

Lecture 15: Advanced topics in Bayesian linear regression

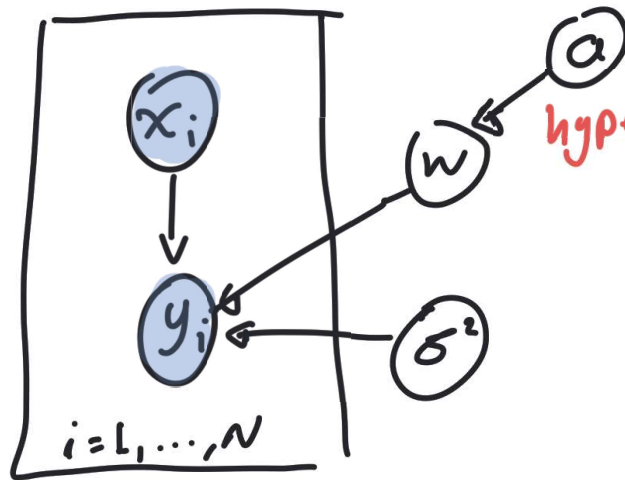
Professor Ilias Bilonis

The evidence approximation

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?

Hyper-priors



hyper-parameter

Prior: $\alpha \sim p(\alpha)$

$$\underline{w} | \alpha \sim p(\underline{w} | \alpha) = N(\underline{w} | 0, \alpha^{-1} \mathbf{I})$$

$$\sigma \sim p(\sigma)$$

Likelihood: $y_i | x_i, \underline{w}, \sigma^2 \sim N(\phi^T(x_i) \underline{w}, \sigma^2)$

$$\hookrightarrow p(y_{1:N} | x_{1:N}, \underline{w}, \sigma^2) = \prod_{i=1}^N \dots$$

Posterior over hyper-parameters and the evidence approximation

$$p(\underline{w}, \alpha, \sigma | x_{1:n}, y_{1:n}) \propto \underbrace{p(y_{1:n} | x_{1:n}, \underline{w}, \sigma) p(\underline{w} | \alpha) p(\alpha) p(\sigma)}$$

$$p(\alpha, \sigma | x_{1:n}, y_{1:n}) = \int \underbrace{p(\underline{w}, \alpha, \sigma | x_{1:n}, y_{1:n})}_{\substack{\propto p(y_{1:n} | x_{1:n}, \underline{w}, \sigma) p(\underline{w} | \alpha) p(\alpha) p(\sigma)}} d\underline{w}$$

$$\propto \int p(y_{1:n} | x_{1:n}, \underline{w}, \sigma) p(\underline{w} | \alpha) \boxed{p(\alpha) p(\sigma)} d\underline{w}$$

$$= \int \underbrace{p(y_{1:n} | x_{1:n}, \underline{w}, \sigma) p(\underline{w} | \alpha)}_{\substack{p(\underline{w} | x_{1:n}, y_{1:n}, \alpha, \sigma) \propto p(y_{1:n} | x_{1:n}, \underline{w}, \sigma) \cdot p(\underline{w} | \alpha)}} d\underline{w} p(\alpha) p(\sigma)$$

$$\mathcal{N}(\underline{w} | \underline{m}(\alpha, \sigma), \Sigma(\alpha, \sigma))$$

$$= Z(\alpha, \sigma) p(\alpha) p(\sigma)$$



PREDICTIVE
SCIENCE LABORATORY

Evidence approximation⁴: Find α, σ by maximizing their posterior (with \underline{w} integrated out).

$$\alpha^*, \sigma^* = \arg \max_{\alpha, \sigma} p(\alpha, \sigma | x_{1:n}, y_{1:n})$$

Example

