

Workshop 11

COMP90051 Statistical Machine Learning Semester 1, 2023

Learning Outcomes

By the end of this workshop you should be able to:

- explain how frequentist and Bayesian treatments of linear regression differ
- 2. implement exact Bayesian inference for a simple linear regression model (assuming const. variance)
- 3. implement approximate Bayesian inference
- 4. perform model selection using the evidence (a.k.a. marginal likelihood)

Review: Bayesian linear regression

Problem: predict a real-valued response y given a feature vector $\mathbf{x} = (x_1, ..., x_d)^{\top} \in \mathbb{R}^d$

Assume a simple probabilistic model:

$$y_i | \mathbf{x}_i \stackrel{\text{ind}}{\sim} \text{Normal}(\mathbf{x}_i^{\mathsf{T}} \mathbf{w}, \sigma^2), \qquad i \in \{1, ..., n\}$$

- * mean response is a *linear combination* of the features
- * variance σ^2 is assumed known
- Adopt a Bayesian approach treat unknown parameter w
 as a random variable and choose a prior:

$$\mathbf{w} \sim \text{Normal}(\mathbf{0}, \gamma^2 \mathbf{I}_d)$$

Review: Bayesian linear regression

- Observe data: $\mathbf{X} = (\mathbf{x}_1^\top, \dots, \mathbf{x}_n^\top)^\top$, $\mathbf{y} = (y_1, \dots, y_n)^\top$
- Bayesian inference for w:

$$p(\mathbf{w}|\mathbf{X},\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{w},\mathbf{X})p(\mathbf{w})}{\int p(\mathbf{y}|\mathbf{w},\mathbf{X})p(\mathbf{w}) d\mathbf{w}}$$

- * posterior distribution accounts for uncertainty in w
- Bayesian inference for response on test instance \mathbf{x}_{\star} :

$$p(y_{\star}|\mathbf{x}_{\star}, \mathbf{X}, \mathbf{y}) = \int p(y_{\star}|\mathbf{w}, \mathbf{x}_{\star}) p(\mathbf{w}|\mathbf{X}, \mathbf{y}) d\mathbf{w}$$
posterior predictive likelihood posterior

- * accounts for uncertainty in ${f w}$ when predicting y_{\star}
- Exact inference is possible for this model, but not in general

Worksheet 11



