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
xiv	line 34	(backpropogation)	(backpropagation)
xvii	line 15	foward/backward	forward/backward
xxii	line 3	contributers	contributors
10	line 5	industiral	industrial
12	line 20	distiniguish	distinguish
13	line 25	fradulent	fraudulent
14	line 19	differnt	different
22	line 18	evaluation of is each	evaluation of each is
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 ...
54	line 14	diretions	directions
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)`
95	Exercise 4.8 (a)	does indeed decrease the evaluation of g, i.e., $g\left(\mathbf{w}^k\right) \leq g\left(\mathbf{w}^{k-1}\right)$	is always a descent direction
109	line 15	shown in black	shown in blue
113	line 14	metrics of around 4500 and 3000	metrics of around 4.7 and 3.1
113	line 20	Autombile	Automobile
119	equation (5.40)	**b** in equation (5.40) is a column vector. It should be a row vector instead.	$\mathbf{b} = \begin{bmatrix} w_{0,0} & w_{0,1} & \cdots & w_{0,C-1} \end{bmatrix}$

page	location	incorrect	correct
120	last line	Next in Section 5.5 we described *weighted regression*, a twist on the standard scheme that allows for complete control over how each point is emphasized during regression. Finally in Section 5.6 we discussed various metrics for quantifying the quality of a trained regression model.	Next in Section 5.4 we discussed various metrics for quantifying the quality of a trained regression model. Then in Section 5.5 we described *weighted regression*, followed by a discussion of *multi-output regression* in Section 5.6.
122	exercise 5.4	circumstancs	circumstances
131	line 15	Here a of run normalized	Here a run of normalized
136	line 5	*mislcassified*	*misclassified*
140	Figure 6.10	The phrase "with three *noisy* data points pointed to by small arrows" should be removed from the figure caption.	
143	equation (6.37)	$g\left(\mathbf{w}\right) = \sum_{p=1}^{P} \log\left(1 + e^{-y_p \mathring{\mathbf{x}}_p^T \mathbf{w}}\right)$	$g\left(\mathbf{w}\right) = \frac{1}{P}\sum_{p=1}^{P} \log\left(1 + e^{-y_p \mathring{\mathbf{x}}_p^T \mathbf{w}}\right)$
145	line 12	since $e^{-C} < 1$ and so $e^{-y_p \mathring{\mathbf{x}}_p^T C\mathbf{w}^0} = e^{-C} e^{-y_p \mathring{\mathbf{x}}_p^T \mathbf{w}^0} < e^{-y_p \mathring{\mathbf{x}}_p^T \mathbf{w}^0}$	since $C\left(-y_p \mathring{\mathbf{x}}_p^T \mathbf{w}^0\right) < -y_p \mathring{\mathbf{x}}_p^T \mathbf{w}^0$ and so $e^{-y_p \mathring{\mathbf{x}}_p^T C\mathbf{w}^0} < e^{-y_p \mathring{\mathbf{x}}_p^T \mathbf{w}^0}$
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. Note: the phrase "(top row)" is removed from the figure caption.
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: $0 - \beta$
152	equations (6.48 - 6.54)	$\mathring{\mathbf{x}}^T \mathbf{w}$	$\mathring{\mathbf{x}}_p^T \mathbf{w}$
152	line 16	$y_p \mathring{\mathbf{x}}^T \mathbf{w} \geq 1$	$y_p \mathring{\mathbf{x}}_p^T \mathbf{w} \geq 1$
152	line 19	Summing up all P equations	Taking the average of all P equations
152	equation (6.50)	$g\left(\mathbf{w}\right) = \sum_{p=1}^{P} \max\left(0,\, 1 - y_p\, \mathring{\mathbf{x}}^T \mathbf{w}\right)$	$g\left(\mathbf{w}\right) = \frac{1}{P}\sum_{p=1}^{P} \max\left(0,\, 1 - y_p\, \mathring{\mathbf{x}}_p^T \mathbf{w}\right)$
152	line 32	$1 - y_p \mathring{\mathbf{x}}^T \mathbf{w}$	$1 - y_p \mathring{\mathbf{x}}_p^T \mathbf{w}$

page	location	incorrect	correct
153	line 4	$1 - y_p\left(\mathring{\mathbf{x}}^T\mathbf{w}\right)$	$1 - y_p\mathring{\mathbf{x}}_p^T\mathbf{w}$
153	equation (6.53)	$g\left(\mathbf{w}\right) = \sum_{p=1}^{P}\max\left(0, \epsilon - y_p\,\mathring{\mathbf{x}}^T\mathbf{w}\right)$	$g\left(\mathbf{w}\right) = \frac{1}{P}\sum_{p=1}^{P}\max\left(0, \epsilon - y_p\,\mathring{\mathbf{x}}_p^T\mathbf{w}\right)$
153	line 14	$\log\left(1 + e^{\epsilon - y_p\mathring{\mathbf{x}}^T\mathbf{w}}\right) \approx \log\left(1 + e^{-y_p\mathring{\mathbf{x}}^T\mathbf{w}}\right)$	$\log\left(1 + e^{\epsilon - y_p\mathring{\mathbf{x}}_p^T\mathbf{w}}\right) \approx \log\left(1 + e^{-y_p\mathring{\mathbf{x}}_p^T\mathbf{w}}\right)$
155	equation (6.57)	$\left(w_0 + \mathbf{x}_1^T\mathbf{w}\right) - \left(w_0 + \mathbf{x}_2^T\mathbf{w}\right) = \left(\mathbf{x}_1 - \mathbf{x}_2\right)^T\boldsymbol{\omega} = 2$	$\left(b + \mathbf{x}_1^T\boldsymbol{\omega}\right) - \left(b + \mathbf{x}_2^T\boldsymbol{\omega}\right) = \left(\mathbf{x}_1 - \mathbf{x}_2\right)^T\boldsymbol{\omega} = 2$
156	equation (6.60)	$g\left(b, \boldsymbol{\omega}\right) = \sum_{p=1}^{P}\max\left(0, 1 - y_p\left(b + \mathbf{x}_p^T\boldsymbol{\omega}\right)\right) + \lambda\left\Vert\boldsymbol{\omega}\right\Vert_2^2$	$g\left(b, \boldsymbol{\omega}\right) = \frac{1}{P}\sum_{p=1}^{P}\max\left(0, 1 - y_p\left(b + \mathbf{x}_p^T\boldsymbol{\omega}\right)\right) + \lambda\left\Vert\boldsymbol{\omega}\right\Vert_2^2$
157	equation (6.61)	$g\left(b, \boldsymbol{\omega}\right) = \sum_{p=1}^{P}\log\left(1 + e^{-y_p\left(b + \mathbf{x}_p^T\boldsymbol{\omega}\right)}\right) + \lambda\left\Vert\boldsymbol{\omega}\right\Vert_2^2.$	$g\left(b, \boldsymbol{\omega}\right) = \frac{1}{P}\sum_{p=1}^{P}\log\left(1 + e^{-y_p\left(b + \mathbf{x}_p^T\boldsymbol{\omega}\right)}\right) + \lambda\left\Vert\boldsymbol{\omega}\right\Vert_2^2.$
158	equation (6.62)	$y_p = 0 \longleftarrow \mathbf{y}_p = \begin{bmatrix}1\\0\end{bmatrix} \quad y_p = 1 \longleftarrow \mathbf{y}_p = \begin{bmatrix}0\\1\end{bmatrix}.$	$y_p = 0 \longrightarrow \mathbf{y}_p = \begin{bmatrix}0\\1\end{bmatrix} \quad y_p = 1 \longrightarrow \mathbf{y}_p = \begin{bmatrix}1\\0\end{bmatrix}.$
162	line 20	we can use an identity function	we can use an indicator function
164	line 3	after a running	after running
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 8	Section 6.24	Section 5.5
168	line 10	notation used used in Section 7.6	notation used in Section 5.4
168	equation (6.83)	$g\left(\mathbf{w}\right) = \sum_{p=1}^{P}\beta_p\log\left(1 + e^{-y_p\text{model}\left(x_p,\mathbf{w}\right)}\right).$	$g\left(\mathbf{w}\right) = \frac{1}{\beta_1+\beta_2+\cdots+\beta_P}\sum_{p=1}^{P}\beta_p\log\left(1 + e^{-y_p\text{model}\left(\mathbf{x}_p,\mathbf{w}\right)}\right).$
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
189	line 22	Percpetron	Perceptron
191	equation (7.33)	Redundant dot before the equality sign should be removed	

page	location	incorrect	correct
197	equation (7.40)	$y_p = 0 \longleftarrow \mathbf{y}_p = \left[1,\, 0,\, \cdots 0,\, 0\right]$ $y_p = 1 \longleftarrow \mathbf{y}_p = \left[0,\, 1,\, \cdots 0,\, 0\right]$ $\vdots$ $y_p = C-1 \longleftarrow \mathbf{y}_p = \left[0,\, 0,\, \cdots 0,\, 1\right].$	$y_p = 0 \longrightarrow \mathbf{y}_p = \left[1,\, 0,\, \cdots\, 0,\, 0\right]$ $y_p = 1 \longrightarrow \mathbf{y}_p = \left[0,\, 1,\, \cdots\, 0,\, 0\right]$ $\vdots$ $y_p = C-1 \longrightarrow \mathbf{y}_p = \left[0,\, 0,\, \cdots\, 0,\, 1\right].$
200	line 9	identity function	indicator function
203	line 5	a sum of P terms	a sum (or an average) of P terms
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 is incorrect	See Figure 9.5 below
251	line 14	descen	descent
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 4	By the time $\lambda \;\approx\; 40$, five major weights remain, corresponding to features 1, 2, 3, 6, and 7 (as illustrated in the bottom panel of Figure 9.22). The first four of these features were also determined to be important via boosting in Example 9.7.	By the time $\lambda \;\approx\; 40$, four major weights remain, corresponding to features 1, 2, 3, and 6 (as illustrated in the bottom panel of Figure 9.22). Note that these features were also determined to be important via boosting in Example 9.7.
268	Figure 9.22	Incorrect figure	See Figure 9.22 below
269	line 15	stregnth	strength
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
297	lines 14, 15	Au-toencder	Au-toencoder
311	line 7	The inverse problem on other hand	The inverse problem on the other hand
363	line 8	coordintes	coordinates
371	line 1	netwok	network

page	location	incorrect	correct
371	Figure 11.47	Some panel titles in the figure are incorrect	See Figure 11.47 below
377	line 9	regluarization	regularization
399	line 1	occured	occurred
421	line 12	dmension	dimension
422	line 3	Example 11.9.2	Example 11.15
425	line 24	ouput	output
438	line 3	graident	gradient
440	line 7	backpropogation	backpropagation
461	line 12	occured	occurred
473	line 3	weaknessess	weaknesses
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.
494	line 6	cab be written	can be written
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.	we update the partial derivative of each parent by multiplying it by the partial derivative of its child node with respect to that parent (when the parent node has multiple children the accumulated partials should be added, not multiplied, since this is what the chain rule requires).
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)
564	[12]	*Marchine*	*Machine*
570	G	graident descent	gradient descent

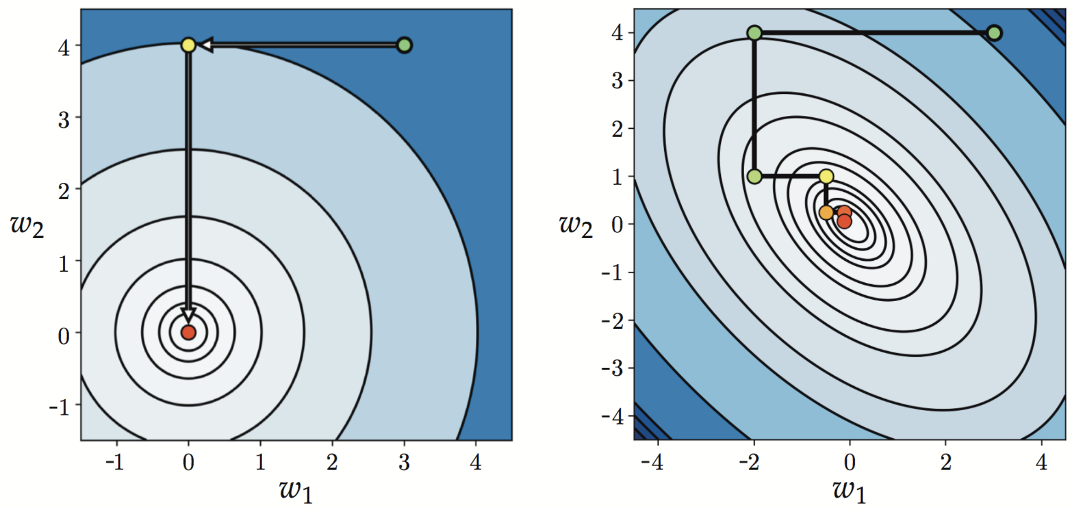

Figure 3.4 *Figure caption remains the same*

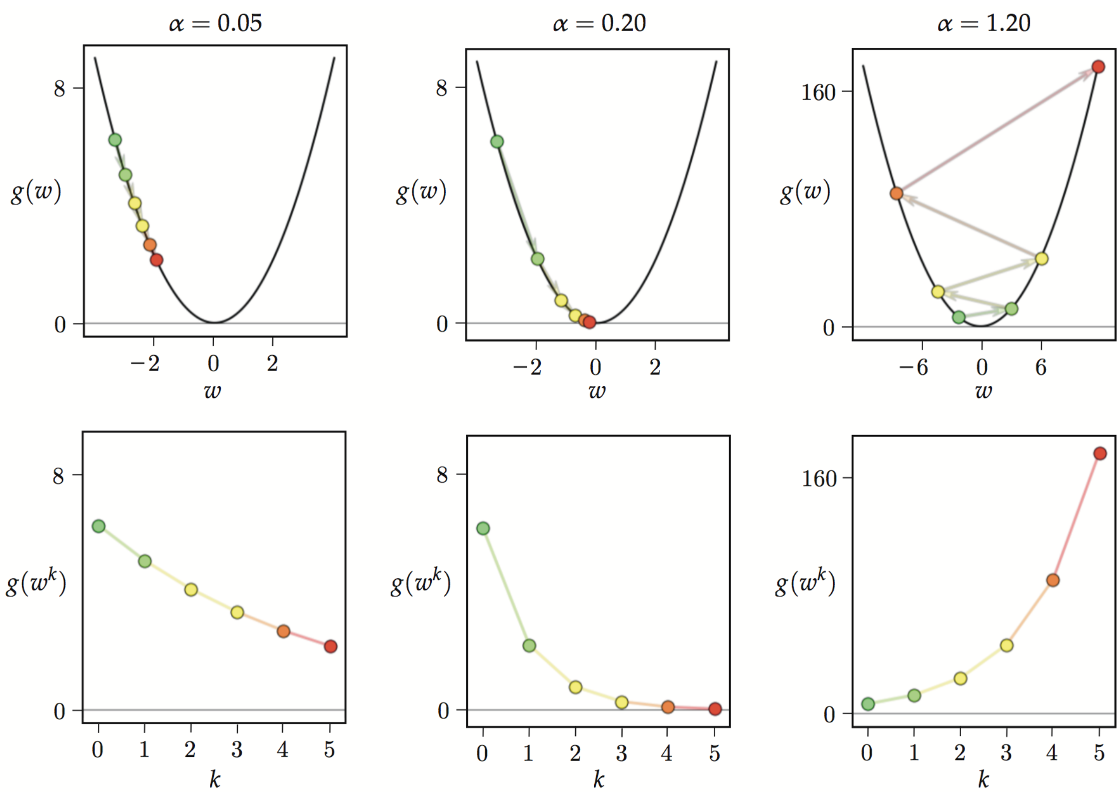

Figure 3.8 *Figure caption remains the same*

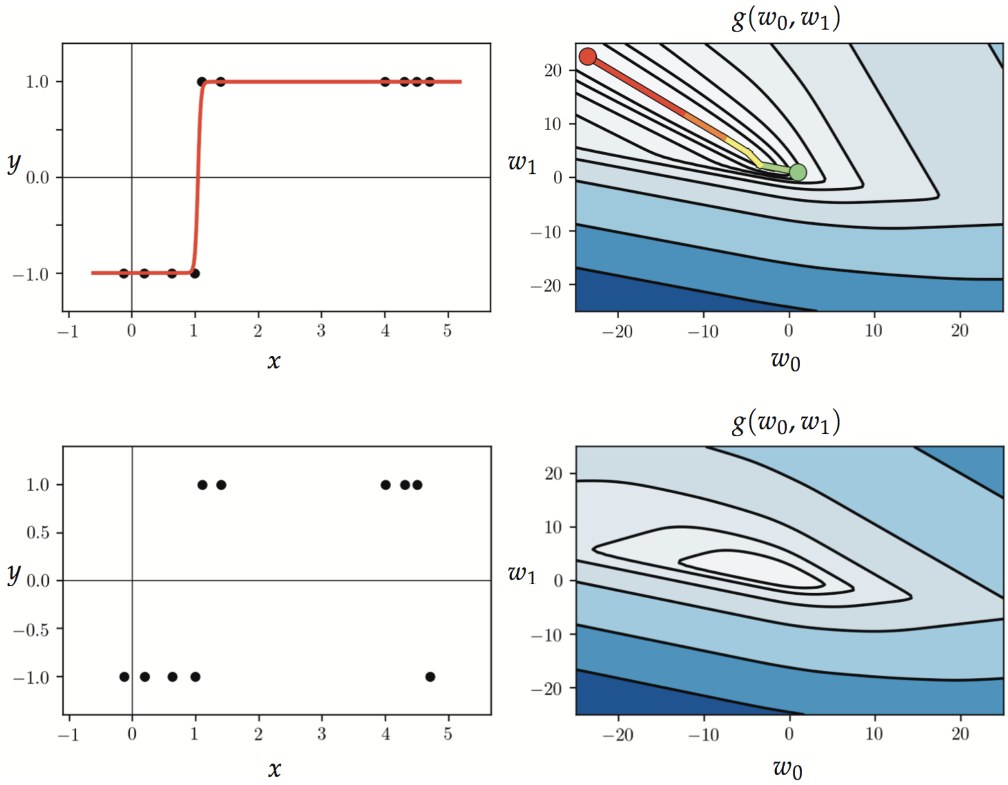

Figure 6.13 Figure associated with Example 6.6. See text for details.

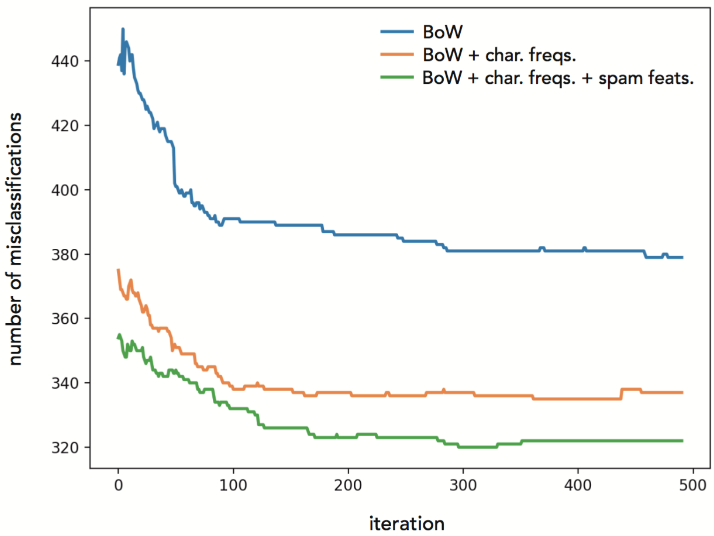

Figure 9.5 *Figure caption remains the same*

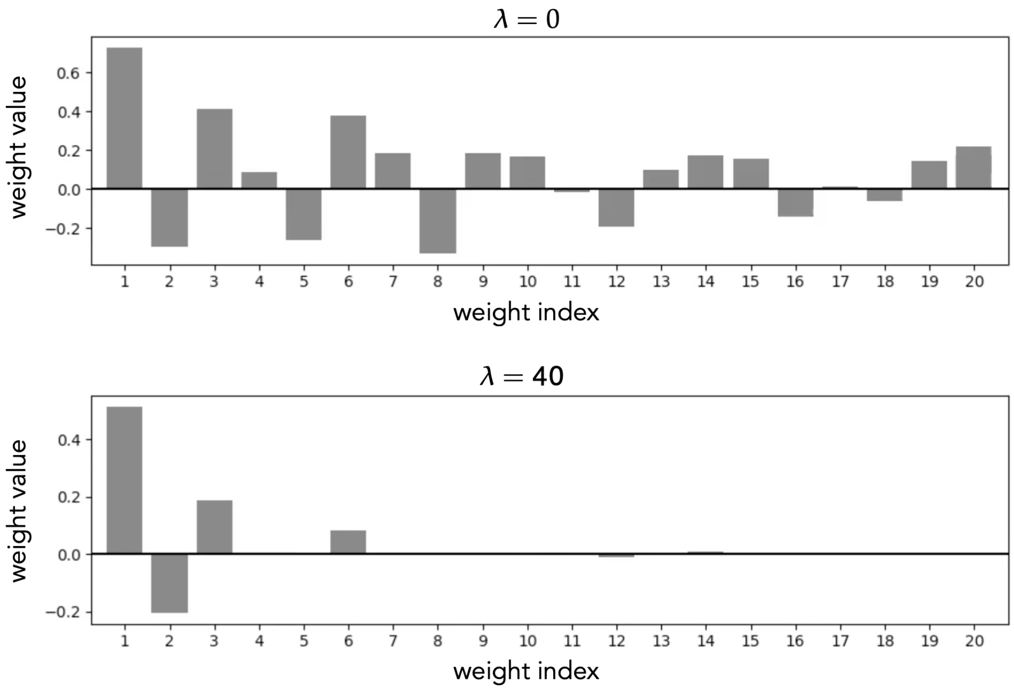

Figure 9.22 *Figure caption remains the same*

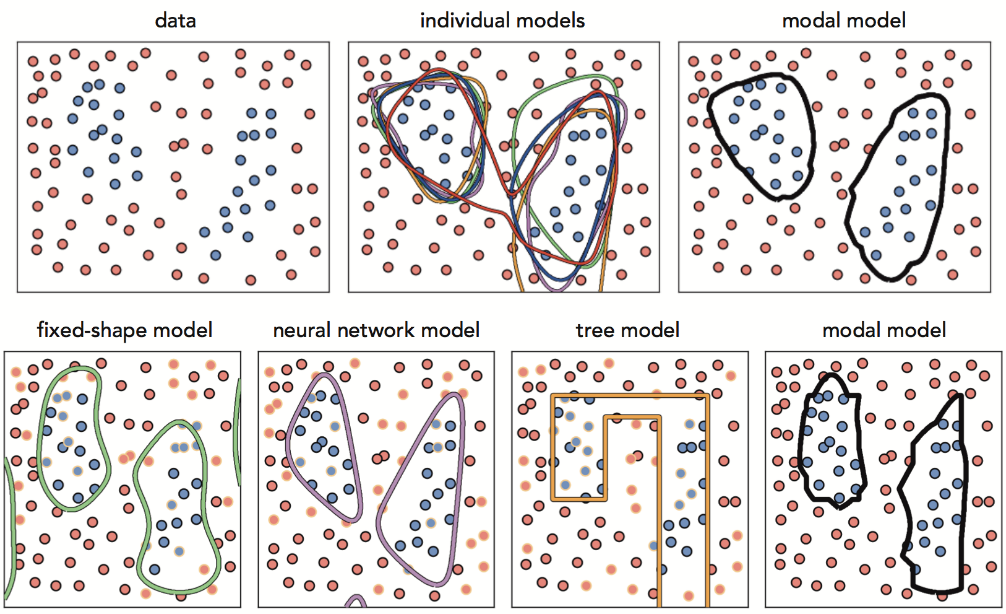

Figure 11.47 *Figure caption remains the same*