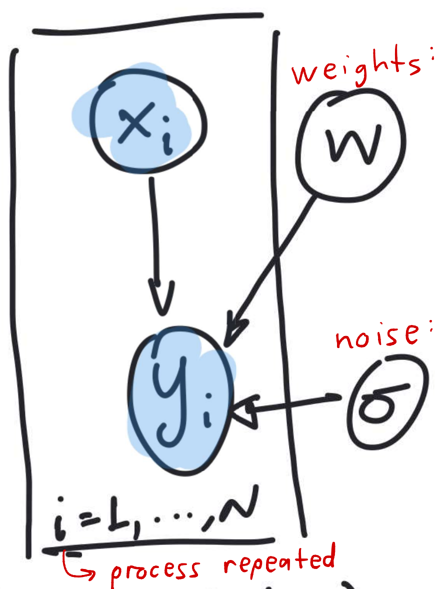
Open questions

- How do I quantify the measurement noise?
- How do we avoid overfitting?
- How do I quantify epistemic uncertainty induced by limited data?
- How do I choose any remaining parameters?
- How do I choose which basis functions to keep?

Probabilistic interpretation



(plate notation)

Prior: $\underline{w} \sim p(\underline{w})$

$\sigma \sim p(\sigma)$

Likelihood:

$$y_i \mid x_i, \underline{w}, \sigma^2 \sim N\Big(\overbrace{\underline{\varphi}^T(x_i)\underline{w}}^{\text{prediction}}, \sigma^2\Big)$$

$$p(y_{1:N} \mid x_{1:N}, \underline{w}, \sigma^2) \overset{\text{conditionally independent}}{=} \prod_{i=1}^{N} p(y_i \mid x_i, \underline{w}, \sigma^2)$$

$$= \prod_{i=1}^{N} (2\pi)^{-1/2} \sigma^{-1} \exp\left\{ - \frac{\left(y_i - \underline{\varphi}^T(x_i)\underline{w}\right)^2}{2\sigma^2} \right\}$$

$$= (2\pi)^{-N/2} \sigma^{-N} \exp\left\{ - \frac{1}{2\sigma^2} \sum_{i=1}^{N} \overbrace{\left(y_i - \underline{\varphi}^T(x_i)\underline{w}\right)^2}^{\text{square error}} \right\}$$

★ **Posterior** $\propto$ Likelihood $\cdot$ Prior

$$p(\underline{w}, \sigma \mid x_{1:N}, y_{1:N}) \propto \boxed{p(y_{1:N} \mid x_{1:N}, \underline{w}, \sigma^2)}\, \underbrace{p(\underline{w})}_{1}\, \underbrace{p(\sigma)}_{1}$$

assume

Maximum likelihood estimate of weights yields least squares

$$\log p(y_{1:N} | x_{1:N}, \underline{w}, \sigma) = -\frac{N}{2}\log 2\pi - N\log\sigma - \frac{1}{2\sigma^2}\underbrace{\sum_{i=1}^{N}\left(y_i - \underline{\phi}^T(x_i)\underline{w}\right)^2}_{= L(\underline{w})}$$

$$\max_{\underline{w}} \log \text{like} \equiv \min_{\underline{w}} L(\underline{w})$$

$$\therefore \quad \underline{\underline{\Phi}}^T \underline{\underline{\Phi}}\, \underline{w} = \underline{\underline{\Phi}}^T \underline{y}$$

For σ:

$$\frac{\partial}{\partial\sigma}(\log \text{like}) = 0 \rightarrow \sigma = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \underline{\phi}^T(x)\underline{w}\right)^2$$

PREDICTIVE SCIENCE LABORATORY