Errata

The following is the errata for the second edition of **”Machine Learning Refined: Foundations, Algorithms, and Applications”** published by Cambridge University Press in 2020.

page	location	incorrect	correct
60	line 10	initialized at the point $w^0 = 2$	initialized at the point $w^0 = 1.75$
80	line 19	as in the first and third panel	as in the first and second panel
81	line 2	as in the second panel	as in the third panel
113	line 14	metrics of around 4500 and 3000	metrics of around 4.7 and 3.1
119	equation (5.49)	**b** in equation (5.49) is a column vector	**b** should be a row vector
122	exercise 5.4	circumstancs	circumstances
140	Figure 6.10	The phrase ”with three *noisy* data points pointed to by small arrows” should be removed from the figure caption.	
145	line 29	dataset shown Figure	dataset shown in Figure
145	line 31	steps in beginning at	steps beginning at
146	Figure 6.13	The bottom row of the figure is missing	See Figure 6.13 below
168	equation (6.83)	model $(x_p, \mathbf{w})$	model $(\mathbf{x}_p, \mathbf{w})$
169	line 17	$\Omega - 1$	Ω_{-1}
178	line 20	*farthest*from	*farthest* from

page	location	incorrect	correct
212	equation (8.10)	$\mathbf{C}\,\mathbf{C}^T = \mathbf{I}_{N\times N}$	$\mathbf{C}^T\mathbf{C} = \mathbf{I}_{N\times N}$
258	line 4	$\mathbf{D}^{-1/2}$	$\mathbf{D}^{-\frac{1}{2}}$
258	equation (9.6)	$\mathbf{D}^{-{}^{1}/_{2}}$	$\mathbf{D}^{-\frac{1}{2}}$
268	line 5	(as illustrated in the bottom panel of Figure 9.22)	(as illustrated in the bottom panel of Figure 9.22 where $\lambda = 130$)
371	Figure 11.47	Some panel titles in the figure are incorrect.	The title for the top-left panel should read: data. The title for top-middle panel should read: individual models. The title for top-right and bottom-right panels should read: modal model.
438	line 3	graident	gradient
474	Figure A.1	The figure caption may be cut off in some electronic versions of the book.	Figure A.1 An example of a time series data, representing the price of a financial stock measured at 450 consecutive points in time.
501 502	line 23 line 11	than value of In this section we discuss a two	than the value of In this section we discuss two
528	line 12	" ... we update the partial derivative of each parent by multiplying it by the partial derivative of its children node with respect to that parent. When the backward sweep is completed we will have recursively constructed the gradient of the function with respect to all of its inputs."	While this is not incorrect per se, the reader should note that in the backward sweep of reverse-mode differentiation, when a parent node has multiple children the accumulated partials should be added, not multiplied, since this is what the chain rule requires. See Example 0.1 below.

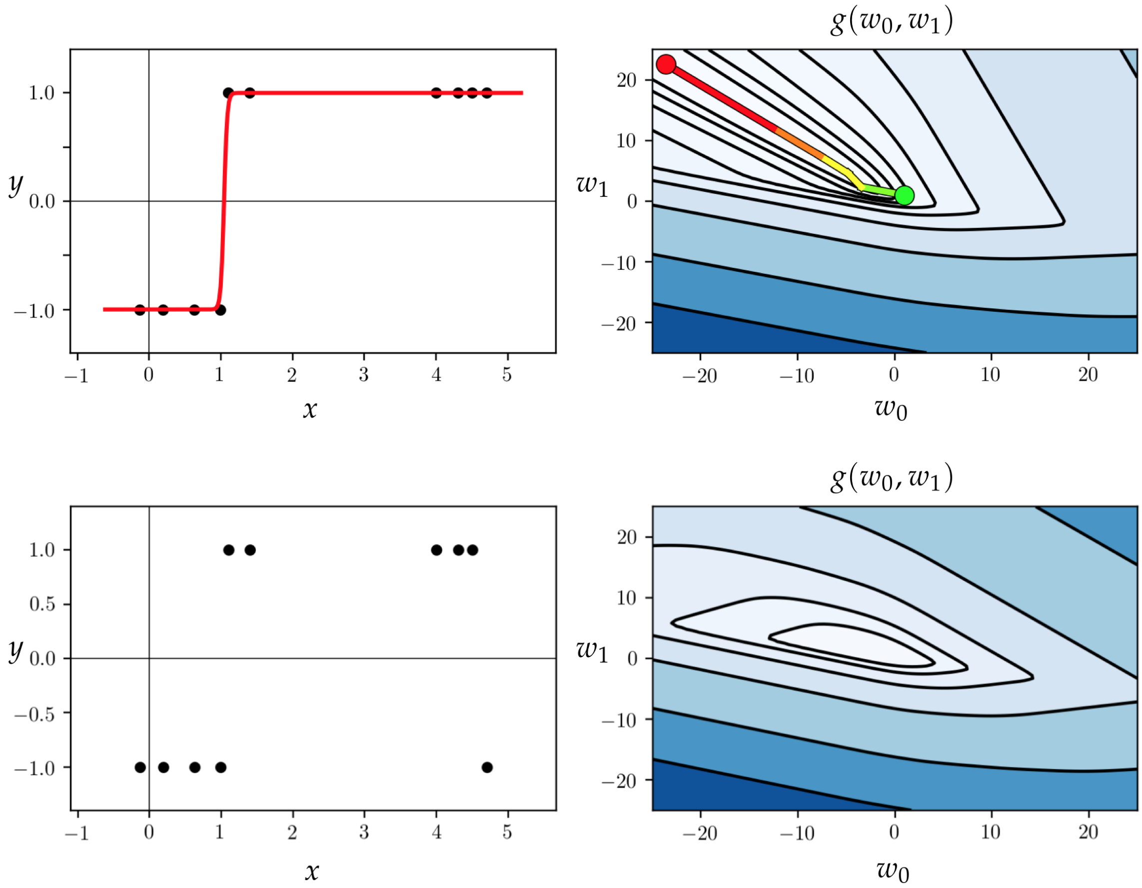

Figure 6.13 Figure associated with Example 6.6. See text for details.

Example 0.1 Reverse-mode differentiation of a simple function

Consider the simple function $g(w) = w^2 + \sin(w)$ whose computation graph is drawn in the top panel of Figure 0.2, and consists of nodes w (input node), a, b, and c. In the forward sweep of the automatic differentiation we compute the partial derivatives of each child node with respect to its parent(s) and store them, as illustrated in the bottom panel of Figure 0.2.

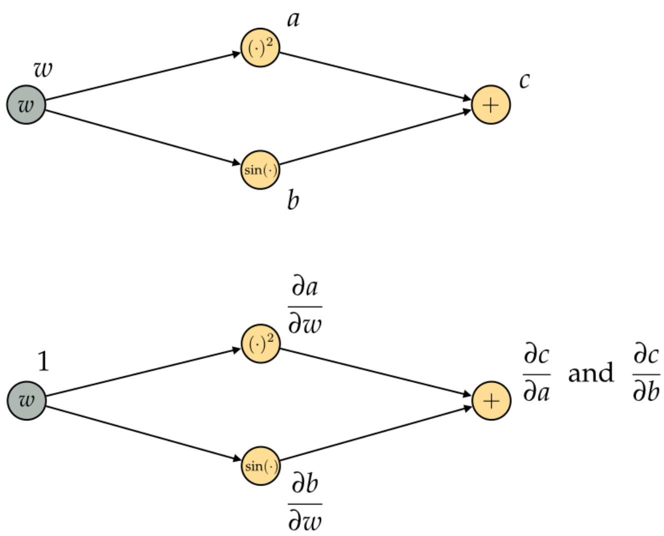

Figure 0.2 (top panel) The computation graph associated with the function $g(w) = w^2 + \sin(w)$. (bottom panel) The result of the full forward sweep of the reverse-mode differentiation of $g(w)$ with respect to w.

The backward sweep then starts at node c, all the way to the right, and at each step, we update the partial derivative of each parent node by multiplying it by the partial derivative of its children node with respect to that parent, as shown from top to bottom in Figure 0.3. Note importantly that at node w – a parent node with more than one child – we must add up the derivative contributions of each of its children to compute the final derivative at this node as

$$\frac{\partial g}{\partial w} = \frac{\partial c}{\partial a}\frac{\partial a}{\partial w} + \frac{\partial c}{\partial b}\frac{\partial b}{\partial w}. \tag{0.1}$$

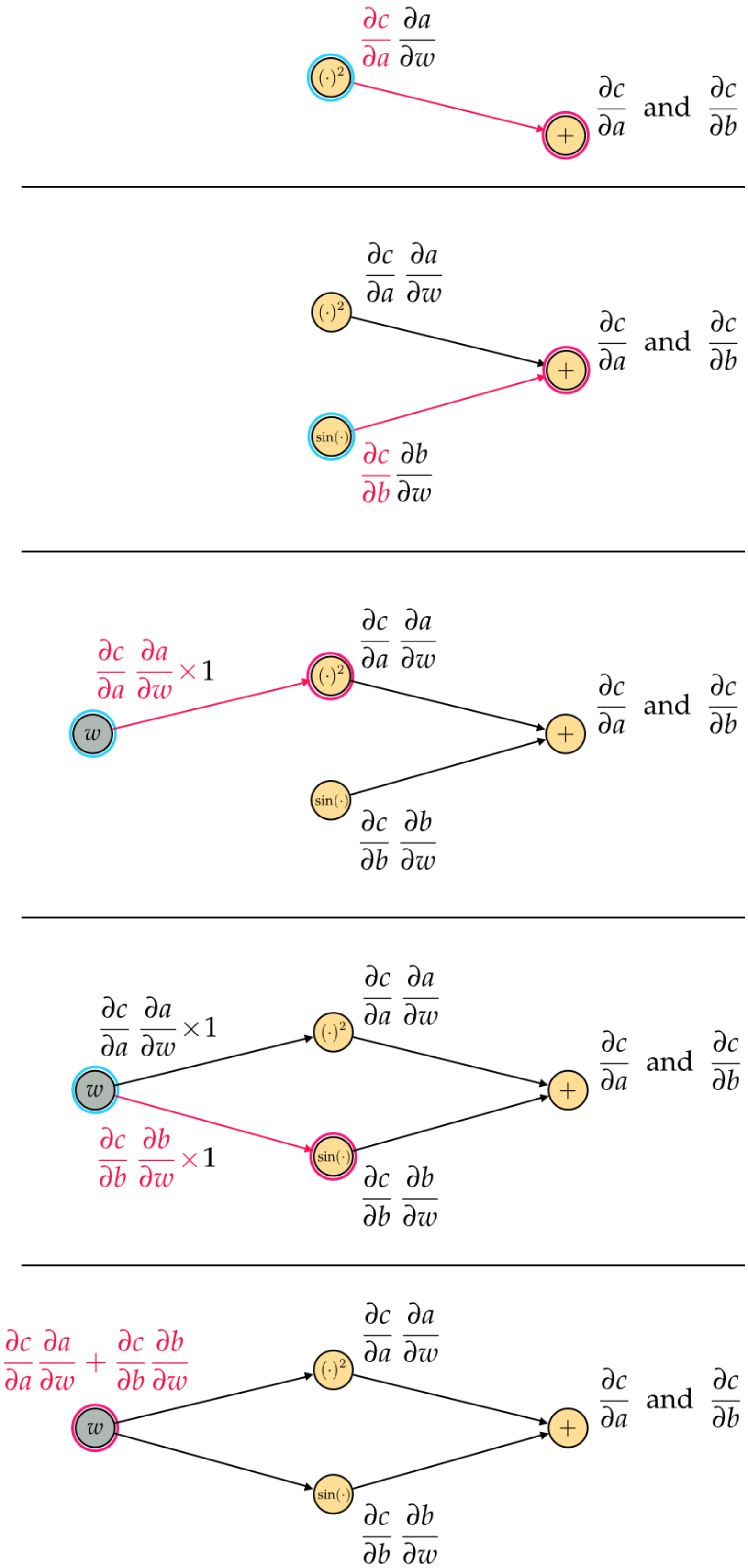

Figure 0.3 From top to bottom, the backward sweep of the reverse-mode differentiation of $g(w)$. Note in the bottom panel that sometimes a parent node might have more than one child (here the node w), in which case the derivative contributions of each of its children must be added – according to the chain rule – to compute the final derivative at that node.