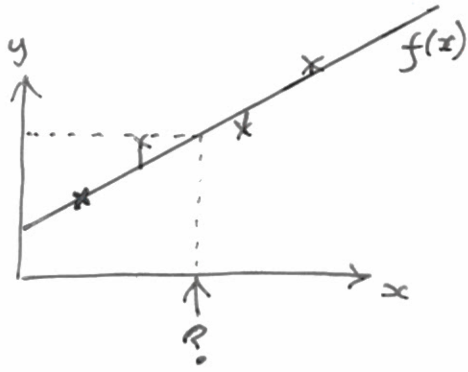
Bayesian Regression

Previously fit functions

$f(x)$ a single guess of output at location x



For classification we fitted

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

For regression we can also write down a probabilistic model:

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

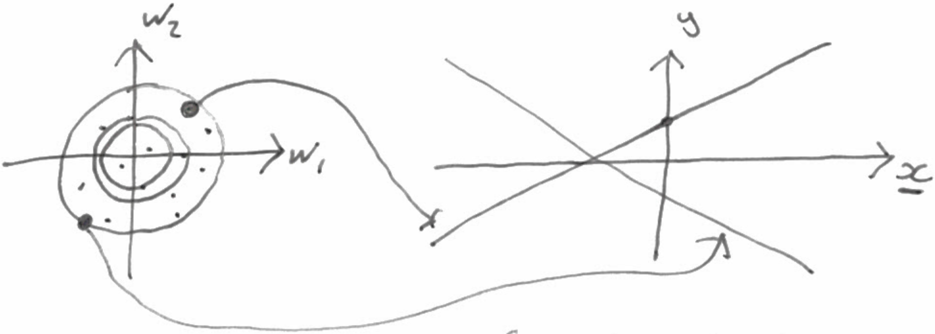
noise variance

Assume σ_y^2 same for each example (for now)

Prior distribution

What model parameters are plausible?

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

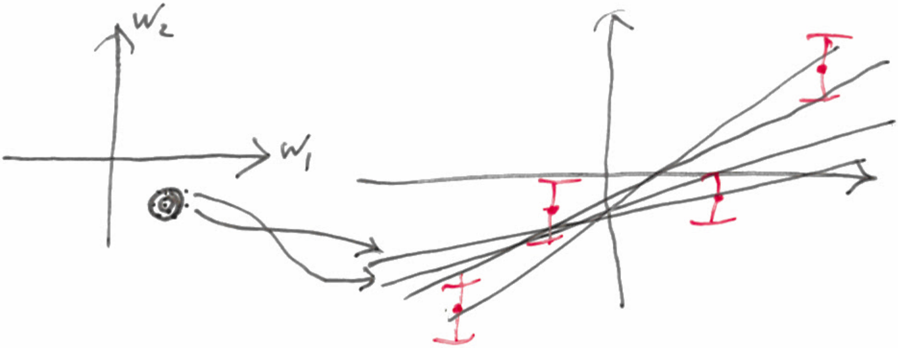


$$f = w_1 x + w_2$$

$\Downarrow$ Bayes' Rule

Updates beliefs using data

$$p(\underline{w} \mid \text{Data})$$



Maximum Likelihood

Minimize negative log-likelihood:

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

$\uparrow$ $N \times 1$

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

$$= \frac{1}{2\sigma_y^2} \underbrace{\sum_n \left(y^{(n)} - f(\underline{x}^{(n)}; \underline{w})\right)^2}_{\text{Minimize}} + \frac{N}{2}\log 2\pi\sigma_y^2$$

That is minimize the sum of squares cost.

Other cost functions?

Variable noise: Could have a different noise variance $(\sigma_y^{(n)})^2$

$\Rightarrow$ Weight each example by $\dfrac{1}{(\sigma_y^{(n)})^2}$

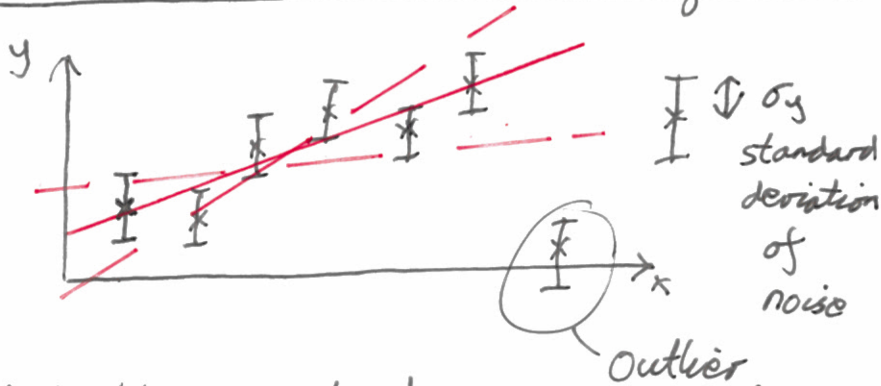
Robust model

Exactly as before:

$p(y \mid \underline{x}, \underline{w})$ is a mixture between the usual model (narrow Gaussian) and a "background model" (maybe a really broad Gaussian, maybe $N(0, 10^6)$)

... Derive likelihood... gradients ... and fit.

We are uncertain about a model given data



We'd like to automate probabilistic reasoning.

Bayesian

→ Probability theory.

Computing posterior

$$p(\underline{w} \mid \underset{\uparrow}{D}) \propto p(\underline{w})\, p(D \mid \underline{w})$$

Data

$\underline{y}$, assume X known.

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

In Gaussian:

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

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

Linear Algebra exercise.