



# Workshop 9

COMP90051 Machine Learning  
Semester 1, 2021

# Learning Outcomes

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

1. 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  $\mathbf{x} = (x_1, \dots, x_d)^\top \in \mathbb{R}^d$

- Assume a simple probabilistic model:

$$y_i | \mathbf{x}_i \stackrel{\text{ind}}{\sim} \text{Normal}(\mathbf{x}_i^\top \mathbf{w}, \sigma^2), \quad i \in \{1, \dots, n\}$$

- \* mean response is a *linear combination* of the features
  - \* variance  $\sigma^2$  is assumed *known*
- Adopt a Bayesian approach – treat unknown parameter  $\mathbf{w}$  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  $\mathbf{w}$ :

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

\* posterior *distribution* accounts for uncertainty in  $\mathbf{w}$

- Bayesian inference for response on test instance  $\mathbf{x}_\star$ :

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

\* accounts for uncertainty in  $\mathbf{w}$  when predicting  $y_\star$

- Exact inference is possible for this model, but not in general

# Worksheet 9