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
45	lines 28, 29	Appendix Sextion	Appendix Section
51	Figure 3.4	The coordinate descent steps shown in the right panel are incorrect.	See Figure 3.4 below.
52	line 1	Here it takes two full sweeps through the variables to find the global minimum ...	Here it takes infinitely many sweeps through the variables to find the global minimum precisely, and at least three full sweeps to get reasonably close to it ...
60	Figure 3.8	In the top-right and bottom-right panels of the figure certain tickmarks are labeled incorrectly.	See Figure 3.8 below for the correct version.
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
81	line 18	(see Section 3.3.	(see Section 3.3).
90	line 2	`def newtons_method(g, max_its, w)`	`def newtons_method(g, max_its, w, **kwargs)`
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
147	equation (6.41)	$\left(\mathbf{x}_p' - \mathbf{x}_p\right)^T \boldsymbol{\omega} = \left\Vert \mathbf{x}_p' - \mathbf{x}_p \right\Vert_2 \left\Vert \boldsymbol{\omega} \right\Vert_2 = d \left\Vert \boldsymbol{\omega} \right\Vert_2$	$\left(\mathbf{x}_p' - \mathbf{x}_p\right)^T \boldsymbol{\omega} = -\left\Vert \mathbf{x}_p' - \mathbf{x}_p \right\Vert_2 \left\Vert \boldsymbol{\omega} \right\Vert_2 = -d \left\Vert \boldsymbol{\omega} \right\Vert_2$
148	equation (6.42)	LHS: $\beta - 0$	LHS: $0 - \beta$
162	line 20	we can use an identity function	we can use an indicator function
164	line 3	after a running	after running

page	location	incorrect	correct
166	equation (6.81)	$\mathcal{A}_{+1} = \frac{A}{A+C}$ $\mathcal{A}_{-1} = \frac{D}{B+D}.$	$\mathcal{A}_{+1} = \frac{A}{A+B}$ $\mathcal{A}_{-1} = \frac{D}{C+D}.$
166	line 13	called *precision*	called *sensitivity*
166	equation (6.82)	$\mathcal{A}_{\text{balanced}} = \frac{1}{2}\frac{A}{A+C} + \frac{1}{2}\frac{D}{B+D}.$	$\mathcal{A}_{\text{balanced}} = \frac{1}{2}\frac{A}{A+B} + \frac{1}{2}\frac{D}{C+D}.$
168	line 10	notation used used in Section 7.6	notation used in Section 5.4
168	equation (6.83)	model $(x_p, \mathbf{w})$	model $(\mathbf{x}_p, \mathbf{w})$
169	line 17	$\Omega - 1$	Ω_{-1}
172	exercise 6.11	*already given in Equation (6.3.2)*	*already given in Equation (6.28)*
178	line 20	*farthest*from	*farthest* from
200	line 9	identity function	indicator function
207	exercise 7.10	If we set the weights the of cost function	If we set the weights of the cost function
212	equation (8.10)	$\mathbf{C}\,\mathbf{C}^T = \mathbf{I}_{N\times N}$	$\mathbf{C}^T\mathbf{C} = \mathbf{I}_{N\times N}$
243	Figure 9.5	Legend reads: BoW, BoW + Char feats, BoW + Char + spam feats	Legend should read: BoW, BoW + char. freqs., BoW + char. freqs. + spam feats.
258	line 4	$\mathbf{D}^{-1/2}$	$\mathbf{D}^{-\frac{1}{2}}$
258	equation (9.6)	$\mathbf{D}^{-{}^{1}/_{2}}$	$\mathbf{D}^{-\frac{1}{2}}$
265	line 32	reguarlizers	regularizers
268	line 5	(as illustrated in the bottom panel of Figure 9.22)	(as illustrated in the bottom panel of Figure 9.22 where $\lambda = 130$)
269	exercise 9.1	experiment described in Example 1.8	experiment described in Example 9.2
280	line 13	modern reenactment[45]	modern reenactment [45]
287	line 19	/low-dimensional	low-dimensional
311	line 7	The inverse problem on other hand	The inverse problem on the other hand
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

page	location	incorrect	correct
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	line 23	than value of	than the value of
502	line 11	In this section we discuss a two	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.
535	equation (B.64)	$h(\mathbf{w}^0)$	$h(\mathbf{w})$
537	equation (B.67)	$g((w)$	$g(w)$
539	line 18	`w = 0`	`w = 0.0`
552	caption of Figure C.4	defined in Equation (C.23)	defined in Equation (C.24)

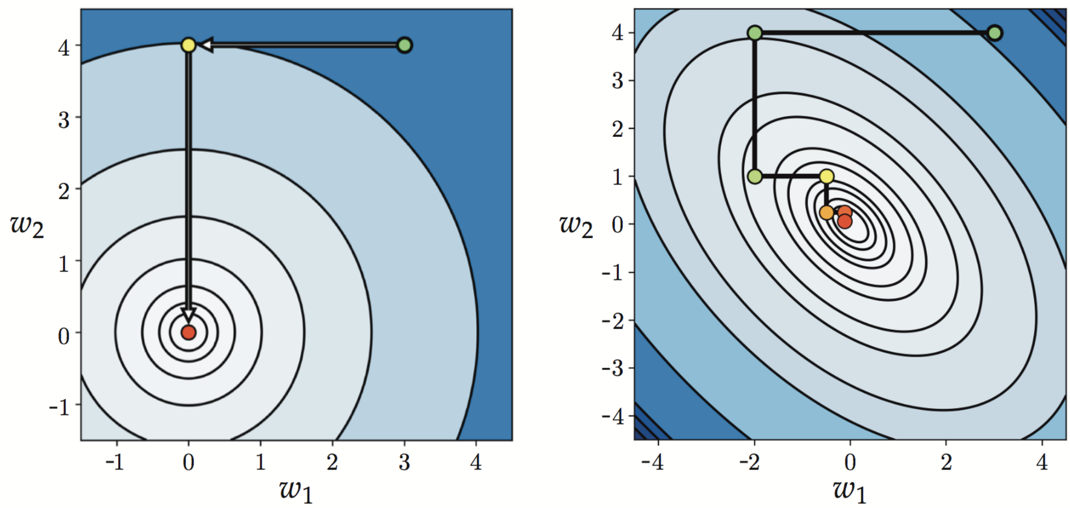

Figure 3.4 Figure associated with Example 3.5. See text for details.

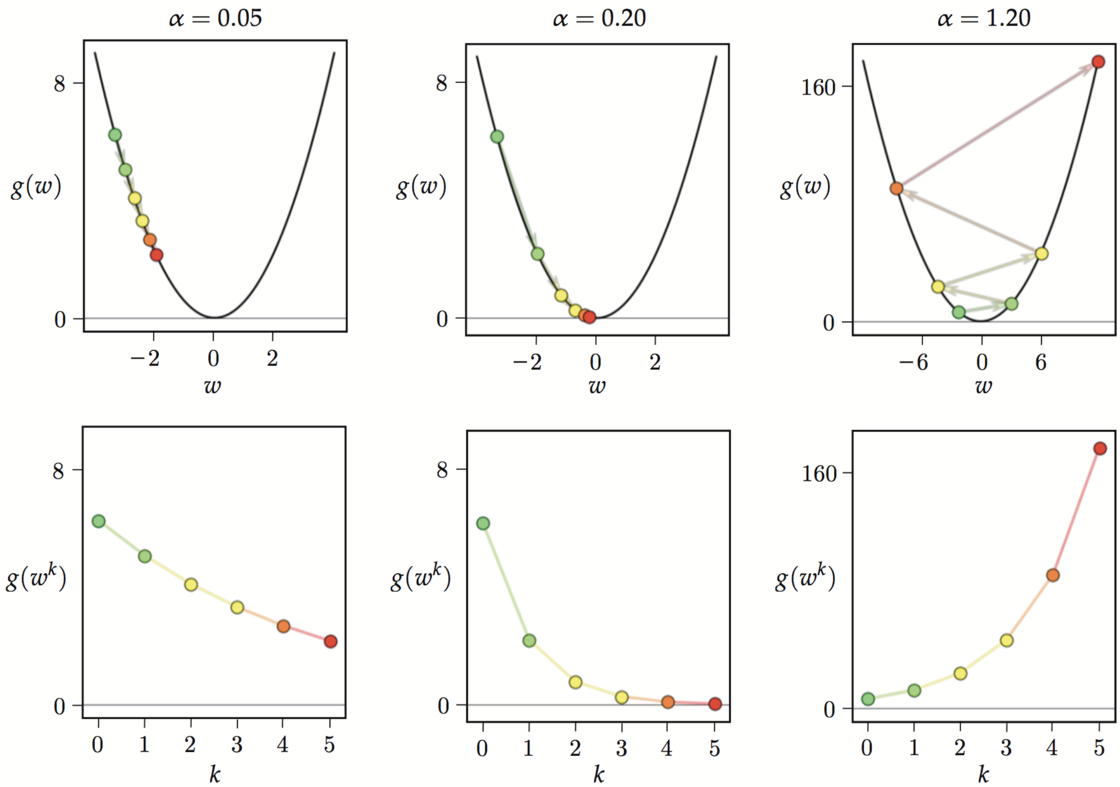

Figure 3.8 Figure associated with Example 3.9. See text for details.

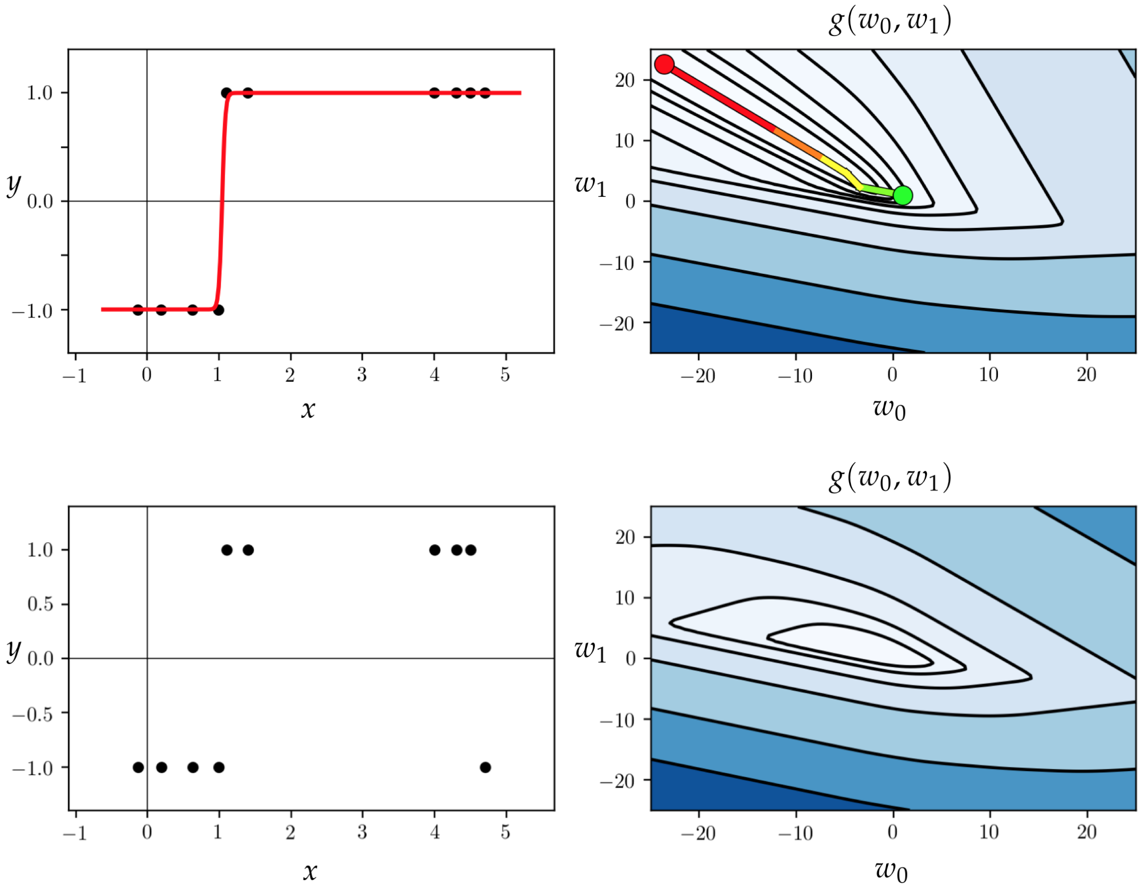

Figure 6.13 Figure associated with Example 6.6. See text for details.

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 0.4, 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 0.4.

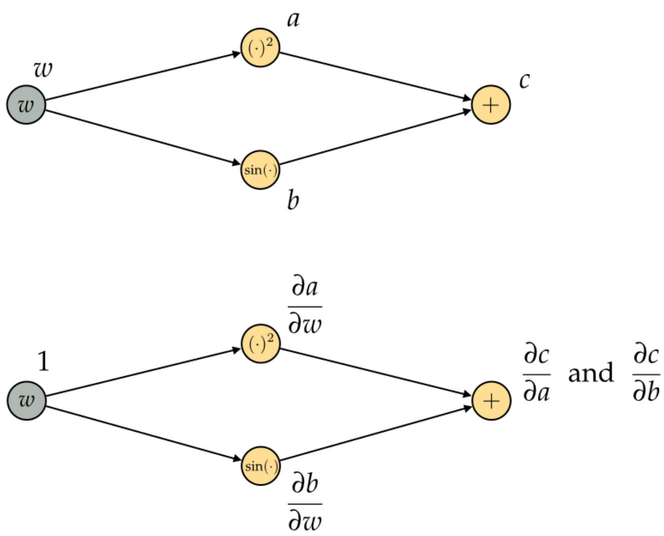

Figure 0.4 (top panel) The computation graph associated with the function $g(w) = w^2 + \sin(w)$. (bottom panel) The result of the full forward sweep of the reverse-mode differentiation of $g(w)$ with respect to w.

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 0.5. 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

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

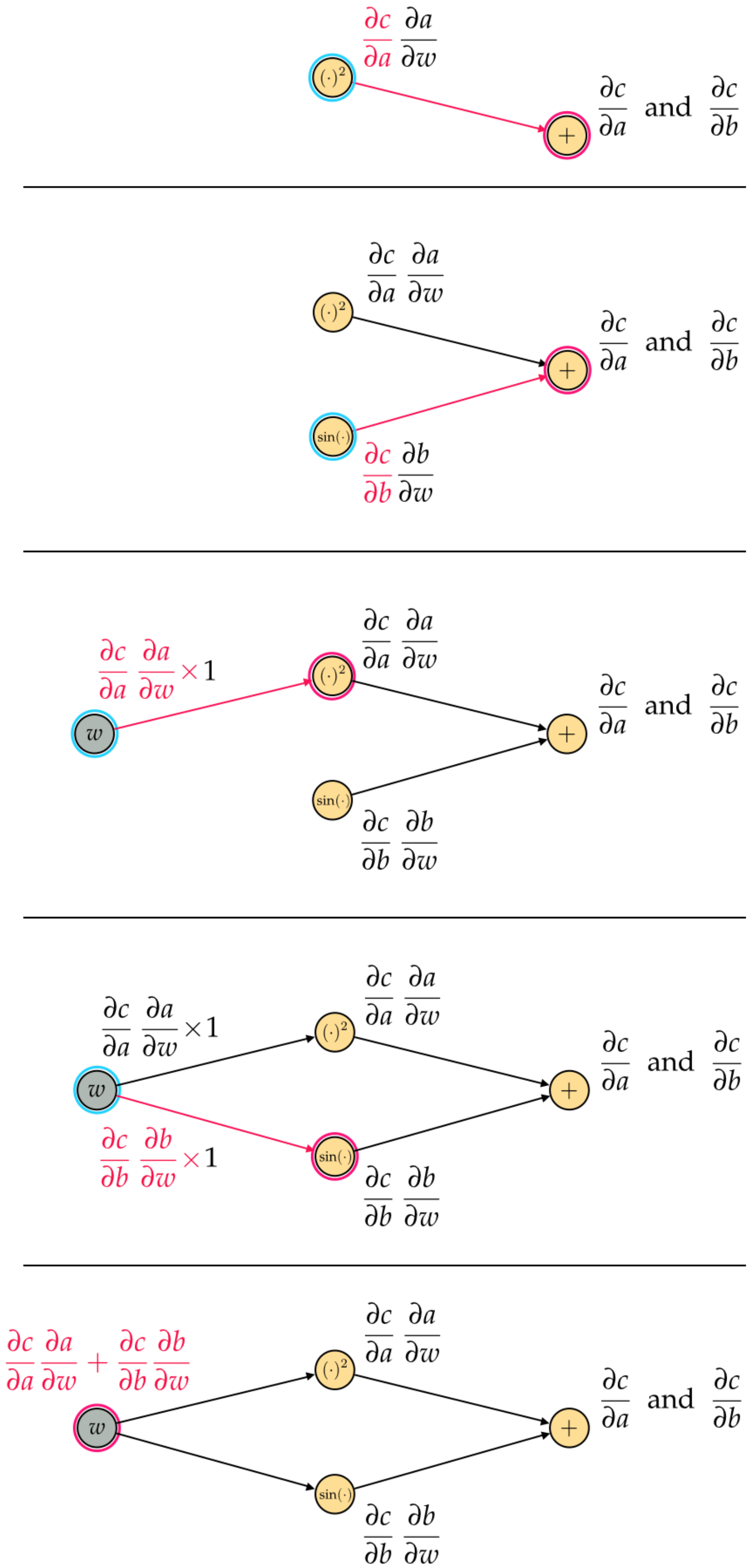

Figure 0.5 From top to bottom, the backward sweep of the reverse-mode differentiation of $g(w)$. Note in the bottom panel that sometimes a parent node might have more than one child (here the node w), in which case the derivative contributions of each of its children must be added – according to the chain rule – to compute the final derivative at that node.