Mechanistic way (long)

$$p(y \mid \underline{x}, D) = \int \underbrace{p(y, \underline{w} \mid \underline{x}, D)}_{\substack{\text{joint distribution} \\ \text{over } y,\, \underline{w} \text{ which} \\ \text{is Gaussian}}} \, dw$$

$= \ldots$ tedious linear algebra

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

$$= \mathcal{N}(y ; \underbrace{m, r^2}_{})$$

obtain from substituting in Gaussians and rearranging terms, and read elements from blocks

$\hat{=}$ integral

Easier way ("bottom-up")

Observation model

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

Beliefs about the function:

$$f = \underline{w}^T \underline{x} = \underline{x}^T \underline{w} \qquad \text{(Form)}$$

$$p(\underline{w} \mid D) = \mathcal{N}(\underline{w}; \underline{w}_N, V_N) \qquad \text{(Beliefs about } \underline{w}\text{)}$$

Distribution over f at test point $\underline{x}$:

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

because $\mathrm{Cov}[A\underline{w} + \underline{b}] = A\,\mathrm{Cov}[\underline{w}]\,A^T$

Beliefs about y:

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

Probabilistic Prediction for Linear Regression

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

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

$$p(y \mid \text{Data}) = \mathcal{N}(y;\ \text{"}\ ,\ \text{"}\ + \sigma_y^2)$$

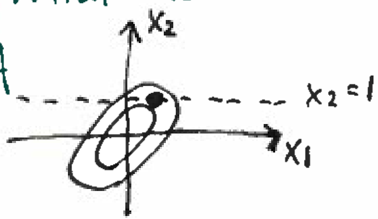
Questions

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

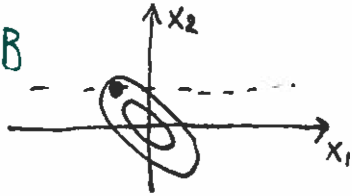
① In the figures of earlier notes, why is the most certain region at $x > 0$ (and not at $x = 0$)?

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

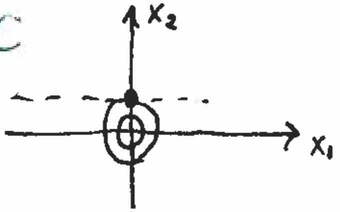
A



B



C

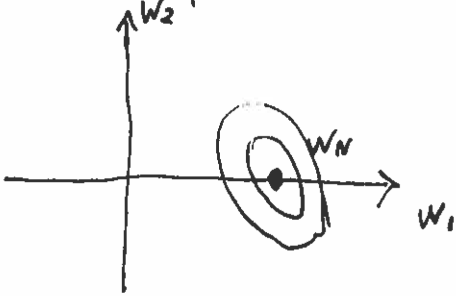


D Other

Z Don't know

Parameter beliefs

V_N : posterior covariance



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

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

Decision - Loss functions

Loss : $L(y, \hat{y})$ "Cost of saying $\hat{y}$ when really we have y"

y: output

$\hat{y}$: estimate of y "point estimate": guess

Minimise the expected loss

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

Find $\hat{y}$ that minimises cost c:

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

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

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

$\Rightarrow$ Point estimate $\hat{y}$ = mean beliefs