MLPR

Lecturer swap !

(for about 2 weeks)

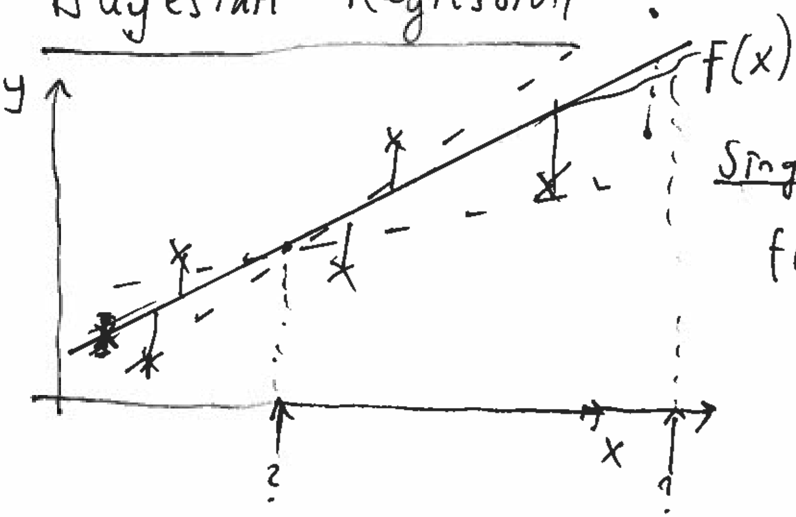
Arno

aonken@inf.ed.ac.uk

"Bayesian methods"

What more can we do with linear / Gaussian models?

Bayesian Regression



<u>Single</u> guess
$f(x)$

For classification, we fitted a probabilistic model $P(\underline{y}|\underline{x})$

For regression, we now write down a probabilistic model:

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

noise variance

For now, simplifying assumption that σ_y^2 is known and same for every output

Maximum likelihood (not Bayesian)

Minimize negative log-likelihood:

$$-\log p(\underset{\substack{\uparrow \\ N\times 1}}{\underline{y}} | X, \underline{w}) = -\sum_n \log p(y^{(n)} | \underline{x}^{(n)}, \underline{w})$$

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

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

⇒ Maximum likelihood for this regression model is least squares.

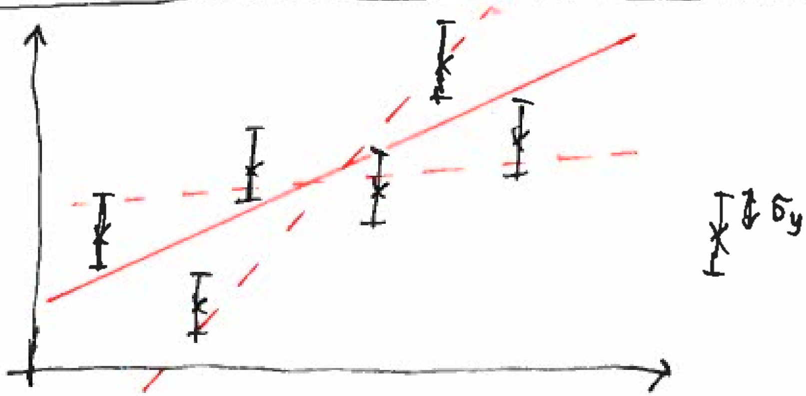
Other costs:

Variable noise: $\sigma_y^{(n)}$ for each data point

⇒ Weight each sample by $\frac{1}{(\sigma_y^{(n)})^2}$

$\frac{1}{(\sigma_y^{(n)})^2}$ called precision

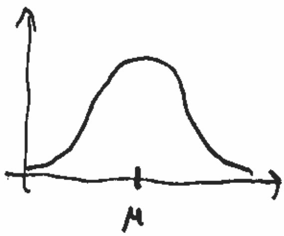
We are uncertain about a model given data



How can we automate this?

Idea: Replace our single guess for $f(x)$ with a distribution over models

The distribution is our mathematical representation of belief.



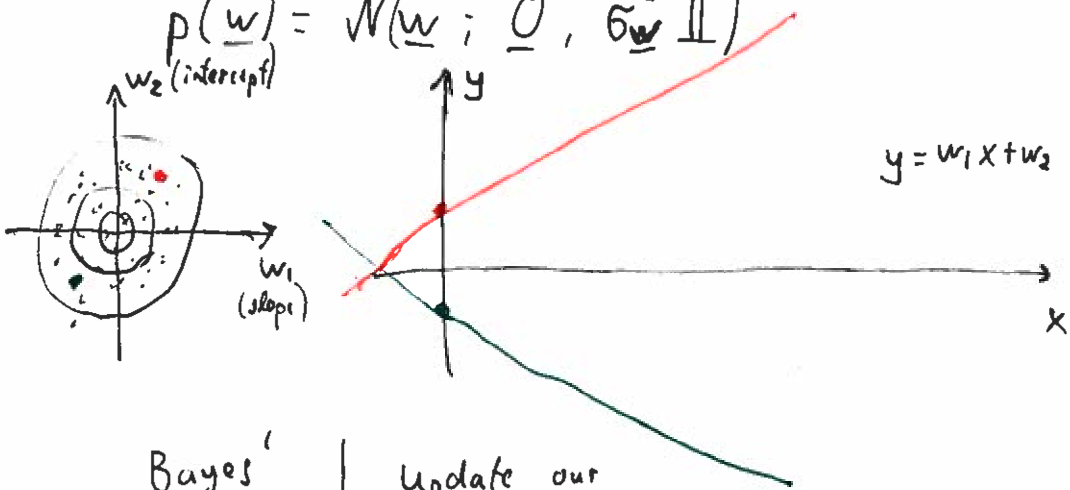
Bayesian Probability Theory

Prior distribution

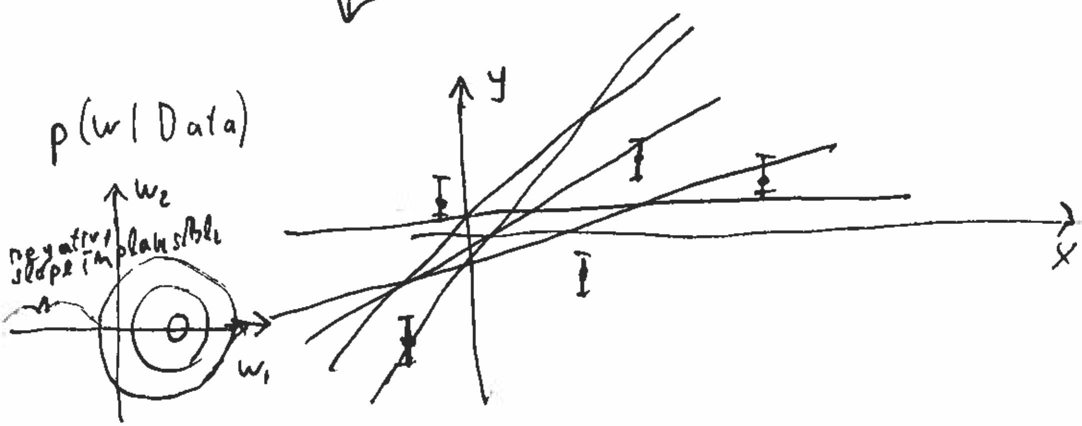
What model parameters are plausible?

Belief could be Gaussian

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



Bayes' rule ↓ Update our beliefs using data



Compute posterior

$$\underbrace{p(w|D)}^{\text{posterior}} \propto \overbrace{p(\underline{w})}^{\text{prior}} \overbrace{\underbrace{p(D|\underline{w})}_{p(y|\underline{w}, X)}}^{\text{likelihood}}$$

Data y

assume X known

for regression

$$\propto \mathcal{N}(\underline{w} \mid \underline{w}_0, V_0) \cdot \mathcal{N}(y; \underbrace{f}_{\Phi w}, \sigma_y^2 \mathbb{I})$$

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

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

obtain via linear algebra (c.f. tut 2 Q2

$$w_N = V_N V_0^{-1} w_0 + \frac{1}{\sigma_y^2} V_N \Phi^T y$$

$$V_N = \sigma_y^2 (\sigma_y^2 V_0^{-1} + \Phi^T \Phi)^{-1}$$