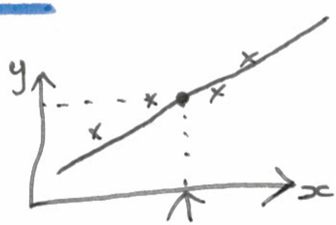
Bayesian Regression



Previously:

fit functions

$f(x)$ is a single guess of what the output will be.

For classification we fitted

$P(y \mid \underline{x})$ by max likelihood.

For regression we can also write down a probabilistic model eg $\underline{w}^T\underline{x}$

$$p(y \mid \underline{x}) = N(y;\ f(\underline{x};\underline{w}),\ \sigma^2)$$

noise variance

or σ_n^2 ← "noise"

For now assume the same for each example.

Maximum Likelihood

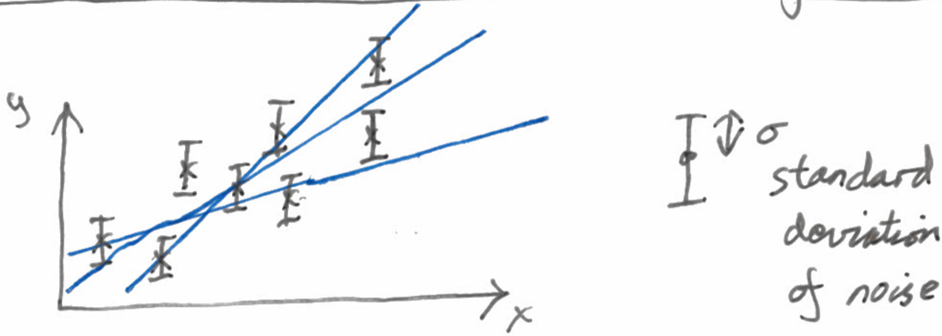
Minimize negative log-likelihood

$$-\log p(\underbrace{\underline{y}}_{\{y^{(n)}\}} / X, \underline{w}) = -\sum_n \log p(y^{(n)} \mid \underline{x}^{(n)}, \underline{w})$$

$$= \frac{1}{2\sigma^2}\sum_n \left[\left(y^{(n)} - f(\underline{x}^{(n)}; \underline{w})\right)^2\right] + \sum_n \left[\frac{1}{2}\log(2\pi\sigma^2)\right]$$

$$= \frac{1}{2\sigma^2} \underbrace{\sum_n \left[\left(y^{(n)} - f(x^{(n)}; \underline{w})\right)^2\right]}_{\text{That is Minimize sum of squares.}} + \frac{N}{2}\log(2\pi\sigma^2)$$

We are uncertain about the model given data



We use probability distributions to express beliefs about parameters

For example:

The "posterior distribution"

$$p(\underline{w} \mid \underbrace{\text{Data}}_{\{\underline{y}, X\}}) = \frac{P(\text{Data} \mid \underline{w})\, P(\underline{w})}{P(\text{Data})}$$

$$\propto P(\text{Data} \mid \underline{w}) \underbrace{P(\underline{w})}_{\text{"Prior beliefs"}}$$

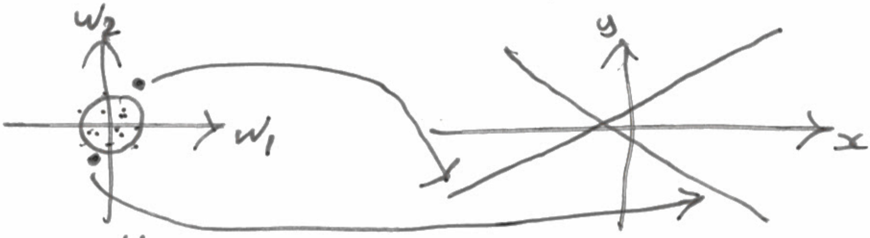
Bayes' Rule

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

$$P(A|B)\,P(B) = P(B|A)\,P(A)$$

Prior distribution

$$p(\underline{w}) = N(\underline{w}; 0, \sigma^2_{prior})$$

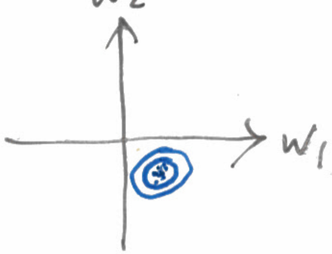


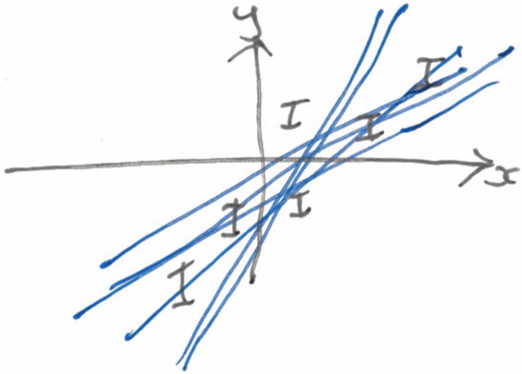
weight space

$$f = w_1 x + w_2$$

Bayes Rule updates using the data

$p(\underline{w}|D)$





Computing the Posterior

For linear regression

$\{ \underline{y}, X \}$

$$p(\underline{w} | D) \propto p(\underline{w}) p(D | \underline{w})$$

$$\propto N(\underline{w}; 0, \sigma_{prior}^2) N(\underline{y}; \underbrace{\Phi \underline{w}}, \sigma^2 I)$$

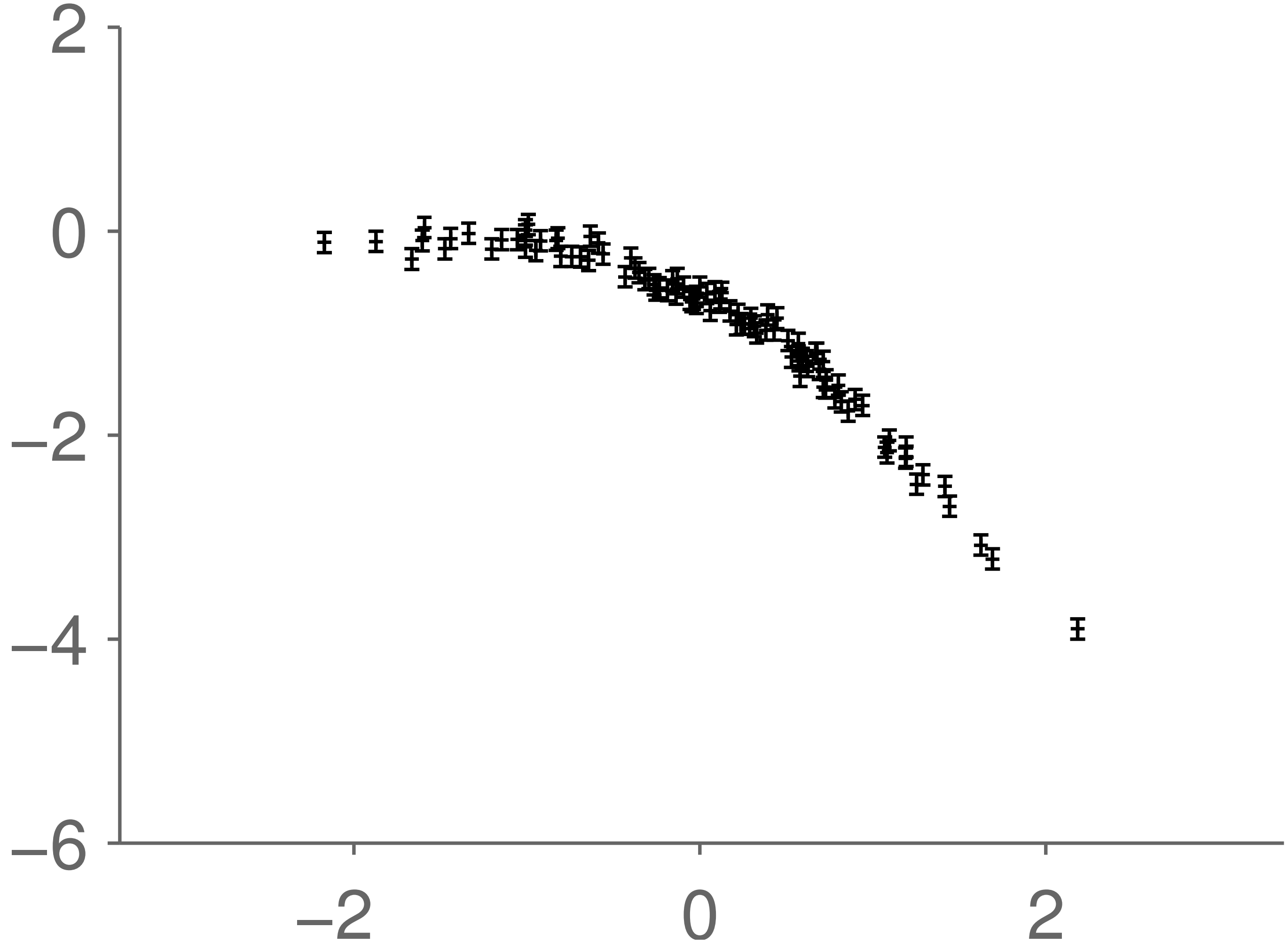
$f(\underline{x}, \underline{w})$ at each input.

Is Gaussian:

$$e^{-\text{some quadratic in } \underline{w}}$$

We find linear & quadratic coeff's of $\underline{w}$ $\Rightarrow$ mean & covariance.

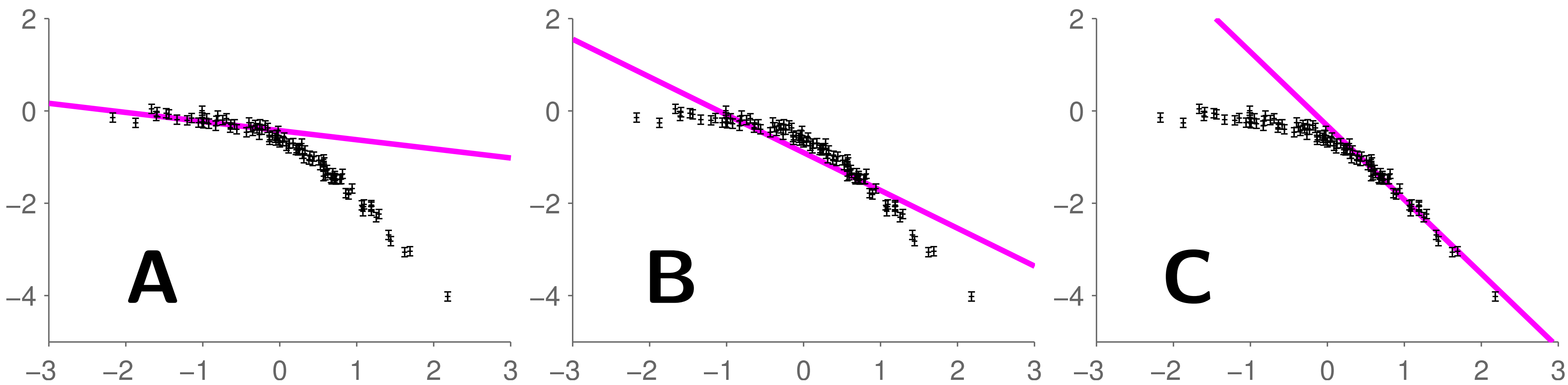
Model mismatch



What will Bayesian linear regression do?

Quiz

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

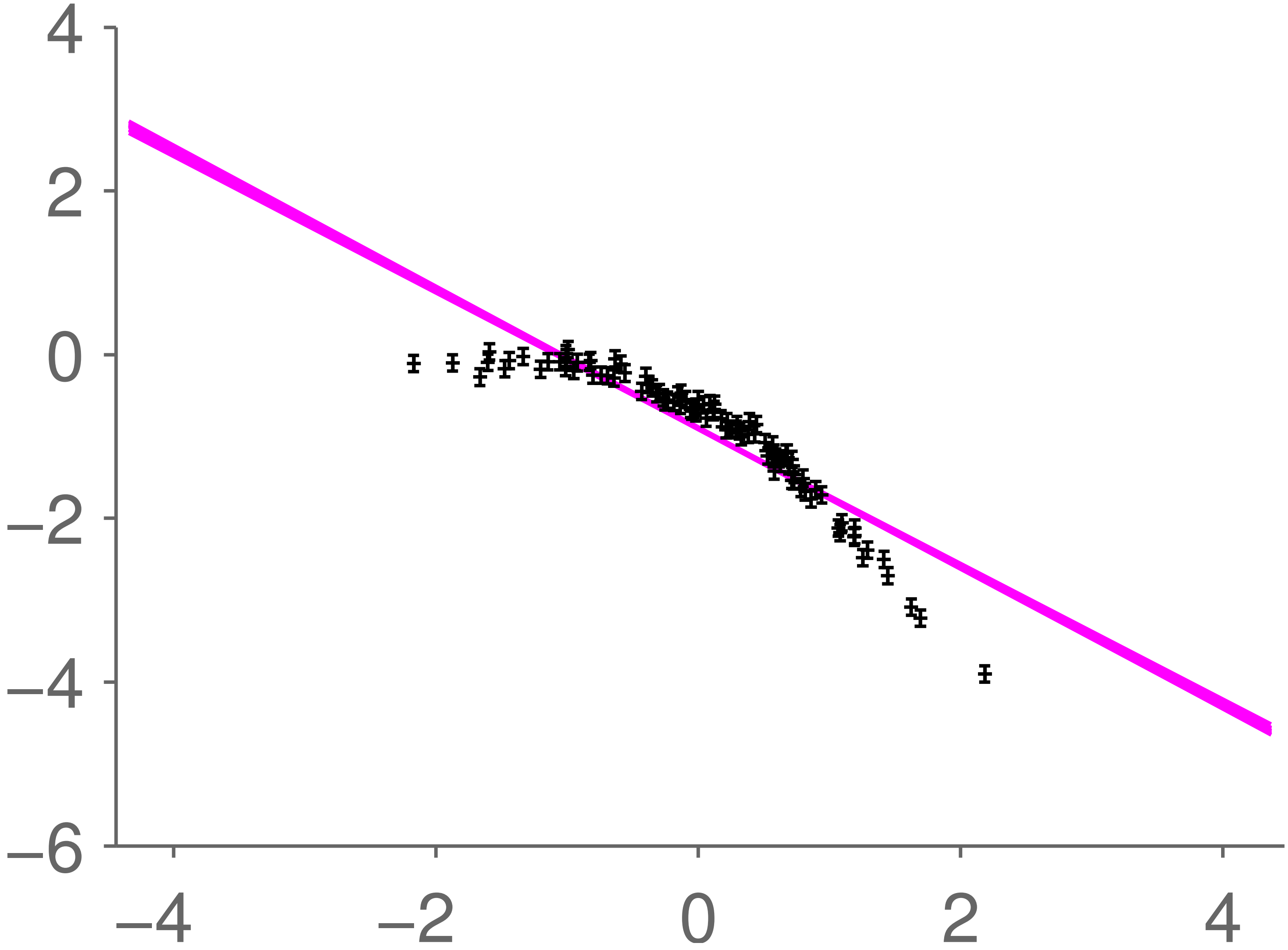


D All of the above

E None of the above

Z Not sure

'Underfitting'



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