PRML Errata

Yousuke Takada

August 28, 2015

Preface

This report communicates some possible errata for PRML (Bishop, 2006) that are not listed in the official errata document (Svensén and Bishop, 2011)[1] at the time of this writing. When specifying the location of an error, I follow the notational conventions adopted by Svensén and Bishop (2011). I have also included in this report some suggestions for improving the readability.

Corrections

Page 51

Equation (1.98): Following the notation (1.93), we should write the left hand side of (1.98) as $\mathrm{H}[X]$ instead of $\mathrm{H}[p]$.

Page 80

Equation (2.52): We usually take eigenvectors $\mathbf{u}_i$ to be the columns of $\mathbf{U}$ as in (C.37). If we follow this convention, Equation (2.52) and the following text should read

$$\mathbf{y} = \mathbf{U}^{\mathrm{T}}(\mathbf{x} - \boldsymbol{\mu}) \tag{1}$$

where $\mathbf{U}$ is a matrix whose columns are given by $\mathbf{u}_i$ so that $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_D)$. From (2.46) it follows that $\mathbf{U}$ is an *orthogonal* matrix, i.e., it satisfies $\mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{I}$ and hence also $\mathbf{U}\mathbf{U}^{\mathrm{T}} = \mathbf{I}$ where $\mathbf{I}$ is the identity matrix.

Page 81

Equations (2.53) and (2.54): If we write the change of variable from $\mathbf{x}$ to $\mathbf{y}$ as (1) instead of (2.52), the Jacobian matrix $\mathbf{J} = (J_{ij})$ we require is simply given by $\mathbf{U}$. Equation (2.53) should read

$$J_{ij} = \frac{\partial x_i}{\partial y_j} = U_{ij} \tag{2}$$

where U_{ij} is the (i, j)-th element of the matrix $\mathbf{U}$. The square of the determinant of the Jacobian matrix (2.54) can be evaluated as follows.

$$|\mathbf{J}|^2 = |\mathbf{U}|^2 = \left|\mathbf{U}^{\mathrm{T}}\right| |\mathbf{U}| = \left|\mathbf{U}^{\mathrm{T}}\mathbf{U}\right| = |\mathbf{I}| = 1. \tag{3}$$

[1] The last line but one of the bibliographic information page of the copy of PRML I have reads "9 8 7 (corrected at 6th printing 2007)." So I refer to Version 2 of the errata.

Page 81

Line -1: Since the determinant of the Jacobian, which is an orthonormal matrix here, can be negative, we should write $|\mathbf{J}| = \pm 1$ instead of $|\mathbf{J}| = 1$.

Page 82

Equation (2.56): We should take the absolute value of the determinant for the same reason given above; the factor $|\mathbf{J}|$ should read $|\det(\mathbf{J})|$. Note that we cannot write $||\mathbf{J}||$ to mean $|\det(\mathbf{J})|$ because it is confusingly similar to the matrix norm $\|\mathbf{J}\|$, which usually refers to the largest singular value of $\mathbf{J}$ (Golub and Van Loan, 2013). This notational inconsistency has been caused by the abuse of the notation $|\cdot|$ for both the absolute value and the matrix determinant. If we always use $\det(\cdot)$ for the determinant, confusion will not arise and the notation be consistent. An alternative solution to this problem would be to explicitly define

$$|\mathbf{A}| \equiv |\det(\mathbf{A})| \tag{4}$$

for any square matrix $\mathbf{A}$. This notation is mostly consistent because we have $|\mathbf{A}| = \det(\mathbf{A})$ for a positive semidefinite matrix $\mathbf{A}$ and most other matrices for which we take determinants are positive (semi)definite in PRML.

Page 104

The text after Equation (2.160): The Gaussian $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Lambda})$ should read $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Lambda}^{-1})$.

Page 141

Equation (3.13): The use of the gradient operator is not consistent here. As in Equation (2.224), the gradient (of a scalar function) is a column vector so that Equation (3.13) should read[2]

$$\nabla \ln p(\mathbf{t}|\mathbf{w}, \beta) = \beta \sum_{n=1}^{N} \left\{ t_n - \mathbf{w}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x}_n) \right\} \boldsymbol{\phi}(\mathbf{x}_n). \tag{5}$$

Moreover, I would like to also suggest that we should give a definition for the gradient before we use it or in an appendix. Although Appendix C defines the vector derivative $\frac{\partial}{\partial \mathbf{x}}$, which is used interchangeably with the gradient $\nabla_{\mathbf{x}}$ in PRML, it has no mention of the gradient.

Page 142

Equation (3.14): The left hand side should be a zero vector $\mathbf{0}$ instead of a scalar zero 0. Thus, Equation (3.14) should read

$$\mathbf{0} = \sum_{n=1}^{N} t_n \boldsymbol{\phi}(\mathbf{x}_n) - \left(\sum_{n=1}^{N} \boldsymbol{\phi}(\mathbf{x}_n) \boldsymbol{\phi}(\mathbf{x}_n)^{\mathrm{T}} \right) \mathbf{w} \tag{6}$$

where we have used the gradient of the form (5) instead of (3.13).

Page 146

Equation (3.31): The left hand side should be $\mathbf{y}(\mathbf{x}, \mathbf{W})$ instead of $\mathbf{y}(\mathbf{x}, \mathbf{w})$.

[2] I have not got a right typeface for the data vector $(t_1, \ldots, t_N)^{\mathrm{T}}$.

Page 166

The second paragraph, Line 1: 'Gamma' should read 'gamma' (without capitalization).

Pages 168–169, and 177

Equations (3.88), (3.93), and (3.117) as well as the text before (3.93): The derivative operators should be partial differentials. For example, Equation (3.117) should read

$$\frac{\partial}{\partial \alpha} \ln |\mathbf{A}| = \operatorname{Tr}\left(\mathbf{A}^{-1}\frac{\partial}{\partial \alpha}\mathbf{A}\right). \tag{7}$$

Page 207

Equation (4.92): On the right hand side, the gradient and the Hessian, which are in general functions of the parameter $\mathbf{w}$, must be evaluated at the previous estimate $\mathbf{w}^{\text{old}}$ for the parameter. Thus, Equation (4.92) should read

$$\mathbf{w}^{\text{new}} = \mathbf{w}^{\text{old}} - \left[\mathbf{H}\left(\mathbf{w}^{\text{old}}\right)\right]^{-1} \nabla E\left(\mathbf{w}^{\text{old}}\right) \tag{8}$$

where $\mathbf{H}(\mathbf{w}) \equiv \nabla\nabla E(\mathbf{w})$ is the Hessian matrix whose elements comprise the second derivatives of $E(\mathbf{w})$ with respect to the components of $\mathbf{w}$.

Page 210

Equation (4.110): The left hand side of (4.110), which is obtained by taking the gradient of $\nabla_{\mathbf{w}_j} E$ given in (4.109) with respect to $\mathbf{w}_k$, refers to the (k, j)-th block of the Hessian, *not* the (j, k)-th. Thus, Equation (4.110) should read

$$\nabla_{\mathbf{w}_k}\nabla_{\mathbf{w}_j} E(\mathbf{w}_1, \ldots, \mathbf{w}_K) = \sum_{n=1}^{N} y_{nj}(I_{kj} - y_{nk})\,\boldsymbol{\phi}_n\boldsymbol{\phi}_n^{\mathrm{T}}. \tag{9}$$

To be clear, we have used the following notation. If we group all the parameters $\mathbf{w}_1, \ldots, \mathbf{w}_K$ into a column vector

$$\mathbf{w} = \begin{pmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_K \end{pmatrix} \tag{10}$$

the gradient and the Hessian of the error function $E(\mathbf{w})$ with respect to $\mathbf{w}$ are given by

$$\nabla_{\mathbf{w}} E = \begin{pmatrix} \nabla_{\mathbf{w}_1} E \\ \vdots \\ \nabla_{\mathbf{w}_K} E \end{pmatrix}, \qquad \nabla_{\mathbf{w}}\nabla_{\mathbf{w}} E = \begin{pmatrix} \nabla_{\mathbf{w}_1}\nabla_{\mathbf{w}_1} E & \cdots & \nabla_{\mathbf{w}_1}\nabla_{\mathbf{w}_K} E \\ \vdots & \ddots & \vdots \\ \nabla_{\mathbf{w}_K}\nabla_{\mathbf{w}_1} E & \cdots & \nabla_{\mathbf{w}_K}\nabla_{\mathbf{w}_K} E \end{pmatrix} \tag{11}$$

respectively.

Page 239

Figure 5.6: The eigenvectors $\mathbf{u}_1$ and $\mathbf{u}_2$ in Figure 5.6 are unit vectors; their orientations should be shown as in Figure 2.7. Or, the scaled vectors should be labeled as $\lambda_1^{-1/2}\mathbf{u}_1$ and $\lambda_2^{-1/2}\mathbf{u}_2$.

Page 251

The second paragraph: The approximation of the form (5.84) is usually referred to as the "*Gauss–Newton*" approximation, but not "*Levenberg–Marquardt*." The Levenberg–Marquardt method is a method that improves the numerical stability of (Gauss–)Newton iterations by correcting the Hessian matrix so as to be more diagonal dominant (Press et al., 1992).

Page 275

The text after Equation (5.154): The identity matrix $\mathbf{I}$ should multiply $\sigma_k^2(\mathbf{x}_n)$.

Page 277

Equation (5.160): The factor L should multiply $\sigma_k^2(\mathbf{x})$ because we have

$$\begin{aligned} s^2(\mathbf{x}) &= \mathbb{E}\left[\mathrm{Tr}\left\{(\mathbf{t} - \mathbb{E}[\mathbf{t}|\mathbf{x}])(\mathbf{t} - \mathbb{E}[\mathbf{t}|\mathbf{x}])^{\mathrm{T}}\right\}\Big|\mathbf{x}\right] && (12) \\ &= \sum_{k=1}^{K} \pi_k(\mathbf{x})\, \mathrm{Tr}\left\{\sigma_k^2(\mathbf{x})\mathbf{I} + (\boldsymbol{\mu}_k(\mathbf{x}) - \mathbb{E}[\mathbf{t}|\mathbf{x}])(\boldsymbol{\mu}_k(\mathbf{x}) - \mathbb{E}[\mathbf{t}|\mathbf{x}])^{\mathrm{T}}\right\} && (13) \\ &= \sum_{k=1}^{K} \pi_k(\mathbf{x}) \left\{L\sigma_k^2(\mathbf{x}) + \|\boldsymbol{\mu}_k(\mathbf{x}) - \mathbb{E}[\mathbf{t}|\mathbf{x}]\|^2\right\} && (14) \end{aligned}$$

where L is the dimensionality of $\mathbf{t}$.

Page 295

Line 1: The vector $\mathbf{x}$ should be a column vector so that $\mathbf{x} = (x_1, x_2)^{\mathrm{T}}$.

Page 318

The text before Equation (6.93) as well as Equations (6.93) and (6.94): The text and the equations should read: We can evaluate the derivative of $a_n^\star$ with respect to θ_j by differentiating the relation (6.84) with respect to θ_j to give

$$\frac{\partial \mathbf{a}_N^\star}{\partial \theta_j} = \frac{\partial \mathbf{C}_N}{\partial \theta_j}(\mathbf{t}_N - \boldsymbol{\sigma}_N) - \mathbf{C}_N \mathbf{W}_N \frac{\partial \mathbf{a}_N^\star}{\partial \theta_j} \tag{15}$$

where the derivatives are Jacobians defined by (C.16) for a vector and analogously for a matrix[3]. Rearranging (15) then gives

$$\frac{\partial \mathbf{a}_N^\star}{\partial \theta_j} = (\mathbf{I} + \mathbf{C}_N \mathbf{W}_N)^{-1} \frac{\partial \mathbf{C}_N}{\partial \theta_j}(\mathbf{t}_N - \boldsymbol{\sigma}_N). \tag{16}$$

Page 355

Equation (7.117): The typeface of the vector $\mathbf{y}$ in (7.117) should be that in (7.110).

Page 414

Figure 8.53, Line 6: The term "max-product" should be "max-sum."

[3] The Jacobian $\frac{\partial \mathbf{A}}{\partial \theta}$ of a matrix $\mathbf{A} = (A_{ij})$ with respect to a scalar θ is a matrix with the same dimensionality as $\mathbf{A}$ whose (i,j)-th element is given by $\frac{\partial A_{ij}}{\partial \theta}$.

Page 425

Equation (9.3): The right hand side should be a zero vector $\mathbf{0}$ instead of a scalar zero 0.

Page 432

The text after Equation (9.13): It should be noted for clarity that, as the prior $p(\mathbf{z})$ over $\mathbf{z}$ is a multinomial distribution (9.10), the posterior $p(\mathbf{z}|\mathbf{x})$ over $\mathbf{z}$ given $\mathbf{x}$ is again a multinomial of the form

$$p(\mathbf{z}|\mathbf{x}) = \prod_{k=1}^{K} \gamma_k^{z_k} \tag{17}$$

where we have written $\gamma_k \equiv \gamma(z_k)$, which can be directly confirmed by inspecting the functional form of the joint distribution

$$p(\mathbf{z})\, p(\mathbf{x}|\mathbf{z}) = \prod_{k=1}^{K} \left\{\pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\right\}^{z_k}. \tag{18}$$

This observation helps the reader to understand that evaluating the responsibilities $\gamma(z_k)$ indeed corresponds to the E step of the general EM algorithm.

Page 434

Equation (9.15): Although the official errata (Svensén and Bishop, 2011) states that σ_j on the right hand side should be raised to a power of D, the whole right hand side should be raised to D so that Equation (9.15) should read

$$\mathcal{N}\left(\mathbf{x}_n \middle| \mathbf{x}_n, \sigma_j^2 \mathbf{I}\right) = \frac{1}{\left(2\pi\sigma_j^2\right)^{D/2}}. \tag{19}$$

Page 435

Equation (9.16): The right hand side should be a zero vector $\mathbf{0}$.

Page 465

Equations (10.6) and (10.7): In PRML, Equation (10.6) will be later recognized as "a negative Kullback-Leibler divergence between $q_j(\mathbf{Z}_j)$ and $\tilde{p}(\mathbf{X}, \mathbf{Z}_j)$" (Page 465, Line −2). However, there is no point in taking a Kullback-Leibler divergence between two probability distributions over different sets of random variables; such a quantity is undefined. Moreover, the discussion here seems to be somewhat redundant. We actually do not have to introduce the probability $\tilde{p}(\mathbf{X}, \mathbf{Z}_j)$ other than $q_j^\star(\mathbf{Z}_j)$. Specifically, we can rewrite Equations (10.6) and (10.7) into

$$\begin{aligned} \mathcal{L}(q) &= \dots && (20)\\ &= \int q_j \ln q_j^\star \mathrm{d}\mathbf{Z}_j - \int q_j \ln q_j \mathrm{d}\mathbf{Z}_j + \text{const} && (21)\\ &= -\,\mathrm{KL}\left(q_j \middle\| q_j^\star\right) + \text{const} && (22) \end{aligned}$$

where we have defined a new distribution $q_j^\star(\mathbf{Z}_j)$ by the relation

$$\ln q_j^\star(\mathbf{Z}_j) = \mathbb{E}_{i\neq j}\left[\ln p(\mathbf{X}, \mathbf{Z})\right] + \text{const}. \tag{23}$$

It directly follows from (22) that, since the lower bound $\mathcal{L}(q)$ is the negative Kullback-Leibler divergence between $q_j(\mathbf{Z}_j)$ and $q_j^\star(\mathbf{Z}_j)$ up to some additive constant, the maximum of $\mathcal{L}(q)$ occurs when $q_j(\mathbf{Z}_j) = q_j^\star(\mathbf{Z}_j)$.

Page 465

The text before Equation (10.8): The latent variable $\mathbf{z}_i$ should read $\mathbf{Z}_i$.

Page 465

Line -1: If we adopt the representation (22), the probability $\tilde{p}(\mathbf{X}, \mathbf{Z}_j)$ should read $q_j^\star(\mathbf{Z}_j)$.

Page 466

Line 1: Again, $\tilde{p}(\mathbf{X}, \mathbf{Z}_j)$ should read $q_j^\star(\mathbf{Z}_j)$. The sentence "Thus we obtain..." should read "Thus we see that we have already obtained a general expression for the optimal solution in (23)."

Page 468

The text after Equation (10.16): The constant term in (10.16) is the *negative* entropy of $p(\mathbf{Z})$.

Page 478

Equation (10.63): The additive constant $+1$ on the right hand side should be omitted so that Equation (10.63) should read

$$\nu_k = \nu_0 + N_k. \tag{24}$$

A quick check for the correctness of the re-estimation equations would be to consider a limit of $N \to 0$, in which the effective number of observations N_k also goes to zero and the re-estimation equations should reduce to identities. Equation (10.63) does not reduces to $\nu_k = \nu_0$, failing the test. Note that the solution for Exercise 10.13 given by Svensén and Bishop (2009) correctly derives the result (24).

Page 489

Equation (10.107): The expectations $\mathbb{E}_\alpha \left[\ln q(\mathbf{w})\right]_\mathbf{w}$ and $\mathbb{E}\left[\ln q(\alpha)\right]$ should read $\mathbb{E}_\mathbf{w}\left[\ln q(\mathbf{w})\right]$ and $\mathbb{E}_\alpha \left[\ln q(\alpha)\right]$, respectively, where the expectation $\mathbb{E}_\mathbf{Z}[\cdot]$ is taken over $q(\mathbf{Z})$.

Page 489

Equations (10.108) through (10.112): The expectations are notationally inconsistent with (1.36); they should be of the forms shown in (10.107) or the ones corrected as above.

Page 490

The third paragraph, Line 2: A comma (,) should be inserted after the ellipsis so that the range of n should read: $n = 1, \ldots, N$.

Page 496

Equation (10.140): In order to be consistent with the mathematical notations in PRML, the differential operator d in (10.140) should be upright d. Moreover, the derivative of x with respect to x^2 should be written with parentheses as $\frac{\mathrm{d}x}{\mathrm{d}(x^2)}$, instead of $\frac{\mathrm{d}x}{\mathrm{d}x^2}$, to avoid ambiguity.

Page 501

The text after Equation (10.162): The variational parameter $\lambda(\xi)$ is a monotonic function of ξ for $\xi \geq 0$, but not that its derivative $\lambda'(\xi)$ is.

Page 503

The text after Equation (10.168): A period (.) should be added at the end of the sentence that follows (10.168).

Page 512

Equation (10.222): The factor $(2\pi v_n)^{D/2}$ in the denominator of the right hand side should be omitted because it has been already included in the Gaussian in (10.213).

Page 513

Equations (10.223) and (10.224): The quantities v^{new} and $\mathbf{m}^{\text{new}}$ in (10.223) and (10.224) are different from those in (10.217) and (10.218)[4]. Thus, we should introduce different notations, say, v and $\mathbf{m}$, with appropriate definitions. Specifically, one can rewrite the approximation to the model evidence in the form

$$p(\mathcal{D}) \simeq (2\pi v)^{D/2} \exp(B/2) \prod_{n=1}^{N} \left\{ s_n (2\pi v_n)^{-D/2} \right\} \tag{25}$$

where

$$B = \frac{\mathbf{m}^{\mathrm{T}}\mathbf{m}}{v} - \sum_{n=1}^{N} \frac{\mathbf{m}_n^{\mathrm{T}}\mathbf{m}_n}{v_n} \tag{26}$$

$$v^{-1} = \sum_{n=1}^{N} v_n^{-1} \tag{27}$$

$$v^{-1}\mathbf{m} = \sum_{n=1}^{N} v_n^{-1}\mathbf{m}_n. \tag{28}$$

Page 515

Equations (10.228) and (10.229): Although Svensén and Bishop (2011) correct (10.228) so that $q^{\backslash b}(\mathbf{x})$ is a normalized distribution, we do not need the normalization of $q^{\backslash b}(\mathbf{x})$ here and, even with this normalization, we cannot ensure that $\hat{p}(\mathbf{x})$ given by (10.229) is normalized. Similarly to (10.195), we can proceed with the unnormalized $q^{\backslash b}(\mathbf{x})$ given by the original (10.228) and, rather than correcting (10.228), we should correct (10.229) so that

$$\hat{p}(\mathbf{x}) \propto q^{\backslash b}(\mathbf{x}) f_b(x_2, x_3) = \dots \tag{29}$$

implying that $\hat{p}(\mathbf{x})$ is a normalized distribution.

Page 515

The text after Equation (10.229): The new distribution $q^{\text{new}}(\mathbf{z})$ should read $q^{\text{new}}(\mathbf{x})$.

Page 516

Equation (10.240): The subscript k of the product $\prod_k \dots$ should read $k \neq j$ because we have already removed the term $\tilde{f}_j(\boldsymbol{\theta}_j)$.

[4]See Svensén and Bishop (2011) for the errata for Equations (10.217) and (10.218).

Page 554

Equation (11.72), Line −2: If we write the expectation $\mathbb{E}_{\mathbf{z}}[\cdot]$ taken over some given distribution $q(\mathbf{z})$ explicitly as $\mathbb{E}_{q(\mathbf{z})}[\cdot]$, the expectation in the last line but one of (11.72) should read

$$\mathbb{E}_{p_G(\mathbf{z})}\left[\exp\left(-E(\mathbf{z})+G(\mathbf{z})\right)\right] \tag{30}$$

where we have written the argument $\mathbf{z}$ for $E(\mathbf{z})$ and $G(\mathbf{z})$ for clarity.

Page 556

Exercise 11.7: The interval should be $[-\pi/2, \pi/2]$ instead of $[0, 1]$.

Page 557

Exercise 11.14, Line 2: The variance should be σ_i^2 instead of σ_i.

Page 564

The text after Equation (12.12): The derivative we consider here is that with respect to b_j (*not* that with respect to b_i).

Page 564

The text after Equation (12.15): The zero should be a zero vector so that we have $\mathbf{u}_j = \mathbf{0}$.

Page 575

The third paragraph, Line 5: The zero vector should be a row vector instead of a column vector so that we have $\mathrm{v}^{\mathrm{T}}\mathbf{U} = \mathbf{0}^{\mathrm{T}}$. Or, the both sides are transposed to give $\mathbf{U}^{\mathrm{T}}\mathbf{v} = \mathbf{0}$.

Page 578

Equation (12.53): As stated in the text preceding (12.53), we should substitute $\boldsymbol{\mu} = \bar{\mathbf{x}}$ into (12.53).

Page 578

The text before Equation (12.56): For the maximization with respect to $\mathbf{W}$, we use (C.25) and (C.27) instead of (C.24).

Page 579

Line 5: The eigendecomposition requires $O(D^3)$ computation*s*.

Page 599

Exercise 12.1, Line −1: The quantity λ_{M+1} is an eigenvalue (not an eigenvector).

Page 602

Exercise 12.25, Line 2: The latent space distribution should read $p(\mathbf{z}) = \mathcal{N}(\mathbf{z}|\mathbf{0}, \mathbf{I})$.

Page 610

The first paragraph, Line −5: The text "our predictions ..." should read: "our predictions for $\mathbf{x}_{n+1}$ depend on all the previous observations."

Page 620

The second paragraph and the following (unlabeled) equation: The last sentence before the equation and the equation each should terminate with a period (.).

Page 621

Figures 13.12 and 13.13: It should be clarified that, similarly to $\alpha(z_{nk})$ and $\beta(z_{nk})$, the notation $p(\mathbf{x}_n|z_{nk})$ is used to denote the value of $p(\mathbf{x}_n|\mathbf{z}_n)$ when $z_{nk} = 1$.

Page 622

The second paragraph, Line −1: "we see" should be omitted.

Page 623

The first paragraph, Line −2: z_{nk} should read $z_{n-1,k}$.

Page 631

Equation (13.73): The equation should read

$$\sum_{r=1}^{R} \ln \left\{ \frac{p(\mathbf{X}_r|\boldsymbol{\theta}_{m_r})\, p(m_r)}{\sum_{l=1}^{M} p(\mathbf{X}_r|\boldsymbol{\theta}_l)\, p(l)} \right\}. \tag{31}$$

Page 637

Equations (13.81), (13.82), and (13.83): The distribution (13.81) over $\mathbf{w}$ should read

$$p(\mathbf{w}) = \mathcal{N}(\mathbf{w}|\mathbf{0}, \boldsymbol{\Gamma}) \tag{32}$$

and so on.

Page 638

The first paragraph, Line 2: "conditional on" should read "conditioned on."

Page 641

The text after Equation (13.103): The form of the Gaussian is unclear. Since a multivariate Gaussian is usually defined over a column vector, we should construct a column vector from the concerned random variables to clearly define the mean and the covariance. Specifically, the text should read for example: . . . , we see that $\xi(\mathbf{z}_{n-1}, \mathbf{z}_n)$ is a Gaussian of the form

$$\xi(\mathbf{z}_{n-1}, \mathbf{z}_n) = \mathcal{N}\left(\begin{pmatrix} \mathbf{z}_{n-1} \\ \mathbf{z}_n \end{pmatrix} \middle| \begin{pmatrix} \hat{\boldsymbol{\mu}}_{n-1} \\ \hat{\boldsymbol{\mu}}_n \end{pmatrix}, \begin{pmatrix} \hat{\mathbf{V}}_{n-1} & \hat{\mathbf{V}}_{n-1,n} \\ \hat{\mathbf{V}}_{n-1,n}^{\mathrm{T}} & \hat{\mathbf{V}}_n \end{pmatrix} \right) \tag{33}$$

where the mean $\hat{\boldsymbol{\mu}}_n$ and the covariance $\hat{\mathbf{V}}_n$ of $\mathbf{z}_n$ are given by (13.100) and (13.101), respectively; and the covariance $\hat{\mathbf{V}}_{n-1,n}$ between $\mathbf{z}_{n-1}$ and $\mathbf{z}_n$ is given by

$$\hat{\mathbf{V}}_{n-1,n} = \mathrm{cov}[\mathbf{z}_{n-1}, \mathbf{z}_n] = \mathbf{J}_{n-1}\hat{\mathbf{V}}_n. \tag{34}$$

Pages 642 and 643

Equation (13.109) and the following equations: If we follow the notation in Chapter 9, the typeface of the Q function should be $\mathcal{Q}$.

Page 642

Equation (13.109): If we follow the notation for a conditional expectation given by (1.37), Equation (13.109) should read

$$\mathcal{Q}\left(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}\right) = \mathbb{E}_{\mathbf{Z}}\left[\ln p\left(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}\right)\middle|\mathbf{X}, \boldsymbol{\theta}^{\text{old}}\right] \tag{35}$$

$$= \int \mathrm{d}\mathbf{Z}\, p\left(\mathbf{Z}\middle|\mathbf{X}, \boldsymbol{\theta}^{\text{old}}\right) \ln p\left(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}\right) \tag{36}$$

which corresponds to (9.30).

Page 643

Equation (13.111): $\mathbf{V}_0^{\text{new}}$ should read $\mathbf{P}_0^{\text{new}}$. Svensén and Bishop (2011) have failed to mention (13.111).

Page 643

Equation (13.114): The size of the opening curly brace '{' should match that of the closing curly brace '}'.

Page 647

Figure 13.23, Line −1: $p(\mathbf{x}_{n+1}|\mathbf{z}_{n+1}^{(l)})$ should read $p(\mathbf{x}_{n+1}|z_{n+1}^{(l)})$.

Page 649

Exercise 13.14, Line 1: (8.67) should be (8.64).

Page 658

Figure 14.1, the equation below: The subscript of the summation in the right hand side should read $m = 1$.

Page 668

Equation (14.37): The arguments of the probability are notationally inconsistent with those of (14.34), (14.35), and (14.36). Specifically, the conditioning on $\boldsymbol{\phi}_n$ should read that on t_n and the probability $p(k|\dots)$ be the value of $p(\mathbf{z}_n|\dots)$ when $z_{nk} = 1$, which we write $p(z_{nk} = 1|\dots)$. Moreover, strictly speaking, the old parameters $\pi_k, \mathbf{w}_k, \beta$ should read $\pi_k^{\text{old}}, \mathbf{w}_k^{\text{old}}, \beta^{\text{old}} \in \boldsymbol{\theta}^{\text{old}}$. In order to solve these problems, we should rewrite Equation (14.37) as, for example,

$$\gamma_{nk} = \mathbb{E}\left[z_{nk}\middle|t_n, \boldsymbol{\theta}^{\text{old}}\right] \tag{37}$$

where we have written the conditioning in the expectation explicitly and the expectation is given by

$$\mathbb{E}\left[z_{nk}|t_n, \boldsymbol{\theta}\right] = p\left(z_{nk} = 1|t_n, \boldsymbol{\theta}\right) = \frac{\pi_k \mathcal{N}\left(t_n\middle|\mathbf{w}_k^{\mathrm{T}}\boldsymbol{\phi}_n, \beta^{-1}\right)}{\sum_j \pi_j \mathcal{N}\left(t_n\middle|\mathbf{w}_j^{\mathrm{T}}\boldsymbol{\phi}_n, \beta^{-1}\right)}. \tag{38}$$

Page 668

The unlabeled equation between (14.37) and (14.38): If we write the implicit conditioning in the expectation explicitly (similarly to the above equations), the unlabeled equation should read

$$\mathcal{Q}\left(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}\right) = \mathbb{E}_{\mathbf{Z}}\left[\ln p(\mathbf{t}, \mathbf{Z}|\boldsymbol{\theta})\middle|\mathbf{t}, \boldsymbol{\theta}^{\text{old}}\right] \tag{39}$$

$$= \ldots \tag{40}$$

where we have again used the typeface $\mathcal{Q}$ for the Q function so that the notation is consistent with that of Chapter 9.

Page 669

Equations (14.40) and (14.41): The left hand sides should read a zero vector $\mathbf{0}$.

Page 669

Equation (14.41): $\boldsymbol{\Phi}$ is undefined. The text following (14.41) should read: ...where $\mathbf{R}_k = \text{diag}(\gamma_{nk})$ is a diagonal matrix of size $N \times N$ and $\boldsymbol{\Phi} = (\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_N)^{\text{T}}$ is an $N \times M$ matrix. Here, N is the size of the data set and M is the dimensionality of the feature vectors $\boldsymbol{\phi}_n$.

Page 669

Equation (14.43): '+const' should be added to the right hand side.

Page 671

The text after Equation (14.46): The text should read: ...where we have omitted the dependence on $\{\boldsymbol{\phi}_n\}$ and defined $y_{nk} = \ldots$. Or, $\boldsymbol{\phi}$ should have been omitted from the left hand side of (14.45).

Page 671

Equation (14.48): The notation should be corrected similarly to the above erratum regarding (14.37).

Page 671

Equation (14.49): The notation should be corrected similarly to the above erratum regarding the unlabeled equation between (14.37) and (14.38).

Page 672

Equation (14.52): The negation must be removed so that $\mathbf{H}_k \equiv \nabla_k \nabla_k \mathcal{Q}$ where

$$\nabla_k \nabla_k \mathcal{Q} = -\sum_{n=1}^{N} \gamma_{nk} y_{nk}(1 - y_{nk}) \boldsymbol{\phi}_n \boldsymbol{\phi}_n^{\text{T}}. \tag{41}$$

Page 674

Exercise 14.1, Line 1: "of" should be inserted after "set."

Page 686

Line -3: The comma in the first inline math should be removed so that the product should read: $m \times (m-1) \times \cdots \times 2 \times 1$.

Page 687

Equation (B.25): The differential operator d should be upright d.

Page 688

Line 1: 'Gamma' should read 'gamma' (without capitalization).

Page 689

Line 1: 'positive-definite' should read 'positive definite' (without hyphenation).

Page 689

Equation (B.49): $\mathbf{x}$ in the right hand side should read $\mathbf{x}_a$.

References

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning.* Springer.

Golub, G. H. and C. F. Van Loan (2013). *Matrix Computations* (Fourth ed.). Johns Hopkins University Press.

Press, W. M., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (1992). *Numerical Recipes in C: The Art of Scientific Computing* (Second ed.). Cambridge University Press.

Svensén, M. and C. M. Bishop (2009). Pattern recognition and machine learning: Solutions to the exercises, web edition. `http://research.microsoft.com/en-us/um/people/cmbishop/prml/pdf/prml-web-sol-2009-09-08.pdf`.

Svensén, M. and C. M. Bishop (2011). Pattern recognition and machine learning: Errata and additional comments. `http://research.microsoft.com/en-us/um/people/cmbishop/prml/pdf/prml-errata-2nd-20110921.pdf`.