

Pattern Recognition and Machine Learning

Errata

Markus Svensén and Christopher M. Bishop

All rights retained. Not to be distributed without permission.
© Markus Svensén and Christopher M. Bishop and (2007)

October 5, 2007

Preface

This document lists corrections for *Pattern Recognition and Machine Learning* by Christopher M. Bishop, published by Springer in 2006. It is intended to be complete, in that it includes even trivial typographical errors. 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 are given in the order they should appear in the book. Each correction starts with a page number in the margin, followed (in the main body of the page) by the location of the mistake and the required correction. 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).

Acknowledgements

We would like to thank all of the readers who have reported mistakes in PRML. In particular, we are grateful to Dr Toshihiro Kamishima of the National Institute of Advance Industrial Science and Technology, Japan, and his colleagues, and also

4 PREFACE

to Dr Xiaobo Jin of the Chinese Academy of Sciences, for particularly thorough feedback.

Notes on the corrections

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

The PRML Companion

Originally it had been intended to write a ‘companion’ textbook to PRML in collaboration with Ian Nabney, accompanied by Matlab software (version 3 of “Netlab”) implementing many of the algorithms discussed in PRML. The companion book was to have served two main purposes: (i) to explain some useful algorithms, mainly concerned with the solution of optimization problems which arise in machine learning, for which there was insufficient space in the main text book, and (ii) to provide an overview and manual for the Netlab v3 software along with guidance on the practical application of machine learning. It has been decided instead to make most of this material, including the software, freely available from the web. In particular the optimization algorithms will be discussed in a tutorial paper, *Optimization Algorithms for Machine Learning*, which will be available as a PDF file from the PRML book web site. The references to Bishop and Nabney (2008) in the book, most which have been kept, now refer to this tutorial paper. The Netlab v3 software, along with supporting documentation, demonstrations and tutorials, will be available directly from the Netlab web site: <http://www.ncrg.aston.ac.uk/netlab/>.

Bayesian “Estimate” of the Variance of a Gaussian

When we estimate the mean, μ , and the variance, σ^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 μ and τ (the inverse variance, $1/\sigma^2$) of the form,

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

We can then integrate over μ 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 σ^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. Indeed, if we chose a different but equally valid prior,

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

the value of $\mathbb{E}[\tau]^{-1}$ will equal the biased maximum likelihood estimate for σ^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 λ to η , while λ has been kept throughout 10.6.

Corrections

- Page viii** Third paragraph: The last sentence, starting “A companion volume . . .”, should be replaced with “Matlab software implementing many of the algorithms discussed in this book, together with example data sets, will be available through the book web site, along with a companion tutorial (Bishop and Nabney, 2008) describing practical algorithms for solving the optimization problems which arise in machine learning.”
- Page xi** Second paragraph, line -3: “roman” should be “Roman”.
- Page xi** Fourth paragraph, middle line: “about it dimensionality” should read “about its dimensionality”.
- Page 8** Table 1.1, column labels: $M = 6$ should be $M = 3$.
- Page 18** Fifth line after Equation (1.26): “suffices” should be “suffixes”.
- Page 19** Second paragraph, first line: “x” should be “ x ”.
- Page 28** First sentence: This sentence should be omitted.
- Page 31** Equation (1.72): $\phi(x_n)\phi(x)^T$ should be $\phi(x_n)\phi(x_n)^T$, inserting the missing index n .
- Page 44** Caption of Figure 1.27, last line: Insert “, assuming the prior class probabilities, $p(C_1)$ and $p(C_2)$, are equal” before the full stop.
- Page 47** Equation (1.90): The integrand of the second integral should be replaced by
- $$\text{var}[t|\mathbf{x}]p(\mathbf{x}).$$
- 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 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 57** Equation (1.119): The r.h.s. should be multiplied by a factor $1/N$.
- Page 61** Exercise 1.16, line 4: $M6th$ should be M^{th} .
- Page 61** Exercise 1.16, Equation (1.139), l.h.s.: $N(d, M)$ should be $N(D, M)$.
- Page 65** Exercise 1.34, line 2: “functional (1.108)” should be “functional preceding (1.108)”.
- Page 71** Section 2.1.1, first paragraph, line 8: “the form of the product” should be “the form of a product” (‘a’ replacing ‘the’).
- Page 75** Equation (2.28): μ_M should be μ_K in the third expression.

8 Corrections for pages 81–99

Page 81 Caption of Figure 2.7, last sentence: Omit the word “major”.

Page 87 Line -2: “. . . is independent of \mathbf{x}_a .” should be “. . . is independent of \mathbf{x}_b .”

Page 89 Equation (2.87): The last line should read

$$+\mathbf{x}_a^T(\mathbf{\Lambda}_{aa} - \mathbf{\Lambda}_{ab}\mathbf{\Lambda}_{bb}^{-1}\mathbf{\Lambda}_{ba})\boldsymbol{\mu}_a + \text{const}$$

(incorrect inverse ($^{-1}$) removed).

Page 89 Last line before Equation (2.88): Omit the “of” before $p(\mathbf{x}_a)$.

Page 89 First line after Equation (2.89): Omit the “in” before (2.88).

Page 96 Equations (2.129): ‘+’ should be changed to ‘−’.

Page 96 Second paragraph, line 3: Insert ‘negative’ before ‘log likelihood’.

Page 96 Equation (2.133): \mathbf{x}_n should be x_n .

Page 96 Equation (2.133): A minus sign (‘−’) should be added to the l.h.s.

Page 96 Equation (2.134): Minus signs (‘−’) should be added to both sides; the correct form is

$$-\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \frac{\partial}{\partial \theta} \ln p(x_n | \theta) = \mathbb{E}_x \left[-\frac{\partial}{\partial \theta} \ln p(x | \theta) \right].$$

Page 96 Equation (2.135): For consistency, the r.h.s. should be rewritten as

$$\theta^{(N)} = \theta^{(N-1)} - a_{N-1} \frac{\partial}{\partial \theta^{(N-1)}} \left[-\ln p(x_N | \theta^{(N-1)}) \right].$$

Page 97 Figure and caption 2.11: The labels μ and μ_{ML} should be exchanged in the figure. The caption should be changed to: “In the case of a Gaussian distribution, with θ corresponding to μ_{ML} , the regression function illustrated in Figure 2.10 takes the form of a straight line, as shown in red. In this case, the random variable z corresponds to the derivative of the negative log likelihood function and is given by $-(x - \mu_{\text{ML}})/\sigma^2$, and its expectation that defines the regression function is a straight line given by $-(\mu - \mu_{\text{ML}})/\sigma^2$. The root of the regression function corresponds to the true mean μ .”

Page 97 Equation (2.136): ‘−’ signs should prefix both the middle and rightmost expressions and on the immediately following line, $\mu - \mu_{\text{ML}}$ should be replaced by $-(\mu - \mu_{\text{ML}})/\sigma^2$.

Page 97–101 Section 2.3.6: \mathbf{X} should be replaced by \mathbf{x} throughout this section.

Page 99 Second paragraph: All instances of \mathbf{x} (with indices n and N) and $\boldsymbol{\mu}$ should be replaced by x and μ , respectively, to indicate univariate data. Moreover, in equation (2.144), D should be replaced by \mathbf{x} , in addition to the changes of \mathbf{x} and $\boldsymbol{\mu}$.

- Page 103** Last paragraph, line 4: There should be a closing paranthesis (‘)’) after 2.3.9, before the full stop.
- Page 106** Equation (2.168): For additional clarity, prefix the left and right expressions with $\bar{x}_1 =$ and $\bar{x}_2 =$, respectively.
- Page 108** Second line before Equation(2.180): “zeroth-order Bessel function” should be “zeroth-order modified Bessel function” .
- Page 109** Equation(2.185): $A(m)$ on the l.h.s. should be $A(m_{\text{ML}})$.
- Page 109** Equation(2.187): ‘−’ should be ‘+’ on the r.h.s..
- Page 111** First sentence after Equation (2.189): This sentence should be changed to: “Also, given that $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \geq 0$, a sufficient condition for the requirement $p(\mathbf{x}) \geq 0$ is that $\pi_k \geq 0$ for all k .”
- Page 114** First line after Equation (2.204): “ $\mathbf{x} = (x_1, \dots, x_N)^{\text{T}}$ ” should be “ $\mathbf{x} = (x_1, \dots, x_M)^{\text{T}}$ ”.
- Page 116** First line after Equation (2.225): Omit the phrase “where we have used (2.194)” and add a full stop after the equation.
- Page 122** Equation (2.243): $(1 - P)^{1-K}$ should be $(1 - P)^{N-K}$.
- 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* $1/2$).
- Page 131** Exercise 2.23, first line: Reference to (2.45) should refer to (2.48).
- Page 133** Equation (2.291): $\mathbb{E}[\mathbf{x}_n \mathbf{x}_m]$ should be $\mathbb{E}[\mathbf{x}_n \mathbf{x}_m^{\text{T}}]$.
- Page 139** Third paragraph, line -2: “fit a different” should be “fitting a different”.
- Page 141** Equation (3.13): The r.h.s. should be multiplied by a factor β .
- Page 144** Equation (3.26): E should be E_D on the l.h.s..
- Page 145** Third paragraph, line 1: “is know as” should be “is known as”.
- Page 145** Third paragraph, line -2: “shows that” should be “shows”.
- Page 152** Third paragraph, line -1: “question model” should be “question of model”.
- Page 156** Line -2 before Equation (3.58): Reference to Section 8.1.4 should refer to Section 2.3.3.
- Page 160** Fourth paragraph, line 1: “effective kernel” should be “equivalent kernel”.
- Page 165** Section 3.5, first paragraph, line 4: Change “over either” to “either over”.

10 Corrections for pages 166–252

Page 166 Second paragraph, line 6: Change “discussed (Section 4.4)” to “discussed in Section 4.4,”.

Page 168–169 Equation (3.89): The full stop after the equation should be removed and the phrase “where we neglect derivatives of $\mathbf{m}_N^T \mathbf{m}_N$ with respect to α .” inserted on the following line (on the following page).

Page 171 First sentence after Equation (3.97): This sentence should be omitted.

Page 173 Exercise 3.1, Equation (3.102):

$$\tanh\left(\frac{x - \mu_j}{s}\right) \text{ should be } \tanh\left(\frac{x - \mu_j}{2s}\right).$$

Page 174 Exercise 3.4: $y(x, \mathbf{w})$ and $y(x_n, \mathbf{w})$ should be $y(\mathbf{x}, \mathbf{w})$ and $y(\mathbf{x}_n, \mathbf{w})$ in Equations (3.105) and (3.106), respectively.

Page 177 Exercises 3.20 and 3.22: Change “Starting from (3.86) verify ..” to “Verify ..”.

Page 185 Line -3: “these point” should be “these points”.

Page 194 Line 1: “where \mathcal{M} ” should be expanded to “where $\phi_n = \phi(\mathbf{x}_n)$ and \mathcal{M} ”.

Page 194 Line 3 after Equation 4.55: “without of” should be “without loss of”.

Page 198 Equation (4.63): A pair of parentheses are missing on the r.h.s.; the correct form is

$$a_k = \ln(p(\mathbf{x}|\mathcal{C}_k)p(\mathcal{C}_k))$$

Page 200 Equation(4.71): $p(\mathbf{t}|\boldsymbol{\pi}, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma})$ should be $p(\mathbf{t}, \mathbf{X}|\boldsymbol{\pi}, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma})$.

Page 210 Equation (4.110): The leading minus (‘-’) sign on the r.h.s. should be removed.

Page 221 Exercise 4.4: (4.23) should be (4.22).

Page 223 Exercise 4.18, line 1: (4.91) should be (4.106).

Page 223 Exercise 4.23, following the equation: Should say “where \mathbf{H} is . . . the negative log likelihood” (insert ‘negative’).

Page 245 Last line before Equation (5.58): “logistic sigmoid” should be “sigmoidal”.

Page 248 Equations (5.75) and (5.76): To conform to indexing in preceding equations, index j should be replaced by l .

Page 251 Equation (5.83): In the first term on the r.h.s., $\nabla y_n \nabla y_n$ should be $\nabla y_n (\nabla y_n)^T$.

Page 251 First line after Equation (5.84): $\mathbf{b}_n = \nabla y_n = \nabla a_n$ should be $\mathbf{b}_n \equiv \nabla a_n = \nabla y_n$.

Page 252 First line after Equation (5.88): The text fragment “where \mathbf{I} is the unit matrix,” should be removed.

Page 254 Equation (5.95): On the r.h.s., $H_{kk'}$ should be $M_{kk'}$. Moreover, the indices j and j' should be swapped on the r.h.s.

Page 263 Figure 5.14: Figure improved (small panels enlarged).

Page 271 Equation (5.144) and the immediately preceding and following lines: η_j should be replaced by ξ_j .

Page 272 Equation (5.147): The r.h.s. should be multiplied by λ .

Page 274 Second paragraph: With the exception of the K on line 4, all instances of K should be replaced by L and vice versa.

Page 275 Equation (5.153): K should replace k as the upper limit of the inner summation on the r.h.s.

Page 275 Equation (5.154): The l.h.s. should be replaced with $\gamma_{nk} = \gamma_k(\mathbf{t}_n | \mathbf{x}_n)$.

Page 275 Equations (5.155)–(5.156): γ_k should be replaced by γ_{nk} . Moreover in (5.156), t_l should be t_{nl} .

Page 275 Equation (5.157): This equation should read

$$\frac{\partial E_n}{\partial a_k^\sigma} = \gamma_{nk} \left\{ L - \frac{\|\mathbf{t}_n - \boldsymbol{\mu}_k\|^2}{\sigma_k^2} \right\}.$$

Page 282 Equation (5.181): On the r.h.s.

$$\sum_n = 1^N \quad \text{should be} \quad \sum_{n=1}^N.$$

Page 284 Exercise 5.1, Line 2: $g(\cdot)$ should be $h(\cdot)$.

Page 289 Exercise 5.32, last line: The constraint equation should read: $\sum_k \gamma_k(w_i) = 1$ for all i .

Page 290 Exercise 5.41, first line: “Section 5.7.1” should be “Sections 5.7.1 and 5.7.2”.

Page 293 Sentence fragment preceding Equation (6.8): This should be changed to “Using (6.3) to eliminate \mathbf{w} from (6.4) and solving for \mathbf{a} we obtain”.

Page 295 Figure 6.1: The figure and caption should be replaced by

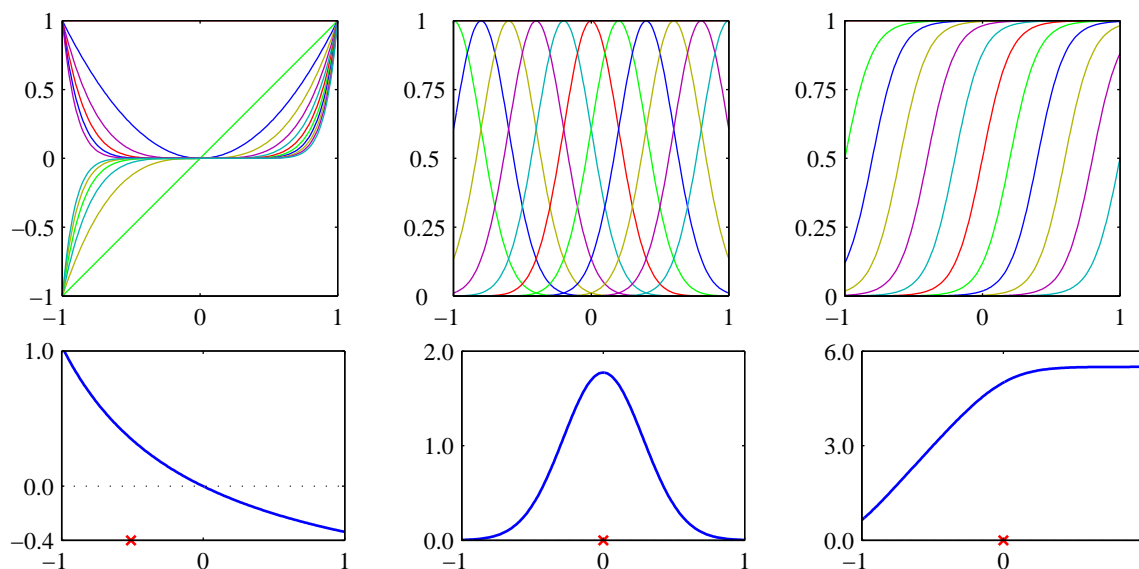


Figure 6.1 Illustration of the construction of kernel functions starting from a corresponding set of basis functions. In each column the lower plot shows the kernel function $k(x, x')$ defined by (6.10) plotted as a function of x , where x' is given by the red cross (\times), while the upper plot shows the corresponding basis functions given by polynomials (left column), ‘Gaussians’ (centre column), and logistic sigmoids (right column).

Page 297 Lines 1–2 after Equation (6.27): It says that $|A|$ denotes the number of subsets in A , it should say: “ $|A|$ denotes the number of elements in A ”.

Page 300 Line 1 after Equation (6.39): $f(\mathbf{x})$ should be $y(\mathbf{x})$.

Page 300 Equation (6.40): On the l.h.s., $y(\mathbf{x}_n)$ should be $y(\mathbf{x})$.

Page 310 Second paragraph, last sentence: This sentence should be omitted.

Page 321 Exercise 6.16: After (6.98), add “where $\mathbf{w}_\perp^\top \phi(\mathbf{x}_n) = 0$ for all n ”.

Page 322 Exercise 6.23, line -2: $\mathbf{x}_1, \dots, \mathbf{x}_{N+1}$ should be $\mathbf{x}_1, \dots, \mathbf{x}_N$.

Page 329 Biography for Lagrange, first column: A “to” is missing: “important contributions to mathematics”.

Page 358 Exercise 7.19, line 1: “approximate log marginal” should be “approximate marginal”.

Page 367 Line -6: The comma before “Similarly” should be replaced by a full stop.

Page 376 Line between the second and third equation: $p(a)p(b)$ should be $p(a|c)p(b|c)$.

Page 378 Line -4: “it has a descendant c because is in the conditioning set” should read “it has a descendant c in the conditioning set”.

Page 379 Second Paragraph, Line 2: “was” should be “way”.

Page 400 Caption of Figure 8.41, Line -1: $f_b(x_1, x_2)$ should be $f_b(x_2, x_3)$.

Page 404 First line after Equation (8.65): f_x should be f_s .

Page 404 First line of Equation(8.66): Last summation sign inside brackets

$$\sum_{X_{xm}} \text{ should be } \sum_{X_{sm}}.$$

Page 410 Equation (8.86), line -2: The middle summation symbol

$$\sum_{x_2} \text{ should be } \sum_{x_3}.$$

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 421 Exercise 8.21, line 2: $f_x(\mathbf{x}_s)$ should be $f_s(\mathbf{x}_s)$.

Page 435 Equation (9.16): There are a matrix inverse missing and an extra minus (‘-’) sign on the r.h.s.; the correct form is

$$0 = \sum_{n=1}^N \frac{\pi_k \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\underbrace{\sum_j \pi_j \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}_{\gamma(z_{nk})}} \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_n - \boldsymbol{\mu}_k).$$

Page 435 Line 3 after Equation (9.16): $\boldsymbol{\Sigma}_k^{-1}$ should be $\boldsymbol{\Sigma}_k$, i.e., no inverse.

Page 444 Equation (9.41): $M/2$ should replace $1/2$ in the denominator of the normalisation constant on the r.h.s.

Page 458 Exercise 9.23, last sentence.: Should read: Show that, at any stationary point, these two sets of re-estimation equations are formally equivalent.

Page 463 Line 1: “We can the introduce” should be “We can then introduce”.

Page 470 Equation (10.20): The integrand on the r.h.s. should be squared, i.e.,

$$D_H(p||q) = \int (p(x)^{1/2} - q(x)^{1/2})^2 dx.$$

14 Corrections for pages 471–496

Page 471 Equation (10.28), second line: An additional $1/2 \ln \tau$ term (arising from the Gaussian-Gamma prior over μ) should be added.

Page 471 Equation (10.29):

$$\frac{N}{2} \text{ should be } \frac{N+1}{2}$$

Page 472 Equation (10.31): This equation should be replaced by

$$\frac{1}{\mathbb{E}[\tau]} = \mathbb{E} \left[\frac{1}{N+1} \sum_{n=1}^N (x_n - \mu)^2 \right] = \frac{N}{N+1} (\overline{x^2} - 2\overline{x}\mathbb{E}[\mu] + \mathbb{E}[\mu^2]) .$$

Page 473 Equation (10.33): This equation should be replaced by

$$\frac{1}{\mathbb{E}[\tau]} = (\overline{x^2} - \bar{x}^2) = \frac{1}{N} \sum_{n=1}^N (x_n - \bar{x})^2 .$$

Furthermore, the sentence immediately following this equation should be replaced by “For a comprehensive treatment of Bayesian inference for the Gaussian distribution, including a discussion of the advantages compared to maximum likelihood, see Minka (1998).”

Page 473 Margin reference to Section 1.2.4: This reference should be omitted.

Page 473 Equations (10.34) and (10.35) and the line in between: \mathcal{L}_m should be \mathcal{L} .

Page 473 Equation(10.36): Remove the full stop after the equation and on the following line insert “where

$$\mathcal{L}_m = \sum_{\mathbf{Z}} q(\mathbf{Z}|m) \ln \left\{ \frac{p(\mathbf{Z}, \mathbf{X}|m)}{q(\mathbf{Z}|m)} \right\} .$$

Page 473 Line -3 and -4: \mathcal{L}_m should be \mathcal{L} .

Page 474 Line 1: Directly after the reference to equation (10.35), insert: “, or equivalently by optimization of \mathcal{L}_m ”.

Page 478 Equation(10.63): The l.h.s. should say $\nu_0 + N_k + 1$.

Page 489 Equation (10.110): The last term, $\ln \Gamma(a_N)$, should be $\ln \Gamma(a_0)$.

Page 490 Equation (10.114): χ_0 should replace \mathbf{v}_0 on the l.h.s.

Page 493–496 Section 10.5: In text, figures and figure captions, from Figure 10.10 up to and including the last line preceding Equation (10.141), λ should be replaced by η .

Page 496 Equation (10.141): This equation should be changed to

$$\eta = -\frac{1}{4\xi} \tanh \left(\frac{\xi}{2} \right) = -\frac{1}{2\xi} \left[\sigma(\xi) - \frac{1}{2} \right] = -\lambda(\xi)$$

and the following text inserted immediately following the equation: “where we have defined $\lambda = -\eta$ to maintain consistency with Jaakkola and Jordan (2000).”

Page 497 Equations (10.142) and (10.143): These equations should be changed to

$$g(\lambda(\xi)) = -\lambda(\xi)\xi^2 - f(\xi) = -\lambda(\xi)\xi^2 + \ln(e^{\xi/2} + e^{-\xi/2}).$$

and

$$f(x) \geq -\lambda(\xi)x^2 - g(\lambda(\xi)) = -\lambda(\xi)x^2 + \lambda(\xi)\xi^2 - \ln(e^{\xi/2} + e^{-\xi/2})$$

respectively.

Page 502 Equation 10.164: There are a number of sign errors in this equation; the correct form is

$$\begin{aligned} \mathcal{L}(\xi) = & \frac{1}{2} \ln \frac{|\mathbf{S}_N|}{|\mathbf{S}_0|} + \frac{1}{2} \mathbf{m}_N^T \mathbf{S}_N^{-1} \mathbf{m}_N - \frac{1}{2} \mathbf{m}_0^T \mathbf{S}_0^{-1} \mathbf{m}_0 \\ & + \sum_{n=1}^N \left\{ \ln \sigma(\xi_n) - \frac{1}{2} \xi_n + \lambda(\xi_n) \xi_n^2 \right\}. \end{aligned}$$

Page 504 Equation (10.177): a_N and b_N should replace all instances of a_0 and b_0 , respectively, on the r.h.s..

Page 505 Equation (10.183): The transposition Ts are in the wrong places; the correct form is

$$\mathbb{E} [\mathbf{w}\mathbf{w}^T] = \mathbf{\Sigma}_N + \boldsymbol{\mu}_N \boldsymbol{\mu}_N^T.$$

Page 521 Exercise 10.29: All instances of λ should be replaced with η .

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

Page 526 Second paragraph: The second half of the sentence forming this paragraph, starting “, and some practical guidance . . .”, should be removed.

Page 528 Equations (11.10) and (11.11): $\ln z_1$ and $\ln z_2$ in the numerators should both be $\ln r^2$.

Page 529 Line -1: The final full stop (‘.’) should be removed.

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

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

Page 541 Equation (11.45): This equation should read

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

Page 544 Line 4: $p(z_i|\{z_{\setminus i}\})$ should be $p(z_i|z_{\setminus i})$ (erroneous ‘{’ removed).

Page 552–553 Equation(11.68) and (11.69): The sign of the argument to the exponential functions forming the second arguments to the min functions need to be changed.

Page 560 Line -2: “w” should be “we”.

Page 567 First line after Equation (12.22): “ σ_i is the variance” should be “ σ_i is the standard deviation”.

Page 573 Equation (12.40): $\sigma^{-1}\mathbf{I}$ on the r.h.s. should be $\sigma^{-2}\mathbf{I}$.

Page 573 Equation (12.42): The covariance on the r.h.s. should be $\sigma^2\mathbf{M}^{-1}$.

Page 575 Second paragraph, line 1: “ $M \times M$ ” should be “ M -dimensional”.

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 579 Second line before Equation (12.58): “ $D \times M$ whose n^{th} row” should be “ $M \times N$ whose n^{th} column”.

Page 579 Equation (12.58): On the r.h.s. $\tilde{\mathbf{X}}$ should be $\tilde{\mathbf{X}}^T$.

Page 581 Caption of Figure 12.12, last line: The caption for the last panel, (f), should read “The converged solution”.

Page 584 Caption of Figure 12.14, line -3: “left-hand” should be “right-hand”.

Page 586 Equations (12.69) and (12.70): On the l.h.s. \mathbf{W}^{new} and Ψ^{new} should be \mathbf{W}_{new} and Ψ_{new} , respectively.

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

Page 603 Exercise 12.29, line 4: The sentence starting “Now . . .” should be replaced by: “Now consider two variables y_1 and y_2 where y_1 is symmetrically distributed around 0 and $y_2 = y_1^2$.”

Page 611 Equation (13.7): On the l.h.s., change $\mathbf{z}_{n-1, \mathbf{A}}$ to $\mathbf{z}_{n-1}, \mathbf{A}$.

Page 612 First line after Equation (13.9): Change “focuss” to “focus”.

Page 621 Caption of Figure 13.12, line 3: Change $\alpha(z_{n1})$ to $\alpha(z_{n,1})$.

Page 622 Caption of Figure 13.13, line 4: Change $\beta(z_{n1})$ to $\beta(z_{n,1})$.

Page 628 Equation (13.65): On the r.h.s., change \mathbf{z}_{-1} to \mathbf{z}_{n-1} .

Page 630 Equation (13.68), r.h.s.: Change $\ln p(\mathbf{x}_{+1}|\mathbf{z}_n)$ to $\ln p(\mathbf{z}_{n+1}|\mathbf{z}_n)$.

Page 632 Equation (13.74): The minus (‘−’) sign in the argument of the exponential in the rightmost expression should be removed.

- Page 635** Section 13.3, first paragraph, line -4: $\mathbf{z}_1, \dots, \mathbf{x}_N$ should be $\mathbf{z}_1, \dots, \mathbf{z}_N$.
- Page 645** Line 1: “which do not have a linear-Gaussian,” should be “which are not linear-Gaussian,”.
- Page 650** Exercise 13.24, first line after Equation (13.128): Change “re-case” to “re-cast”.
- Page 651** Exercise 13.27, line 3: The sentence should start “Show that, in the case $\mathbf{C} = \mathbf{I}$, the posterior ...”.
- Page 659** First paragraph after AdaBoost algorithm, line 5: “decreased” should be “unchanged”.
- Page 674** Exercise 14.4, line 2: “hods” should be “holds”.
- Page 681** Paragraph 1, Line -4: “six distinct segments” should be “ten distinct segments”.
- Page 688** Equation (B.29): α should be a .
- Page 691** Equation (B.59): Two instances of M on the r.h.s. should both be replaced by K . The correct form is

$$\text{Mult}(m_1, m_2, \dots, m_K | \boldsymbol{\mu}, N) = \binom{N}{m_1 m_2 \dots m_K} \prod_{k=1}^K \mu_k^{m_k}.$$

- Page 698** Equation (C.20): The vector in all the denominators should be a scalar, i.e.,

$$\frac{\partial}{\partial x} (\mathbf{A}\mathbf{B}) = \frac{\partial \mathbf{A}}{\partial x} \mathbf{B} + \mathbf{A} \frac{\partial \mathbf{B}}{\partial x}.$$

- Page 698** First line after Equation (C.28): (C.26) should be (C.24).
- Page 701** Last paragraph, line 2: “all values” should be “all non-zero values”.
- Page 704** Line 2: $E[f]$ should be $F[y]$ and both instances of $f(x)$ should be $y(x)$.
- Page 704** Equation (D.4): δE should be δF in the numerator of the integrand.
- Page 713** Column 1, entry -2: This entry should be changed to: Bishop, C. M. and I. T. Nabney (2008). Optimization Algorithms for Machine Learning. In preparation.
- Page 714** Column 1, entry 2: “J. M. B.” should be “J. M. Bernardo”.
- Page 722** Column 1, entry -2: Before this entry, a new entry should be inserted: Minka, T. (1998) Inferring a Gaussian distribution. MIT Media Lab note. Available from <http://research.microsoft.com/~minka/>.
- Page 737** Index entry “Shur complement”: Should be “Schur complement”.