T \underline{w}$$

Beliefs about

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

Beliefs are Gaussian

$$p(f | D, \underline{x}) = N(f; \underline{x}^T \underline{w}_N; \underline{x}^T V_N \underline{x})$$

Beliefs about prediction

$$p(y | D, \underline{x}) = N(y; \underline{x}^T \underbrace{\underline{w}_N}_{\text{Posterior mean weights}}; \underline{x}^T V_N \underline{x} + \sigma^2)$$