Reading Notes for ch3 Probability Distributions

Xiang Pan

September 11, 2021

1 Linear Basis Function Models

This section just reviews the basic linear regression setting.

2 Bias-Variance Decomposition

$$\text{expected loss } = (\text{ bias })^2 + \text{ variance } + \text{ noise}$$

$$(\text{ bias })^2 = \int \{\mathbb{E}_{\mathcal{D}}[y(\mathbf{x};\mathcal{D})] - h(\mathbf{x})\}^2 p(\mathbf{x})\mathrm{d}\mathbf{x}$$

$$\text{variance } = \int \mathbb{E}_{\mathcal{D}}\left[\{y(\mathbf{x};\mathcal{D}) - \mathbb{E}_{\mathcal{D}}[y(\mathbf{x};\mathcal{D})]\}^2\right] p(\mathbf{x})\mathrm{d}\mathbf{x}$$

$$\text{noise } = \int \{h(\mathbf{x}) - t\}^2 p(\mathbf{x}, t)\mathrm{d}\mathbf{x}\mathrm{d}t$$

Very flexible models having low bias and high variance, and relatively rigid models having high bias and low variance.

3 Bayesian Liear Regression

$$p(t \mid \mathbf{t}, \alpha, \beta) = \int p(t \mid \mathbf{w}, \beta) p(\mathbf{w} \mid \mathbf{t}, \alpha, \beta)\mathrm{d}\mathbf{w}$$

$$p(t \mid \mathbf{x}, \mathbf{t}, \alpha, \beta) = \mathcal{N}\left(t \mid \mathbf{m}_N^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x}), \sigma_N^2(\mathbf{x})\right)$$

$$\sigma_N^2(\mathbf{x}) = \text{noise on the data} + \text{uncertainty associated with the parameters w.}$$

$$\sigma_N^2(\mathbf{x}) = \frac{1}{\beta} + \phi(\mathbf{x})^{\mathrm{T}} \mathbf{S}_N \phi(\mathbf{x})$$

When dataset size get unlimited, the second term goes to zero.