Lecture 15: Advanced topics in Bayesian linear regression

Professor Ilias Bilionis

The evidence approximation

PREDICTIVE
SCIENCE LABORATORY

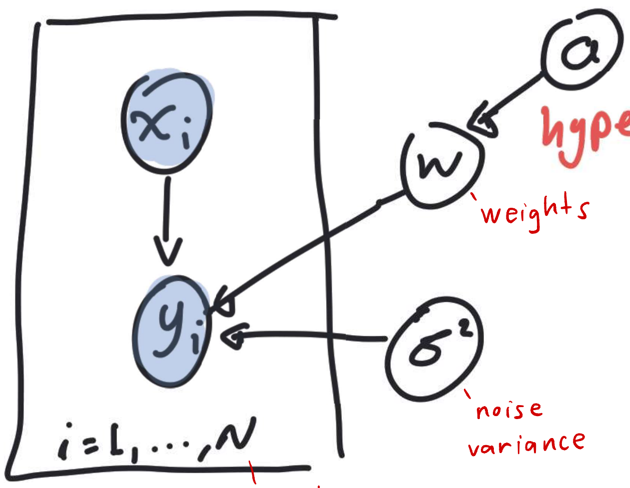
Open questions

- 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?
  ↳ evidence approximation
- How do I choose which basis functions to keep?

remaining problems to address

Hyper-priors

blue = observed



repeat N times

hyper-parameter - parameter of the prior of w

Prior:

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

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

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

Posterior over hyper-parameters and the evidence approximation

likelihood, prior, hyper-prior, prior

$$p(\underline{w}, \alpha, \sigma | x_{1:N}, y_{1:N}) \propto \underbrace{p(y_{1:N} | x_{1:N}, \underline{w}, \sigma)}_{\text{likelihood}} \underbrace{p(\underline{w}|\alpha)}_{\text{prior}} \underbrace{p(\alpha)}_{\text{hyper-prior}} \underbrace{p(\sigma)}_{\text{prior}}$$

$$p(\alpha, \sigma | x_{1:N}, y_{1:N}) \overset{*}{=} \int p(\underline{w}, \alpha, \sigma | x_{1:N}, y_{1:N}) d\underline{w}$$

marginalize:

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

$$= \int p(y_{1:N} | x_{1:N}, \underline{w}, \sigma) p(\underline{w}|\alpha) d\underline{w} \; p(\alpha) p(\sigma)$$

prop. to

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

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

∴ this integral is the normalization constant

the normalization constant of the posterior of the weights conditioned on α & σ

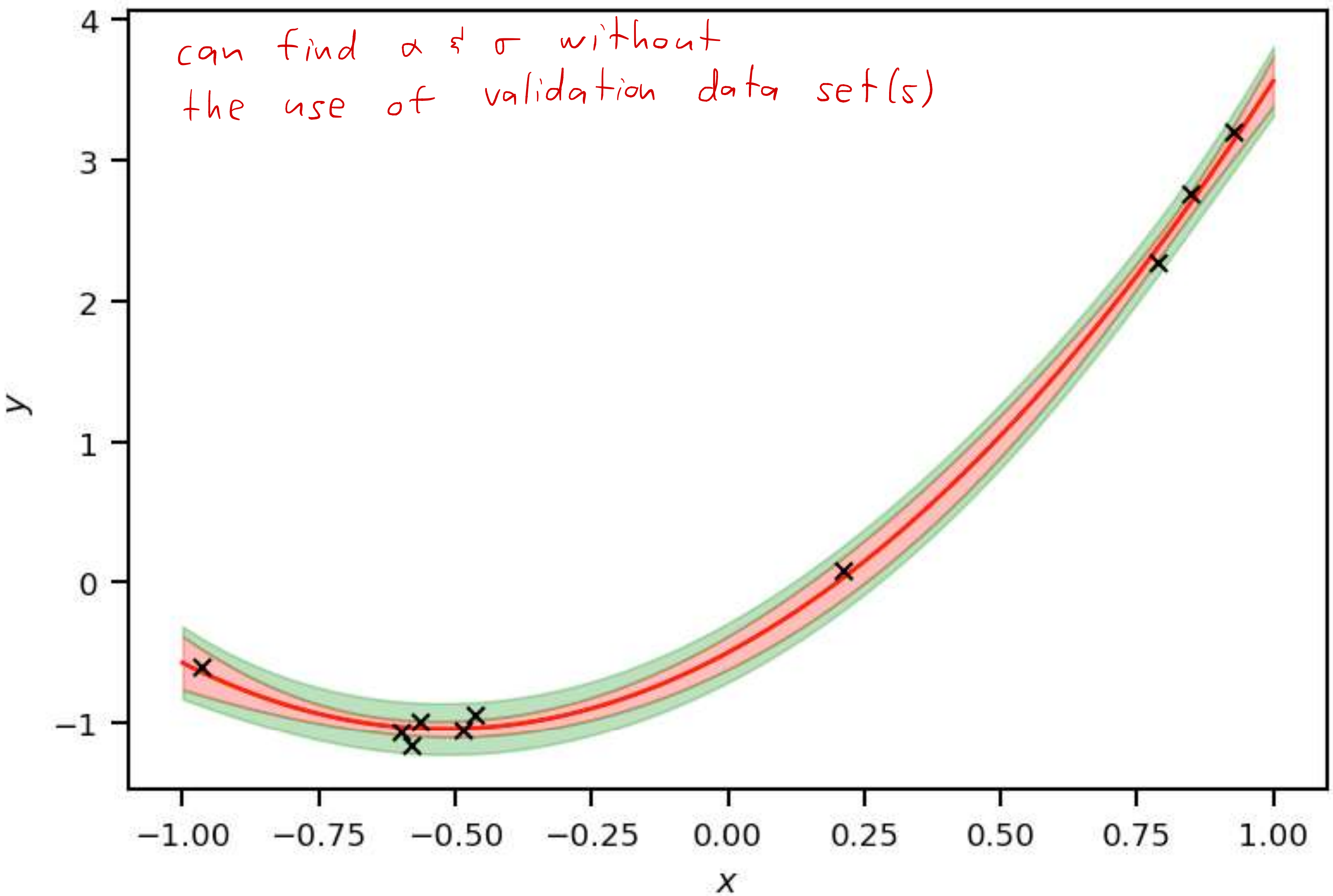
$$= Z(\alpha, \sigma) \, p(\alpha) p(\sigma)$$

PREDICTIVE SCIENCE LABORATORY

<u>Evidence approximation</u>: 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



PREDICTIVE
SCIENCE LABORATORY