Pattern Recognition and Machine Learning
Errata and Additional Comments

Markus Svensén and Christopher M. Bishop

September 21, 2011

Preface

This document lists corrections and clarifications for the first printing[1] of *Pattern Recognition and Machine Learning* by Christopher M. Bishop, first 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

<div align="center">

`prml-fb@microsoft.com`

</div>

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 ammendment. 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' referes to the last paragraph started, but not necessarily completed on a page.

---

[1]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, if it reads "9 8 (corrected at 8th printing 2009)" you have a copy from the third printing.

- 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

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, 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,

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

If we assume a fully factorized posterior over μ and τ, 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 maxiumum 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 and posterior. If we make an equally valid choice, given by the Gaussian-Gamma prior,

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

and again assume a fully factorised posterior, the value of $\mathbb{E}[\tau]^{-1}$ will equal the biased maximum likelihood estimate for σ^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 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): $\phi(x_n)\phi(x)^{\mathrm{T}}$ should be $\phi(x_n)\phi(x_n)^{\mathrm{T}}$, inserting the missing index n.

Page 44 Caption of Figure 1.27, last line: Insert ", assuming the prior class probabilities, $p(\mathcal{C}_1)$ and $p(\mathcal{C}_2)$, are equal" before the full stop.

Page 47 Equation (1.90): The integrand of the second integral should be replaced by

$$\operatorname{var}[t|\mathbf{x}]\, p(\mathbf{x}).$$

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 49 Second paragraph, line –2: Both occurrences of $\ln$ should be replaced by $\log_2$.

Page 52 Equation (1.100): $\partial \widetilde{H}$ should be $\partial^2 \widetilde{H}$ 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 57 Equation (1.119): The r.h.s. should be multiplied by a factor $1/N$.

Page 57 Line –4: $I(\mathbf{x}, \mathbf{y})$ should be $\mathrm{I}[\mathbf{x}, \mathbf{y}]$.

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 63   Exercise 1.20, Equation (1.149):

$$-\frac{3\epsilon^2}{2\sigma^2} \qquad \text{should be} \qquad -\frac{\epsilon^2}{\sigma^2}$$

in the argument of the exponential on the r.h.s.

Page 65   Exercise 1.32, line -1: This line should read "by $H[\mathbf{y}] = H[\mathbf{x}] + \ln|\det(\mathbf{A})|$ where $|\det(\mathbf{A})|$ denotes the absolute value of the determinant of $\mathbf{A}$."

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.

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

Page 81   Second paragraph, Line 2: The reference to (2.51) should refer to (2.50).

Page 83   Line 2: $\boldsymbol{\mu}^{\mathrm{T}}\mathbf{z}$ should be $\mathbf{z}\boldsymbol{\mu}^{\mathrm{T}}$.

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

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

(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 90   Equation (2.96): $\mathbf{x}$ should be $\mathbf{x}_a$ on the r.h.s.

Page 91   Second line after Equation (2.102): The sentence should end "hence $p(\mathbf{z})$ is a Gaussian distribution." ('a' inserted).

Page 96   Equation (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

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

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 $-(x - \mu_{\mathrm{ML}})/\sigma^2$, and its expectation that defines the regression function is a straight line given by $-(\mu - \mu_{\mathrm{ML}})/\sigma^2$. 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, $\mu - \mu_{\mathrm{ML}}$ should be replaced by $-(\mu - \mu_{\mathrm{ML}})/\sigma^2$.

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 101   First line after Equation (2.154): $a = 1 + \beta/2$ should read $a = (1 + \beta)/2$

Page 103   Last paragraph, line 4: There should be a closing paranthesis ('$)$') after 2.3.9, before the full stop.

Page 103   Line –3: A space should be inserted before the sentence starting "Note that . . . ".

Page 106   Equation (2.168): For additional clarity, prefix the left and right expressions with $\overline{x}_1 =$ and $\overline{x}_2 =$, respectively.

Page 106   Figure 2.17: $\bar{\mathbf{x}}$, $\bar{r}$ and $\bar{\theta}$ should be $\overline{\mathbf{x}}$, $\overline{r}$ and $\overline{\theta}$, 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 $A(m_{\mathrm{ML}})$.

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 $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \geqslant 0$, a sufficient condition for the requirement $p(\mathbf{x}) \geqslant 0$ is that $\pi_k \geqslant 0$ for all k."

Page 114   First line after Equation (2.204): "$\mathbf{x} = (x_1, \ldots, x_N)^{\mathrm{T}}$" should be "$\mathbf{x} = (x_1, \ldots, x_M)^{\mathrm{T}}$".

Page 115   Line before Equation (2.215): $(\eta_1, \ldots, \eta_{M-1})^{\mathrm{T}}$ should be $(\eta_1, \ldots, \eta_{M-1}, 0)^{\mathrm{T}}$.

Page 116   Equation (2.222): $h(\mathbf{x})$ should be $h(x)$ on the l.h.s.

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 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 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 129   Exercise 2.7, line 3: "mean value of x" should be "mean value of μ".

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 $\mathbb{E}[\mathbf{x}_n\mathbf{x}_m^{\mathrm{T}}]$.

Page 139   Third paragraph, line –2: "fit a different" should be "fitting a different".

Page 139   Second line after Equation (3.6): The inline equation should read: $\tanh(a) = 2\sigma(2a) - 1$.

Page 141   Equation (3.13): The r.h.s. should be multiplied by a factor β.

Page 143   Section 3.1.2, line 8: $\phi_j(\mathbf{x}_n)$ should be $\boldsymbol{\varphi}_j$.

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 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 152   Third paragraph, line –1: "question model" should be "question of model".

Page 156   Equation (3.56): $\sum_{j=1}^{M}$ should be replaced by $\sum_{j=0}^{M-1}$ in the argument to the exponential on the r.h.s.

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 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 occurence of "increase" should exchange their positions in the sentence.

Page 165 Section 3.5, first paragraph, line 4: Change "over either" to "either over".

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

Page 168 Figure 3.14, caption: "model evidence" should be "model log evidence".

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 173 Exercise 3.1, last two lines: u_1 and w_1 should be replaced by u_0 and w_0, respectively.

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 176 Exercise 3.14, line –2: $\{\psi_1, \ldots, \psi_M\}$ should be replaced by $\{\psi_0, \ldots, \psi_{M-1}\}$.

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

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 $\widehat{\mathbf{x}}$.

Page 185 Line –3: "these point" should be "these points".

Page 192 Equations (4.47), (4.48) and (4.50): $\mathbf{s}_W$ and $\mathbf{s}_B$ should be $\mathbf{S}_W$ and $\mathbf{S}_B$, respectively.

Page 192 Equation (4.51): This equation should read:

$$J(\mathbf{W}) = \text{Tr}\left\{(\mathbf{W}^{\text{T}}\mathbf{S}_W\mathbf{W})^{-1}(\mathbf{W}^{\text{T}}\mathbf{S}_B\mathbf{W})\right\}.$$

Page 192 Second paragraph, line –3: $J(\mathbf{w})$ should be $J(\mathbf{W})$.

Page 193 Biography of Frank Rosenblatt: Frank Rosenblatt died in 1971.

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 197 Figure 4.9, caption: "probit function" should be "inverse probit function".

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

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

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

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

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

$$\text{erf}(a) = \frac{2}{\sqrt{\pi}} \int_0^a \exp(-\theta^2)\, d\theta$$

(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

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

Note that $\boldsymbol{\Phi}$ should be Φ (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 221   Exercise 4.4: (4.23) should be (4.22).

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   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 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 238 Equation (5.37): "for all $\mathbf{v}$" should be "for all $\mathbf{v} \neq \mathbf{0}$".

Page 239 Figure 5.6, last line of the caption: "'eigenvectors" should be "eigenvalues".

Page 239 First line after Equation (5.39): "'strictly" should be inserted before the second "positive".

Page 241 Second paragraph, line 2: "To see, this" should be "To see this,".

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

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

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

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 256 Section 5.5, line 1: "outputs" should be "output".

Page 257 Section 5.5.1, line 1: An "it" should be inserted before the last "is".

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, "$\alpha_1^{\mathrm{b}} = 100$", should be "$\alpha_1^{\mathrm{b}} = 100^{-2}$" or, simpler, "$\alpha_1^{\mathrm{b}} = 10^{-4}$".

Page 262 Line –1: "approach 2" should be "approach 1".

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

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

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

Page 266 Equation (5.132): This equation should read

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

Page 267 First paragraph: Both occurrences of $O(\xi)$ should be replaced by $O(\xi^2)$. On the first line following Equation (5.133), "to leading order in ξ" should be replaced by "to order ξ^2"

Page 268 Line 1: Insert "a" before "whole".

Page 270–272 Section 5.5.7, from Equation (5.139) onwards: With the introduction of the σ_j^2s, the regularization coefficient becomes irrelevant and hence it can be dropped from text and equations.

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

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

Page 273 Equation (5.148): $\mathbf{I}$ should multiply $\sigma_k^2(\mathbf{x})$ on the r.h.s.

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 (5.153): $\mathbf{I}$ should multiply $\sigma_k^2(\mathbf{x}_n, \mathbf{w})$ 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 282   Equation (5.183): $+\mathrm{const}$ on the r.h.s. should be omitted.

Page 284   Equation (5.190): $\mathbf{b}^{\mathrm{T}}\mathbf{w}_{\mathrm{MAP}}$ should be replaced by a_{MAP} on the r.h.s.

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

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 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".

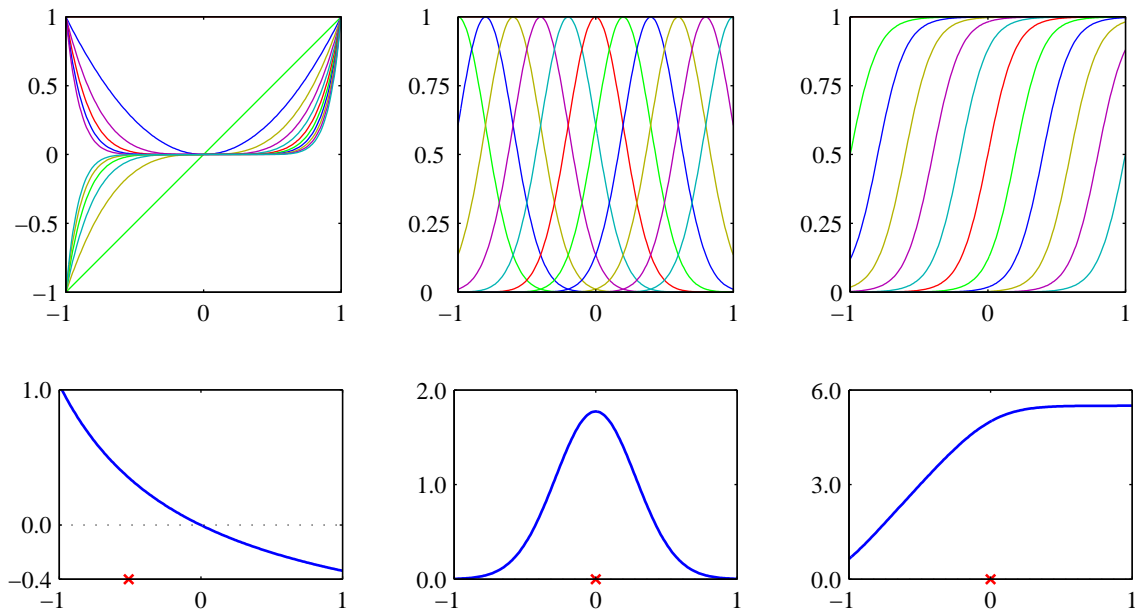**Page 295**   Figure 6.1: The figure and caption should be replaced by

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 295   Second paragraph, Line –2: The period ('.') should be moved up to the previous line.

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

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

Page 329   Second paragraph, Line –2: "bounded below" should be "bounded above".

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

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 347   Second paragraph: In the one but last sentence, following $\phi_i(\mathbf{x}_n)$, insert "for $i = 1, \ldots, N$ and $\Phi_{nM} = 1$ for $n = 1, \ldots, N$" before the comma. The last sentence should be omitted.

Page 350   Caption of Figure 7.10, line –2: "contrition" should be "contribution".

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

Page 351   Equation (7.94):
$$|1 + \alpha_i^{-1}\boldsymbol{\varphi}_i^{\mathrm{T}}\mathbf{C}_{-i}^{-1}\boldsymbol{\varphi}_i|$$
should be
$$\left(1 + \alpha_i^{-1}\boldsymbol{\varphi}_i^{\mathrm{T}}\mathbf{C}_{-i}^{-1}\boldsymbol{\varphi}_i\right)$$
on the r.h.s.

Page 352   Line 3: $\boldsymbol{\varphi}_n$ should be $\boldsymbol{\varphi}_i$.

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

Page 355   Equation (7.118): $, \beta$ should be omitted on the l.h.s.

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

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 $\widehat{x}$ and $\widehat{t}$, 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 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 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 ψ 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 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

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

Page 405–406 Last paragraph on 405 upto and including Equation (8.69): X_{ml} should be replaced by X_{lm} throughout (text, equations and Figure 8.48).

Page 409 Equation (8.79): "$\mu_{x_2 \to f_b}$." should be "$\mu_{x_2 \to f_b}(x_2)$." on the r.h.s.

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

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

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 413 Equation (8.93): f_s should be f under the summation operator on the r.h.s.

Page 414 Last unnumbered equation before Equation (8.99) as well as Equation (8.101): $\mu_{x_{n-1} \to f_{n-1,n}}(x_n)$ should be $\mu_{x_{n-1} \to f_{n-1,n}}(x_{n-1})$ on the r.h.s.

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 $0 \leqslant \mu_i \leqslant 1$ for $i = 0, \ldots, M$.". Moreover, the last sentence of the exercise should be: Discuss the interpretation of the μ_is.

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.

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

Page 434   Equation (9.15): σ_j should be σ_j^D in the denominator on the r.h.s.

Page 435   Third paragraph, line 3: "will play" and "discuss" should be "played" and "discussed', respectively.'

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

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

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

Page 440   Second paragraph, line 4: Insert "log" before "likelihood".

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: α should be $\boldsymbol{\alpha}$.

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

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

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: $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(\mathrm{old})})$ should be $\mathcal{L}(q, \boldsymbol{\theta}^{(\mathrm{old})})$.

Page 453   Line 6: "convex" should be "concave".

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

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

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

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

$$\mathrm{D}_{\mathrm{H}}(p\|q) = \int \left(p(x)^{1/2} - q(x)^{1/2}\right)^2 \mathrm{d}x.$$

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

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

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

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

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

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)–(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 followng 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 479 Equation (10.69): α_k should be α_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 ϕ_n should be $\boldsymbol{\phi}_n$.

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

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

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

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

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

Page 491 Equation (10.121): "$\boldsymbol{\chi}_N = \boldsymbol{\chi}_0 \ldots$" should be "$\nu_N\boldsymbol{\chi}_N = \nu_0\boldsymbol{\chi}_0 \ldots$".

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

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 495 Last line before Equation (10.132): An opening quote (') is missing before max'.

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

Page 497   Equations (10.142)–(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

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

respectively.

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

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

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   Line 4: (10.159) should be (10.160).

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

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

Page 508   Figure 10.14, caption, line −1: "obtained by' should be inserted between "that" and "variational".

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

Page 512   Equations (10.217) and (10.218): On the left hand sides, $\mathbf{m}$ and v should be $\mathbf{m}^{\mathrm{new}}$ and v^{new}, respectively.

Page 513   First line after Equation (10.224): An "of" should be inserted after "Examples".

Page 515   Figure 10.18: $\tilde{f}$ should be $\widetilde{f}$ 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 10.29: All instances of λ should be replaced with η.

Page 521   Exercise 10.30, line 3: "second order" should be "first order".

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

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

Page 526 Equation (11.5):

$$\frac{dz}{dy} \quad \text{should be} \quad \frac{\mathrm{d}z}{\mathrm{d}y}.$$

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

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

Page 529 Line –1: The final full stop ('.') should be removed.

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

$$q(z) = k_i \lambda_i \exp\left\{-\lambda_i\left(z - z_i\right)\right\} \qquad \widehat{z}_{i-1,i} < z \leqslant \widehat{z}_{i,i+1}$$

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

Page 534 First line: $f(\mathbf{z})$ should be ommitted.

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

Page 536 Last line of Equation (11.27): $\mathbf{z}_l$ should be $\mathbf{z}^{(l)}$.

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_τ should be $z^{(\tau)}$.

Page 541 Equation (11.45): This equation should read

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

Page 544 Line 4: $p(z_i|\{\mathbf{z}_{\backslash i})$ should be $p(z_i|\mathbf{z}_{\backslash i})$ (erroneous '{' removed).

Page 545 Equation (11.50): α_i^2 should be α^2 in the last term on the r.h.s.

Page 547 Figure 11.13: Both instances of $\tilde{p}$ should be $\widetilde{p}$.

Page 550 Equation (11.62), second line: $+$ and $-$ should be swapped.

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

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 556 Exercise 11.7: The roles of y and z in the text of the exercise should be swapped in order to be consistent with the notation used in Section 11.1.2, including Equation (11.16).

Page 560 Line –2: "w" should be "we".

Page 564 First line after Equation (12.13): Insert "in (12.10)" after b_i.

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 yellow 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): "σ_i is the variance" should be "σ_i is the standard deviation".

Page 572 Figure 12.9: All instances of $\hat{z}$ should be $\widehat{z}$.

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 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^{th} row" should be "$M \times N$ whose n^{th} column".

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

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

Page 581　Last line: The fragment "introduced by' should be ommitted'.

Page 582　Third paragraph, line 1: "log marginal" should be just "marginal".

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

Page 586　Equations (12.69)–(12.70): On the l.h.s. $\mathbf{W}^{\text{new}}$ and $\boldsymbol{\Psi}^{\text{new}}$ should be $\mathbf{W}_{\text{new}}$ and $\boldsymbol{\Psi}_{\text{new}}$, 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 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 607　Equation (13.1): The r.h.s. should read

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

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

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 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 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 624   Second paragraph, last sentence: This sentence, starting "Since the observed variables ...", should be omitted.

Page 624   Third paragraph, line –2: "seen" should be "occur".

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 626   Equation (13.51): f should be replaced by $\mathbf{z}$ to the right of the arrow in the (subscript) message indices on both sides.

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

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

Page 630   Equation (13.70): This should read

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

(missing $\ln$ inserted).

Page 630   Equation (13.71): This should read

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

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

Page 632   Line –2: "excessive the number" should be just "excessive number".

Page 635   Section 13.3, first paragraph, line –4: $\mathbf{z}_1,\ldots,\mathbf{x}_N$ should be $\mathbf{z}_1,\ldots,\mathbf{z}_N$.

Page 636   Fourth paragraph, lines 3–4: The first instance of $\mathbf{z}_n$ 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 639   Last paragraph: All instances of $\mathbf{CAz}_{n-1}$ should be replaced by $\mathbf{CA}\boldsymbol{\mu}_{n-1}$.

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

Page 666 Equation (14.32): Insert a minus sign('$-$') before the summation on the r.h.s.

Page 666 First sentence after Equation (14.33): This sentence should read: "These both vanish if $p_{\tau k} = 1$ for any one $k = 1, \ldots, K$ (in which case $p_{\tau j} = 0$ for all $j \neq k$) and have their maxima at $p_{\tau k} = 1/K$ for all $k = 1, \ldots, K$."

Page 674 Exercise 14.4, line 2: "hods" should be "holds".

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 678 Paragraph 1, Line 4: "ideas" should be "idea".

Page 681 Paragraph 1, Line –4: "six distinct segments" should be "ten distinct segments".

Page 688 Equation (B.29): α should be a.

Page 690 Equation (B.57): The r.h.s.should read $-\mu_j \mu_k, j \neq k$.

Page 690 Equation (B.58): K should replace M as the upper limit of the sum on the r.h.s. and the equation should end in a period.

Page 691 First line: The sentence fragment starting the page is irrelevant and should be omitted.

Page 691 Equation (B.59): Two instances of M on the r.h.s. should both be replaced by K. The correct form is

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

Page 691 Equation (B.62): The r.h.s.should read $-N\mu_j \mu_k, j \neq k$.

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

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

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 711 Column 2, entry 2: "S. I. Amari" should be "S. Amari".

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 714 Column 1, entry –4: "Tao" should be "Tiao".

Page 719 Column 1, entry 7: "Jeffries" should be "Jeffreys".

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".