

# Pattern Recognition and Machine Learning

## Errata and Additional Comments

Markus Svensén and Christopher M. Bishop

September 9, 2009



## Preface

This document lists corrections and clarifications for the second printing<sup>1</sup> of *Pattern Recognition and Machine Learning* by Christopher M. Bishop, published by Springer in 2006. It is intended to be complete, in that it includes also trivial typographical errors and provides clarifications that some readers may find helpful. However, it is not assumed to include all mistakes that exist in the book and the author welcomes reports of any remaining potential mistakes, along with any other feedback on the book, which should be sent to

`prml-fb@microsoft.com`

Corrections and clarifications are given in the order they should appear in the book. Each entry starts with a page number in the margin, followed (in the main body of the page) by the location of the mistake or ambiguity and the required amendment. In specifying the location, the following conventions are used:

- Paragraphs are numbered from 1 on each page. The first paragraph is usually the one continuing from the previous page, but if the first line on a page starts a new paragraph, this will be the first paragraph. In the book, the first line of each paragraph is indented, with the exception of paragraphs that follow immediately after a chapter or a section (but not a sub-section) heading, which are not indented.
- Line and paragraph numbers preceded by a minus (–) sign are counted from the bottom of the paragraph or page. ‘Paragraph –1’ refers to the last paragraph started, but not necessarily completed on a page.
- The following abbreviations are used in this document: PRML (Pattern Recognition and Machine Learning), l.h.s. (left hand side) and r.h.s. (right hand side).

---

<sup>1</sup>To identify which printing your copy of the book is from, consult the page with bibliographic information (immediately preceding the dedication page); if the one but last line reads “9 8 7 6 5 4 3 2 1” you have a copy from the first printing, if it reads “9 8 7 6 5 (corrected printing 2007)” you have a copy from the second printing.

### **Acknowledgements**

We would like to thank all of the readers who have reported mistakes in PRML. In particular, we are grateful to the Japanese translation team, Dr Xiaobo Jin of the Chinese Academy of Sciences, and also to Makoto Otsuka of Okinawa Institute of Science and Technology, Japan, and his colleagues in Neural Computation Unit, for particularly thorough feedback.

## Additional Notes

Although the majority of the changes are relatively straightforward, such as typographical mistakes, accidental changes of sign or missing terms in equations and incorrect references, a few call for a bit more explanation. Citations below refer to the References section of PRML.

### Bayesian “Estimate” of the Variance of a Gaussian

When we estimate the mean,  $\mu$ , and the variance,  $\sigma^2$  of a Gaussian from a data set using maximum likelihood, the estimate for the mean is unbiased, whereas the estimate for the variance is biased, as discussed in section 1.2.4. The bias in the variance is due to the use of the maximum likelihood estimate for the mean and disappears if the true mean is known, yielding the corresponding unbiased estimate for the variance.

Instead suppose we take a Bayesian approach and chose a *particular* prior distribution over  $\mu$  and  $\tau$  (the inverse variance,  $1/\sigma^2$ ) of the form,

$$p(\mu, \tau) = \mathcal{N}(\mu|\mu_0, \lambda_0^{-1}) \text{Gam}(\tau|a_0, b_0).$$

If we assume a fully factorized posterior over  $\mu$  and  $\tau$ , we can then integrate over  $\mu$  in the posterior distribution for the parameters to obtain a marginal  $p(\tau)$ . From this we can calculate  $\mathbb{E}[\tau]^{-1}$ , whose value equals that of the unbiased maximum likelihood estimate of  $\sigma^2$ . This is analogous to the result discussed by MacKay (2003). However, this is not a general consequence of taking Bayesian approach, but depends on the choice of prior and posterior. If we make an equally valid choice, given by the Gaussian-Gamma prior,

$$p(\mu, \tau) = \mathcal{N}(\mu|\mu_0, (\tau\lambda_0)^{-1}) \text{Gam}(\tau|a_0, b_0)$$

and again assume a fully factorised posterior, the value of  $\mathbb{E}[\tau]^{-1}$  will equal the biased maximum likelihood estimate for  $\sigma^2$ . However, if we consider the exact posterior, which is given by a Gaussian-Gamma distribution, we again obtain a value of  $\mathbb{E}[\tau]^{-1}$  equal to the unbiased maximum likelihood estimate for  $\sigma^2$ . Changes have been incorporated, in particular in Section 10.1.3 (pages 470–473), to reflect this.

### Variational Logistic Regression

Section 10.5 discusses local variational methods and a particular example, in the form of variational logistic regression, is discussed in section 10.6. Section 10.5 largely uses conventions from Jordan *et al.* (1999), whereas 10.6 largely follows Jaakkola and Jordan (2000). Unfortunately, the use of different conventions regarding the sign of the variational parameters lead to inconsistencies in some of the equations in Section 10.5. In order to correct these, while adhering to conventions from existing literature as far as possible, the symbol used for the variational parameters in Section 10.5 (pages 493–498) has been changed from  $\lambda$  to  $\eta$ , while  $\lambda$  has been kept throughout 10.6.

---

## Corrections

---

- Page 47** Line –2: “marginalize to find” should be replaced by “calculate”.
- Page 48** Line 2: “marginalize to find” should be replaced by “calculate”.
- Page 52** Equation (1.100):  $\partial\tilde{H}$  should be  $\partial^2\tilde{H}$  in the numerator on the l.h.s.
- Page 53** Equation (1.103): A minus sign (‘–’) should be added to the l.h.s.
- Page 53** Biography of L. Boltzmann, column 1, line –1: “they day” should be “the day”.
- Page 53** Biography of L. Boltzmann, column 2, line –6: “lead” should be “led”.
- Page 63** Exercise 1.20, Equation (1.149):
- $$-\frac{3\epsilon^2}{2\sigma^2} \quad \text{should be} \quad -\frac{\epsilon^2}{\sigma^2}$$
- in the argument of the exponential on the r.h.s.
- Page 81** Second paragraph, Line 2: The reference to (2.51) should refer to (2.50).
- Page 83** Line 2:  $\boldsymbol{\mu}^T \mathbf{z}$  should be  $\mathbf{z} \boldsymbol{\mu}^T$ .
- Page 89** Equation (2.87): The last line should read
- $$+\mathbf{x}_a^T (\boldsymbol{\Lambda}_{aa} - \boldsymbol{\Lambda}_{ab} \boldsymbol{\Lambda}_{bb}^{-1} \boldsymbol{\Lambda}_{ba}) \boldsymbol{\mu}_a + \text{const}$$
- (incorrect inverse (‘<sup>-1</sup>’) removed).
- Page 96** Equation (2.133):  $\mathbf{x}_n$  should be  $x_n$ .
- Page 106** Figure 2.17:  $\bar{\mathbf{x}}$ ,  $\bar{r}$  and  $\bar{\boldsymbol{\theta}}$  should be  $\bar{\mathbf{x}}$ ,  $\bar{r}$  and  $\bar{\boldsymbol{\theta}}$ , respectively.
- Page 108** Second line before Equation (2.180): “zeroth-order Bessel function” should be “zeroth-order modified Bessel function”.
- Page 114** First line after Equation (2.204): “ $\mathbf{x} = (x_1, \dots, x_N)^T$ ” should be “ $\mathbf{x} = (x_1, \dots, x_M)^T$ ”.
- Page 115** Line before Equation (2.215):  $(\eta_1, \dots, \eta_{M-1})^T$  should be  $(\eta_1, \dots, \eta_{M-1}, 0)^T$ .
- Page 116** Equation (2.222):  $h(\mathbf{x})$  should be  $h(x)$  on the l.h.s.
- Page 116** Last line before Equation (2.227):  $\mathbf{x}_n$  should be  $\mathbf{x}_N$ .
- Page 117** Section 2.4.2, line –3: “a effective” should be “an effective”.
- Page 118** Line 4: “Jeffries” should be “Jeffreys” and “Tao” should be “Tiao”.
- Page 123** Equation (2.250): The exponent in the denominator of normalizing constant of the Gaussian kernel on the r.h.s. should be  $D/2$  (not  $1/2$ ).

## 8 Corrections for pages 129–211

- Page 129** Exercise 2.7, line 3: “mean value of  $x$ ” should be “mean value of  $\mu$ ”.
- Page 133** Equation (2.291):  $\mathbb{E}[\mathbf{x}_n \mathbf{x}_m]$  should be  $\mathbb{E}[\mathbf{x}_n \mathbf{x}_m^T]$ .
- Page 139** Second line after Equation (3.6): The inline equation should read:  $\tanh(a) = 2\sigma(2a) - 1$ .
- Page 143** Section 3.1.2, line 8:  $\phi_j(\mathbf{x}_n)$  should be  $\varphi_j$ .
- Page 148** Equation (3.37): The second integral on the r.h.s. should be a double integral.
- Page 149** Equation (3.44): The integral on the r.h.s. should be a double integral.
- Page 161** Second paragraph, line 2: “the form an” should be “the form of an”.
- Page 163** Second sentence after Equation (3.72): “decrease” and the last occurrence of “increase” should exchange their positions in the sentence.
- Page 168** Figure 3.1, caption: “model evidence” should be “model log evidence”.
- Page 168–169** Equation (3.89): The final part of the sentence containing the equation (the part following after the equation) should be removed and the equation should end in a period (‘.’).
- Page 182** Section 4.1.2, line 2: “be tempted be” should be “be tempted”.
- Page 184** Figure 4.3:  $\hat{\mathbf{x}}$  should be  $\hat{\mathbf{x}}$ .
- Page 197** Figure 4.9, caption: “probit function” should be “inverse probit function”.
- Page 203** Equations (4.85) and (4.86): There is a factor of  $1/s$  missing in the first term on the r.h.s. in both equations.
- Page 207** First paragraph, line 7: “concave” should be “convex”.
- Page 208** First paragraph, line –2: “concave” should be “convex”.
- Page 211** End of sentence following Equation (4.114): This should read: “which is known as the *inverse probit* function.”
- Page 211** Equation (4.115): This equation should read
- $$\operatorname{erf}(a) = \frac{2}{\sqrt{\pi}} \int_0^a \exp(-\theta^2) d\theta$$
- (factor of  $1/2$  removed from the argument of the exponential on the r.h.s.).
- Page 211** Last line before Equation (4.116): “probit function” should be “inverse probit function”.

**Page 211** Equation (4.116): This equation should read

$$\Phi(a) = \frac{1}{2} \left\{ 1 + \operatorname{erf} \left( \frac{a}{\sqrt{2}} \right) \right\}.$$

Note that  $\Phi$  should be  $\Phi$  (i.e. not bold) on the l.h.s.

**Page 213** Equation (4.124):  $\nabla \ln E(\mathbf{w})$  should be  $\nabla E(\mathbf{w})$  on the l.h.s.

**Page 218** Equation (4.143):  $\mathbf{S}_N$  should be  $\mathbf{S}_N^{-1}$  on the l.h.s.

**Page 219–220** Section 4.5.2: All instances of “probit function” should be replaced by “inverse probit function”.

**Page 222** Exercise 4.15, last line: “concave” should be “convex”.

**Page 222** Exercise 4.16, line 4:  $t$  should be  $t_n$ .

**Page 223–224** Exercises 4.21, 4.25 and 4.26: All instances of “probit function” should be replaced by “inverse probit function”.

**Page 227** Second paragraph, line 1: “described a” should be “described as a”.

**Page 235** Equation (5.24):  $t_{kn}$  should be  $t_{nk}$ .

**Page 238** Equation (5.32):  $=$  should be  $\simeq$ .

**Page 250** Equation (5.80):

$$\frac{\partial E^n}{\partial a_k} \quad \text{should be} \quad \frac{\partial E_n}{\partial a_k}.$$

**Page 250** Last line: “be” should be “by”.

**Page 256** Section 5.5, line 1: “outputs” should be “output”.

**Page 259** Line –2: The word ‘to’ should be omitted.

**Page 260** Figure 5.11: In all sub-figure titles, all numbers (1, 10, 100, 1000) on the r.h.s. of the ‘=’-signs should be raised to  $-2$ ; e.g. in the title of the lower left sub-figure, “ $\alpha_1^b = 100$ ”, should be “ $\alpha_1^b = 100^{-2}$ ” or, simpler, “ $\alpha_1^b = 10^{-4}$ ”.

**Page 262** Line –1: “approach 2” should be “approach 1”.

**Page 265** Figure 5.16, caption, line 5: Before ‘(c)’, insert: “where blue and yellow correspond to positive and negative values, respectively”.

**Page 266** Last Equation before Equation (5.131): The third term on the r.h.s. should be

$$\mathbb{E}[\xi^2] \frac{1}{2} \iint \left[ \{y(\mathbf{x}) - t\} \left\{ (\boldsymbol{\tau}')^T \nabla y(\mathbf{x}) + \boldsymbol{\tau}^T \nabla \nabla y(\mathbf{x}) \boldsymbol{\tau} \right\} + (\boldsymbol{\tau}^T \nabla y(\mathbf{x}))^2 \right] p(t|\mathbf{x}) p(\mathbf{x}) \, d\mathbf{x} \, dt.$$

10 Corrections for pages 266–366

**Page 266** Equation (5.132): This equation should read

$$\Omega = \frac{1}{2} \int \left[ \{y(\mathbf{x}) - \mathbb{E}[t|\mathbf{x}]\} \{(\boldsymbol{\tau}')^T \nabla y(\mathbf{x}) + \boldsymbol{\tau}^T \nabla \nabla y(\mathbf{x}) \boldsymbol{\tau}\} + (\boldsymbol{\tau}^T \nabla y(\mathbf{x}))^2 \right] p(\mathbf{x}) \, d\mathbf{x}$$

**Page 271** Equation (5.142): The numerator on the r.h.s. should read  $(\mu_j - w_i)$ .

**Page 282** Equation (5.183):  $+\text{const}$  on the r.h.s. should be omitted.

**Page 287** Exercise 5.21: The text in the exercise could be misunderstood; a less ambiguous formulation is: “Extend the expression (5.86) for the outer product approximation of the Hessian matrix to the case of  $K > 1$  output units. Hence, derive a form that allows (5.87) to be used to incorporate sequentially contributions from individual outputs as well as individual patterns. This, together with the identity (5.88), will allow the use of (5.89) for finding the inverse of the Hessian by sequentially incorporating contributions from individual outputs and patterns.”

**Page 314** Second paragraph, line 2:  $\mathbf{t}$  should be  $\mathbf{t}_N$ .

**Page 314** Second paragraph, line 6–7:  $\mathbf{t}_{N+1}$  and  $\mathbf{t}_N$  should be  $\mathbf{t}_{N+1}$  and  $\mathbf{t}_N$ , respectively.

**Page 316** Equation (6.80):  $+\text{const}$  on the r.h.s. should be omitted.

**Page 332** Equation (7.22):  $L(\mathbf{w}, b, \mathbf{a})$  on the l.h.s. should be  $L(\mathbf{w}, b, \boldsymbol{\xi}, \mathbf{a}, \boldsymbol{\mu})$ .

**Page 333** Two lines above Equation (7.33): ‘minimize’ should be ‘maximize’.

**Page 346** Equation (7.79):  $p(t_n|\mathbf{x}_n, \mathbf{w}, \beta^{-1})$  should be  $p(t_n|\mathbf{x}_n, \mathbf{w}, \beta)$  on the r.h.s.

**Page 350** Caption of Figure 7.10, line –2: “contribution” should be “contribution”.

**Page 350** Second paragraph, lines 8, 9 and 11:  $\mathbf{t}$  should be  $\mathbf{t}$ .

**Page 351** Equation (7.94):

$$|1 + \alpha_i^{-1} \boldsymbol{\varphi}_i^T \mathbf{C}_{-i}^{-1} \boldsymbol{\varphi}_i|$$

should be

$$(1 + \alpha_i^{-1} \boldsymbol{\varphi}_i^T \mathbf{C}_{-i}^{-1} \boldsymbol{\varphi}_i)$$

on the r.h.s.

**Page 352** Line 3:  $\varphi_n$  should be  $\varphi_i$ .

**Page 352** Line –2:  $j \neq i$  should be  $j \neq 1$ .

**Page 364** Equation (8.7):  $\mathbf{T}$  should be  $\mathbf{t}$  on the l.h.s.

**Page 365** Figure 8.7: The node labels  $\hat{x}$  and  $\hat{t}$  should read  $\hat{x}$  and  $\hat{t}$ , respectively.

**Page 366** Second paragraph, last line: “show” should be “shown”.

- Page 367** Line –6: The comma before “Similarly” should be replaced by a full stop.
- Page 380** Equation (8.35):  $\int_0^\infty$  should be  $\int_{-\infty}^\infty$  on the r.h.s.
- Page 383** Figure 8.26: The label  $x_i$  should be  $\mathbf{x}_i$  in the graph.
- Page 390** Figure 8.32(b): The labels of the two rightmost nodes,  $x_N$  and  $x_{N-1}$  should be swapped to match the ordering of the nodes in Figure 8.32(a).
- Page 390** Third paragraph, Line 2: “max-product” should be “max-sum”.
- Page 397** Equation (8.57): The ordering of the indices and the arguments of the  $\psi$  functions disagrees with the corresponding ordering used in other equations in this section. The correct form is

$$\begin{aligned} \mu_\beta(x_n) &= \sum_{x_{n+1}} \psi_{n,n+1}(x_n, x_{n+1}) \left[ \sum_{x_{n+2}} \cdots \right] \\ &= \sum_{x_{n+1}} \psi_{n,n+1}(x_n, x_{n+1}) \mu_\beta(x_{n+1}). \end{aligned}$$

- Page 398** Line 3:  $O(N^2 M^2)$  should be  $O(N^2 K^2)$ .
- Page 409** Equation (8.79): “ $\mu_{x_2 \rightarrow f_b}$ .” should be “ $\mu_{x_2 \rightarrow f_b}(x_2)$ .” on the r.h.s.
- Page 412** Unlabelled Equation between Equation (8.90) and Equation (8.91): The second line should read

$$= \frac{1}{Z} \max_{x_1} \left[ \max_{x_2} \left[ \psi_{1,2}(x_1, x_2) \left[ \cdots \max_{x_N} \psi_{N-1,N}(x_{N-1}, x_N) \right] \cdots \right] \right].$$

- Page 416** Paragraph 2, line 10:  
 $A-C-B-D-A$  is chord-less a link could be  
 should be  
 $A-C-B-D-A$  is chord-less and so a link should be
- Page 416** Paragraph 2, line 15: “join tree” should be “junction tree”.
- Page 416** Paragraph 2, line 22–23: The sentence starting “If the tree is condensed, . . .” should be omitted.
- Page 419** Exercise 8.6: The sentence fragment following (8.104) should read “where  $0 \leq \mu_i \leq 1$  for  $i = 0, \dots, M$ ”. Moreover, the last sentence of the exercise should be: Discuss the interpretation of the  $\mu_i$ s.
- Page 421** Exercise 8.16, line 1:  $p(\mathbf{x}_n | \mathbf{x}_N)$  should be  $p(x_n | x_N)$ , in order to agree with notation used in Section 8.4.1.

## 12 Corrections for pages 443–491

**Page 443** Equation (9.39): The first line of this equation should read

$$\mathbb{E}[z_{nk}] = \frac{\sum_{\mathbf{z}_n} z_{nk} \prod_{k'} [\pi_{k'} \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_{k'}, \boldsymbol{\Sigma}_{k'})]^{z_{nk'}}}{\sum_{\mathbf{z}_n} \prod_j [\pi_j \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)]^{z_{nj}}}.$$

**Page 444** Equation (9.41):  $D/2$  should replace  $M/2$  in the denominator of the normalisation constant on the r.h.s.

**Page 446** Equation (9.56): The first line of this equation should read

$$\gamma(z_{nk}) = \mathbb{E}[z_{nk}] = \frac{\sum_{\mathbf{z}_n} z_{nk} \prod_{k'} [\pi_{k'} p(\mathbf{x}_n | \boldsymbol{\mu}_{k'})]^{z_{nk'}}}{\sum_{\mathbf{z}_n} \prod_j [\pi_j p(\mathbf{x}_n | \boldsymbol{\mu}_j)]^{z_{nj}}}$$

**Page 449** Line 2: The final clause “, and  $y(\mathbf{x}, \mathbf{w})$  is given by (3.3)” should be omitted.

**Page 449** Last paragraph, line 4:  $\alpha$  should be  $\alpha$ .

**Page 449** Equation (9.66): A pair of braces is missing; the correct form is

$$\mathbb{E}_{\mathbf{w}} [\ln \{p(\mathbf{t} | \mathbf{X}, \mathbf{w}, \beta) p(\mathbf{w} | \alpha)\}].$$

**Page 450** Equation (9.68):  $\mathbf{m}_N$  should be  $\mathbf{m}$  on the r.h.s.

**Page 452** First line after Equation (9.74): The word “negative” should be omitted.

**Page 453** Line 3:  $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(\text{old})})$  should be  $\mathcal{L}(q, \boldsymbol{\theta}^{(\text{old})})$ .

**Page 462** Equation (10.1): A minus-sign ( $-$ ) should be inserted before the integral on the r.h.s.

**Page 467** Equation (10.12):  $q^*(z_1)$  should be  $q_1^*(z_1)$  on the l.h.s.

**Page 479** Equation (10.69):  $\alpha_k$  should be  $\alpha_0$  in the numerator on the r.h.s.

**Page 483** Equations (10.80) and (10.81):  $=$  should be  $\simeq$ .

**Page 483** Second line after Equation (10.80): A full stop (‘.’) should be inserted after  $j \neq k$ .

**Page 484** Line 1: The reference to Figure 10.2 should refer to Figure 10.3.

**Page 487** Figure 10.8: Label  $\phi_n$  should be  $\phi_n$ .

**Page 489** Line -7:  $p(\mathbf{t} | M)$  should be  $\ln p(\mathbf{t} | M)$ .

**Page 491** Equation (10.118):  $\boldsymbol{\eta}^T \boldsymbol{\chi}_0$  should be  $\nu_0 \boldsymbol{\eta}^T \boldsymbol{\chi}_0$ .

**Page 491** Equation (10.119):  $\boldsymbol{\eta}^T \boldsymbol{\chi}_N$  should be  $\nu_N \boldsymbol{\eta}^T \boldsymbol{\chi}_N$ .

**Page 491** Equation (10.121): “ $\chi_N = \chi_0 \dots$ ” should be “ $\nu_N \chi_N = \nu_0 \chi_0 \dots$ ”.

**Page 492** Line 7 after Equation (10.124):  $q^*(\mathbf{x}_j)$  should be  $q_j^*(\mathbf{x}_j)$ .

**Page 495** Last line before Equation (10.132): An opening quote (‘) is missing before max’.

**Page 501** Equation (10.160):  $\ln h(\mathbf{w}, \boldsymbol{\xi})p(\mathbf{w})$  should be  $\ln \{h(\mathbf{w}, \boldsymbol{\xi})p(\mathbf{w})\}$  on the r.h.s.

**Page 505** Line 4: (10.159) should be (10.160).

**Page 509** Line 1:  $q^{\setminus i}(\boldsymbol{\theta})$  should be  $q^{\setminus j}(\boldsymbol{\theta})$

**Page 515** Figure 10.18:  $\tilde{f}$  should be  $\tilde{f}$  in all factor labels in the right graph.

**Page 515** Equation (10.228): = should be  $\propto$ .

**Page 520** Exercise 10.27, line 2: “, defined by (10.107),” should be omitted.

**Page 521** Exercise 10.30, line 3: “second order” should be “first order”.

**Page 522** Equation (10.245): A term  $v^{\setminus n} D$  should be added to the r.h.s.

**Page 526** Equation (11.5):

$$\frac{dx}{dy} \quad \text{should be} \quad \frac{dx}{dy}.$$

**Page 529** Figure 11.4:  $\tilde{p}$  should be  $\tilde{p}$ .

**Page 531** Equation (11.17): This equation and end of sentence need to modified as follows:

$$q(z) = k_i \lambda_i \exp \{-\lambda_i (z - z_i)\} \quad \widehat{z}_{i-1,i} < z \leq \widehat{z}_{i,i+1}$$

where  $\widehat{z}_{i-1,i}$  is the point of intersection of the tangent lines at  $z_{i-1}$  and  $z_i$ ,  $\lambda_i$  is the slope of the tangent at  $z_i$  and  $k_i$  accounts for the corresponding offset.

**Page 535** First line after Equation (11.25):  $I(\cdot)$  should be  $I(\cdot)$ .

**Page 539** First line after Equation (11.36): “ $z^{(1)} = 0$ ” should be “ $z^{(0)} = 0$ ” (superscript index changed).

**Page 541** Equation (11.43):

$$\sum_{\mathbf{z}_{n-1}} \quad \text{should be} \quad \sum_{\mathbf{z}_{K-1}}.$$

**Page 541** Last line before Equation (11.44):  $z_\tau$  should be  $z^{(\tau)}$ .

**Page 541** Equation (11.45): This equation should read

$$\begin{aligned} p(\mathbf{z})q_k(\mathbf{z}'|\mathbf{z})A_k(\mathbf{z}', \mathbf{z}) &= \min(p(\mathbf{z})q_k(\mathbf{z}'|\mathbf{z}), p(\mathbf{z}')q_k(\mathbf{z}|\mathbf{z}')) \\ &= \min(p(\mathbf{z}')q_k(\mathbf{z}|\mathbf{z}'), p(\mathbf{z})q_k(\mathbf{z}'|\mathbf{z})) \\ &= p(\mathbf{z}')q_k(\mathbf{z}|\mathbf{z}')A_k(\mathbf{z}, \mathbf{z}') \end{aligned}$$

## 14 Corrections for pages 545–586

- Page 545** Equation (11.50):  $\alpha_i^2$  should be  $\alpha^2$  in the last term on the r.h.s.
- Page 546** Figure 11.13: Both instances of  $\tilde{p}$  should be  $\tilde{p}$ .
- Page 550** Equation (11.62), second line: + and – should be swapped.
- Page 554** Equation (11.72): A factor of  $1/L$  is missing on the last line.
- Page 555** Equation (11.73): A factor of  $1/L$  is missing on the r.h.s.
- Page 564** First line after Equation (12.12):  $b_i$  should be  $b_j$ .
- Page 565** Line –1: Before the period (‘.’) of the sentence ending “digits data set”, insert “, restricting our attention to images of the digit three”.
- Page 566** Figure 12.3, caption.: On the first line, before “off-line”, insert “digit three from the”. At the end of the caption, add the sentence: Blue corresponds to positive values, white is zero and green corresponds to negative values.
- Page 566** Line 2: “first five” should be “first four”.
- Page 566** Figure 12.4, caption, line 1: Before “off-line”, insert “digit three from the”.
- Page 567** First paragraph, last sentence: This sentence should read: Examples of reconstructions of a sample from the digit three data set are shown in figure 12.5.
- Page 567** First line after Equation (12.22): “ $\sigma_i$  is the variance” should be “ $\sigma_i$  is the standard deviation”.
- Page 572** Figure 12.9: All instances of  $\hat{z}$  should be  $\hat{z}$ .
- Page 575** Second paragraph, line 1: “ $M \times M$ ” should be “ $M$ -dimensional”.
- Page 575** Second paragraph, line 6–7: “variance parameter  $\lambda_i - \sigma^2$ ” should be “square root of the variance parameter  $\sqrt{\lambda_i - \sigma^2}$ ”.
- Page 577** Line –3: “distribution of the latent distribution” should be “distribution of the latent variable”.
- Page 578** Equation (12.53): A term  $M/2 \ln(2\pi)$  should be added to the summand (i.e. inside the braces) on the r.h.s.
- Page 578** Second paragraph: “M-step” should be “M step”.
- Page 579** Second line before Equation (12.58): “ $D \times M$  whose  $n^{\text{th}}$  row” should be “ $M \times N$  whose  $n^{\text{th}}$  column”.
- Page 579** Equation (12.58): On the r.h.s.  $\tilde{\mathbf{X}}$  should be  $\tilde{\mathbf{X}}^T$ .
- Page 582** Third paragraph, line 1: “log marginal” should be just “marginal”.
- Page 586** Equations (12.69)–(12.70): On the l.h.s.  $\mathbf{W}^{\text{new}}$  and  $\Psi^{\text{new}}$  should be  $\mathbf{W}_{\text{new}}$  and  $\Psi_{\text{new}}$ , respectively.

**Page 588** Equation (12.78): The upper limit of the second summation on the l.h.s. should be  $N$  and not  $m$ .

**Page 588** First line after Equation (12.79):  $a_{ni}$  should be  $a_{in}$ .

**Page 592** Equation (12.90): The numerator in the rightmost expression should be 2.

**Page 593** First line after Equation (12.91): “activations” should be “activation”.

**Page 607** Equation (13.1): The r.h.s. should read

$$p(\mathbf{x}_1) \prod_{n=2}^N p(\mathbf{x}_n | \mathbf{x}_1, \dots, \mathbf{x}_{n-1}).$$

**Page 609** First paragraph, line –3:  $K^{M-1}(K-1)$  should be  $K^M(K-1)$ .

**Page 616** Equation (13.15): The summation should run over  $\mathbf{z}_n$  in the rightmost expression.

**Page 616** Equation (13.16): The the rightmost expression should read

$$\sum_{\mathbf{z}_{n-1}, \mathbf{z}_n} \xi(\mathbf{z}_{n-1}, \mathbf{z}_n) z_{n-1,j} z_{nk}.$$

**Page 619** Second line after Equation (13.31): “in the first of these results” should be “in the second of these results”.

**Page 620** First paragraph, line –3: “represent set” should be “represent a set”.

**Page 624** Second paragraph, last sentence: This sentence, starting “Since the observed variables ...”, should be omitted.

**Page 624** Fourth paragraph, line 4: The reference to (13.29) should be referring to (13.30).

**Page 625–626** Figures 13.14 and 13.15: All  $x$  and  $z$  node labels should be bold (i.e.  $\mathbf{x}$  and  $\mathbf{z}$ ).

**Page 628** Equation (13.65): On the r.h.s.,  $c_n$  should be  $c_n^{-1}$ .

**Page 629** Line 2: “based” should be “based on”.

**Page 630** Equation (13.70): This should read

$$\omega(\mathbf{z}_n) = \max_{\mathbf{z}_1, \dots, \mathbf{z}_{n-1}} \ln p(\mathbf{x}_1, \dots, \mathbf{x}_n, \mathbf{z}_1, \dots, \mathbf{z}_n)$$

(missing  $\ln$  inserted).

**Page 630** Equation (13.71): This should read

$$k_{n-1}^{\max} = \psi(k_n^{\max}).$$

**Page 632** Line –2: “excessive the number” should be just “excessive number”.

## 16 Corrections for pages 636–675

**Page 636** Fourth paragraph, lines 3–4:  $\mathbf{z}_n$  and  $\mathbf{x}_n$  should be  $\mathbf{z}_{n-1}$  and  $\mathbf{x}_{n-1}$ , respectively.

**Page 637–643** Equation (13.77)–Equation (13.110): All instances of  $\mathbf{V}_0$  should be replaced by  $\mathbf{P}_0$ , in equations as well as in the text.

**Page 641** Equation (13.100): On the r.h.s., change  $\boldsymbol{\mu}_N$  to  $\boldsymbol{\mu}_n$ .

**Page 641** Equation (13.103), first line: On the r.h.s., change  $\mathbf{z}_{-1}$  to  $\mathbf{z}_{n-1}$ .

**Page 641** Line –2: This line should read: “Gaussian with mean given by  $[\widehat{\boldsymbol{\mu}}_{n-1}, \widehat{\boldsymbol{\mu}}_n]^\top$  and a covariance”.

**Page 641** Equation (13.104): The order of  $\mathbf{z}_n$  and  $\mathbf{z}_{n-1}$  should be swapped on the l.h.s.

**Page 642** Equation (13.106):  $\mathbf{J}_{n-1} \widehat{\mathbf{V}}_n$  should be  $\widehat{\mathbf{V}}_n \mathbf{J}_{n-1}^\top$  on the r.h.s..

**Page 643** Equation (13.114): The first instance of  $\mathbf{A}^{\text{new}}$  on the second line of the equation should be transposed.

**Page 643** Equation (13.116): The first and third instances of  $\mathbf{C}^{\text{new}}$  on the second line of the equation should be transposed.

**Page 645** Paragraph –2, line –4:  $0 \leq w_n^{(l)} 1$  should be  $0 \leq w_n^{(l)} \leq 1$ .

**Page 646** Equation (13.119), last line: = should be  $\simeq$ .

**Page 651** Lines 2–3: Sentence fragment starting “...in which  $\mathbf{C}$ ” should be changed to “in which  $\mathbf{C} = 1$ ,  $\mathbf{A} = 1$  and  $\boldsymbol{\Gamma} = 0$ .”

**Page 651** Exercise 13.26, last line: Insert “, assuming  $\boldsymbol{\mu} = \mathbf{0}$  in (12.42)” before the end of the sentence.

**Page 651** Line 3:  $\mathbf{m}_0$  should be  $\boldsymbol{\mu}_0$ .

**Page 651** Exercises 13.25, 13.28 and 13.32: All instances of  $\mathbf{V}_0$  should be replaced by  $\mathbf{P}_0$ .

**Page 651** Exercise 13.28, line 3: Insert “ $\mathbf{C} = \mathbf{I}$  and that” before  $\mathbf{P}_0$ .

**Page 659** Second paragraph after AdaBoost algorithm, line –2: “parallel” should be “perpendicular”.

**Page 663** Paragraph 3, line 1: “Figure 14.5 shows” should be “Figures 14.5 and 14.6 show”.

**Page 666** Line 3: Insert “negative” before “cross-entropy”.

**Page 675** Exercise 14.11: The text of this exercise should be changed to “Consider a data set comprising 400 data points from class  $\mathcal{C}_1$  and 400 data points from class  $\mathcal{C}_2$ . Suppose that a tree model A splits these into (300, 100) assigned to the first leaf node (predicting  $\mathcal{C}_1$ ) and (100, 300) assigned to the second leaf node (predicting  $\mathcal{C}_2$ ), where  $(n, m)$  denotes that  $n$  points come from class  $\mathcal{C}_1$  and  $m$  points come from class  $\mathcal{C}_2$ . Similarly, suppose that a second tree model B splits them into (200, 400) and (200, 0), respectively. Evaluate the misclassification rates for the two trees and

hence show that they are equal. Similarly, evaluate the pruning criterion (14.31) for the cross-entropy case (14.32) and the Gini index case (14.33) for the two trees and show that they are both lower for tree B than for tree A.”.

- Page 690** Equation (B.57): The r.h.s. should read  $-\mu_j \mu_k, j \neq k$ .
- Page 691** Equation (B.62): The r.h.s. should read  $-N \mu_j \mu_k, j \neq k$ .
- Page 698** First line after Equation (C.28): (C.26) should be (C.24).
- Page 711** Column 2, entry 2: “S. I. Amari” should be “S. Amari”.
- Page 714** Column 1, entry 2: “J. M. B.” should be “J. M. Bernardo”.
- Page 714** Column 1, entry -4: “Tao” should be “Tiao”.
- Page 719** Column 1, entry 7: “Jeffries” should be “Jeffreys”.