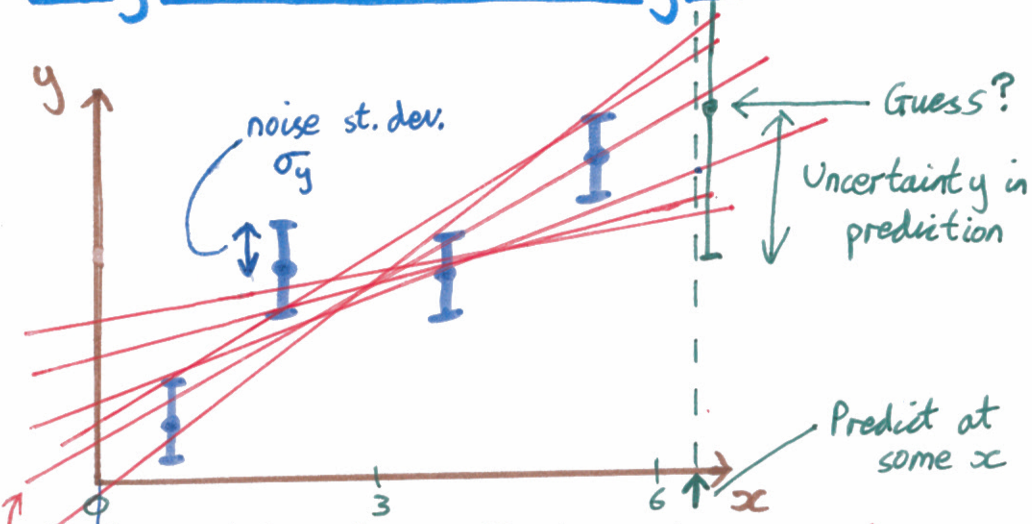
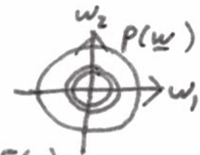
Bayesian Linear Regression



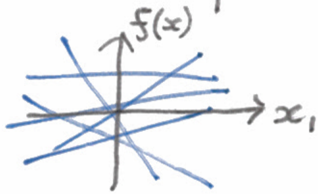
Some different plausible lines $(w_1 x + w_2 x_2)$

$\underbrace{w_1 x + w_2 x_2}_{\mathbf{w}^\top \mathbf{x},\ \text{with } x_2 = 1}$

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



$\Rightarrow$ Broad range functions plausible before see data



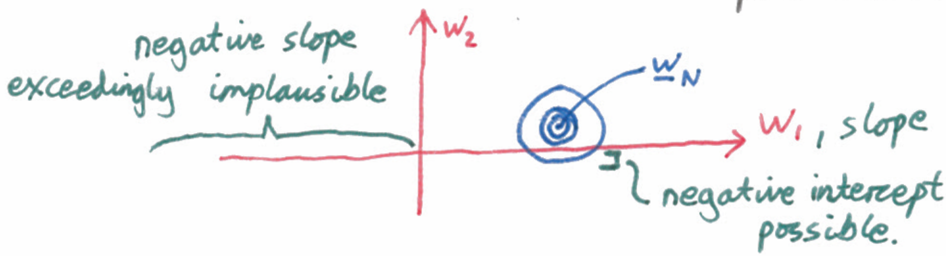
Posterior

"data"

$$p(\mathbf{w} \mid D) \propto p(\mathbf{w}) \underbrace{p(\mathbf{y} \mid X, \mathbf{w})}_{\text{likelihood}}$$

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

(for Gaussian prior and noise)



The Bayesian Update

Prior: $p(\underline{w}) = \mathcal{N}(\underline{w};\ \overbrace{\underline{w}_0,\ V_0}^{\text{We choose}})$

Bayes' rule

Posterior: $p(\underline{w} \mid D) = \mathcal{N}(\underline{w};\ \underbrace{\underline{w}_N,\ V_N}_{\text{From several lines linear algebra}})$

Don't memorize

$$\left[\begin{aligned} V_N &= \sigma_y^2\left(\sigma_y^2 V_0^{-1} + X^\top X\right)^{-1} \\ \underline{w}_N &= V_N V_0^{-1} \underline{w}_0 + \frac{1}{\sigma_y^2} V_N X^\top \underline{y} \end{aligned}\right.$$

(Replace X with Φ if you want.)

Prediction

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

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

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

... several lines later...

$$= \int N\left(\begin{bmatrix} \underline{w} \\ y \end{bmatrix} ; \begin{bmatrix} \underline{w}_N \\ m \end{bmatrix}, \begin{bmatrix} V_N & \Sigma_{\underline{w}, y} \\ \Sigma_{y, \underline{w}} & r^2 \end{bmatrix} \right) d\underline{w}$$

$$= N(y ; \underbrace{m, r^2}_{\text{lots of work to find these}})$$

Yuck!

Instead...

$$y = f(\underline{x}) + \nu, \quad \nu \sim N(0, \sigma_y^2)$$

What do we believe about the function?

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

Beliefs about $\underline{w}$

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

Beliefs about f:

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

Beliefs about y:

$$p(y \mid \underline{x}, D) = N(y; \underline{x}^T \underline{w}_N, \underline{x}^T V_N \underline{x} + \sigma_y^2)$$

Probabilistic Prediction

$$f(\underline{x}) = \underline{w}^\top \underline{x} = \underline{x}^\top \underline{w}$$

$$p(\,f(\underline{x})\,|\,\text{Data}) = N(f;\ \underline{w}_N^\top \underline{x},\ \ \underline{x}^\top V_N \underline{x})$$

$$p(\,y\,|\,\text{Data}) = N(y;\ \ ''\ ,\ \ \ ''\ + \sigma_y^2)$$

Questions

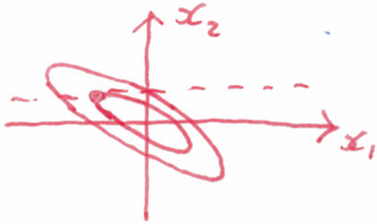
Uncertainty $\underline{x}^\top V_N \underline{x}$ grows with $\underline{x}$

1. Why in figure is most certain region at $x > 0$ (around $x = 3$)?

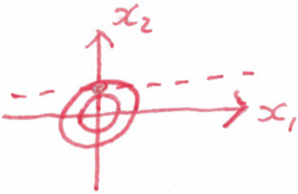
2. What do contours of $\underline{x}^\top V_N \underline{x}$ look like?

A

B



C



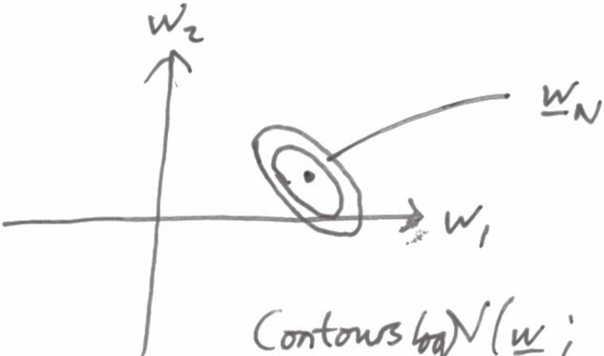
D Other

Z ???

Show MLPR 2017 L19 ① again as L20 ①.

Parameter beliefs?

V_N was posterior covariance of $\underline{w}$



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

$$-\tfrac{1}{2}(\underline{w}-\underline{w}_N)^T V_N^{-1}(\underline{w}-w_N)$$

Decision - Loss functions

$$L(y, \hat{y})$$

loss — y: what happens — $\hat{y}$: "point estimate"/ guess

Minimize expected loss:

$$c = \mathbb{E}_{P(y|D)}\left[L(y, \hat{y})\right]$$

Find $\hat{y}$ that minimizes this cost, c:

Eg square loss $L(y, \hat{y}) = (y - \hat{y})^2$

$$\frac{\partial c}{\partial \hat{y}} = \mathbb{E}_{P(y|D)}\left[2(y - \hat{y})\right]$$

$$= 0 \quad \text{if} \quad \mathbb{E}[y] = \hat{y}$$

$\Rightarrow$ Estimate $\hat{y}$ = mean beliefs