03 Bayesian Linear Regression

June 20, 2021

Information: *Basic concepts and simple examples of Bayesian linear regression*

Written by: *Zihao Xu*

Last update date: *06.20.2021*

1 Maximum Likelihood Estimation

1.1 Motivation

In the chapter talking about ***Generalization and Regularization***, the concepts of parameter estimation, bias and variance are used to formally characterize notions of generalization, underfitting and overfitting. Here are some important remarks:

- View the parameter estimator $\hat{\boldsymbol{\theta}}$ as a **function** of the sampled training dataset

$$\hat{\boldsymbol{\theta}} = g\left(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(m)}\right)$$

- The datasets (training, testing and probably validation) are generated by a **i.i.d.** probability distribution over datasets called the **data-generating process**
- Assume that the true parameter value $\boldsymbol{\theta}$ is fixed but unknown
- Since the **data** is drawn from a **random process**, any function of the data is random, which means the parameter estimator $\hat{\boldsymbol{\theta}}$ is a **random variable**

The concepts of **bias** and **variance** are used to measure the performance of a parameter estimator. However, **for obtaining a good estimator**, it's not a good idea to guess that some function might make a good estimator and then to analyze its bias and variance. This motivated some principles from which specific functions that are good estimators for different models can be derived.

1.2 Basic Concepts

1.2.1 Definition

Consider a set of m examples $\mathbf{X} = \left\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(m)}\right\}$ drawn independently from the true but unknown data-generating distribution $p_{data}\left(\mathbf{x}\right)$

```
[ ]:
```