

so there is no need to store moving averages. Layer normalization is compared with batch normalization in Figure 7.8.

Exercises

- 7.1** (★) By substituting (7.10) into (7.7) and using (7.8) and (7.9), show that the error function (7.7) can be written in the form (7.11).
- 7.2** (★) Consider a Hessian matrix \mathbf{H} with eigenvector equation (7.8). By setting the vector \mathbf{v} in (7.14) equal to each of the eigenvectors \mathbf{u}_i in turn, show that \mathbf{H} is positive definite if, and only if, all its eigenvalues are positive.
- 7.3** (★★) By considering the local Taylor expansion (7.7) of an error function about a stationary point \mathbf{w}^* , show that the necessary and sufficient condition for the stationary point to be a local minimum of the error function is that the Hessian matrix \mathbf{H} , defined by (7.5) with $\hat{\mathbf{w}} = \mathbf{w}^*$, is positive definite.
- 7.4** (★★) Consider a linear regression model with a single input variable x and a single output variable y of the form

$$y(x, w, b) = wx + b \quad (7.61)$$

together with a sum-of-squares error function given by

$$E(w, b) = \frac{1}{2} \sum_{n=1}^N \{y(x_n, w, b) - t_n\}^2. \quad (7.62)$$

Derive expressions for the elements of the 2×2 Hessian matrix given by the second derivatives of the error function with respect to the weight parameter w and bias parameter b . Show that the trace and the determinant of this Hessian are both positive. Since the trace represents the sum of the eigenvalue and the determinant corresponds to the product of the eigenvalues, then both eigenvalues are positive and hence the stationary point of the error function is a minimum.

- 7.5** (★★) Consider a single-layer classification model with a single input variable x and a single output variable y of the form

$$y(x, w, b) = \sigma(wx + b) \quad (7.63)$$

where $\sigma(\cdot)$ is the logistic sigmoid function defined by (5.42) together with a cross-entropy error function given by

$$E(w, b) = \sum_{n=1}^N \{t_n \ln y(x_n, w, b) + (1 - t_n) \ln(1 - y(x_n, w, b))\}. \quad (7.64)$$

Derive expressions for the elements of the 2×2 Hessian matrix given by the second derivatives of the error function with respect to the weight parameter w and bias parameter b . Show that the trace and the determinant of this Hessian are both positive.