4.1. Linear Regression

The goal of regression is to predict the value of one or more continuous *target* variables t given the value of a D-dimensional vector $\mathbf{x}$ of *input* variables. Typically we are given a training data set comprising N observations $\{\mathbf{x}_n\}$, where $n = 1, \ldots, N$, together with corresponding target values $\{t_n\}$, and the goal is to predict the value of t for a new value of $\mathbf{x}$. To do this, we formulate a function $y(\mathbf{x}, \mathbf{w})$ whose values for new inputs $\mathbf{x}$ constitute the predictions for the corresponding values of t, and where $\mathbf{w}$ represents a vector of parameters that can be learned from the training data.

The simplest model for regression is one that involves a linear combination of the input variables:

$$y(\mathbf{x}, \mathbf{w}) = w_0 + w_1 x_1 + \ldots + w_D x_D \tag{4.1}$$

where $\mathbf{x} = (x_1, \ldots, x_D)^{\mathrm{T}}$. The term *linear regression* sometimes refers specifically to this form of model. The key property of this model is that it is a linear function of the parameters $w_0, \ldots, w_D$. It is also, however, a linear function of the input variables x_i, and this imposes significant limitations on the model.

4.1.1 Basis functions

We can extend the class of models defined by (4.1) by considering linear combinations of fixed nonlinear functions of the input variables, of the form

$$y(\mathbf{x}, \mathbf{w}) = w_0 + \sum_{j=1}^{M-1} w_j \phi_j(\mathbf{x}) \tag{4.2}$$

where $\phi_j(\mathbf{x})$ are known as *basis functions*. By denoting the maximum value of the index j by $M - 1$, the total number of parameters in this model will be M.

Section 4.3

The parameter w_0 allows for any fixed offset in the data and is sometimes called a *bias* parameter (not to be confused with bias in a statistical sense). It is often convenient to define an additional dummy basis function $\phi_0(\mathbf{x})$ whose value is fixed at $\phi_0(\mathbf{x}) = 1$ so that (4.2) becomes

$$y(\mathbf{x}, \mathbf{w}) = \sum_{j=0}^{M-1} w_j \phi_j(\mathbf{x}) = \mathbf{w}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x}) \tag{4.3}$$

where $\mathbf{w} = (w_0, \ldots, w_{M-1})^{\mathrm{T}}$ and $\boldsymbol{\phi} = (\phi_0, \ldots, \phi_{M-1})^{\mathrm{T}}$. We can represent the model (4.3) using a neural network diagram, as shown in Figure 4.1.

By using nonlinear basis functions, we allow the function $y(\mathbf{x}, \mathbf{w})$ to be a nonlinear function of the input vector $\mathbf{x}$. Functions of the form (4.2) are called linear models, however, because they are linear in $\mathbf{w}$. It is this linearity in the parameters that will greatly simplify the analysis of this class of models. However, it also leads to some significant limitations.

Section 6.1