Main

Errata

Please email us if you find any errors in the book. We will list known errata on this page.

Chapter 1: Data Mining and Analysis

- p4, Section 1.3, line 13: "as *linear combination*"should be "as a *linear combination*"
- p9, Example 1.3, 3rd line from end: "$(153)^{1/3}$" should be "$(152)^{1/3}$"
- p9, Example 1.3, last line: "$(4^3 + (-1)^3)^{1/3} = (63)^{1/3} = 3.98$" should be "$(4^3 + |-1|^3)^{1/3} = (65)^{1/3} = 4.02$"
- p24, Section 1.4.3, last line of subsection **Univariate Sample**: "where $f_{\mathbf{X}}$ is the probability mass or density function for $\mathbf{X}$" should be "where f_X is the probability mass or density function for X"
- p30, Section 1.7, Q1: "in (1.5)" should be "in Eq. (1.5)"

Chapter 2: Numeric Attributes

- p34, Equation (2.2): "$\hat{F}(x) \geq q$" should be "$F(x) \geq q$"
- p34, Line after Equation (2.2):

  "That is, the inverse CDF gives the least value of X, for which q fraction of the values are higher, and $1 - q$ fraction of the values are lower."

  should be

  "That is, the inverse CDF gives the least value of X, for which q fraction of the values are **lower**, and $1 - q$ fraction of the values are **higher**."
- p53, Example 2.6, line 1: "... range for `Income` is $2700 - 300 = 2400$" should be "... range for `Income` is $6000 - 300 = 5700$"
- p55, In Eq (2.32): "$P(-k \leq z \leq k) = P\big(0 \leq t \leq k/\sqrt{2}\big)$" should be "$P(-k \leq z \leq k) = 2 \cdot P\big(0 \leq t \leq k/\sqrt{2}\big)$"
- p58, Total and Generalized Variance, Line 2: "...product of its eigenvectors" should be "...product of its eigenvalues"
- p58, two lines above Example 2.8: "$tr(\Lambda)$" should be "$tr(\mathbf{\Lambda})$"
- p61, Q3: "mu" should be "μ" so that it reads

$$\sum_{i=1}^{n}(x_i - \mu)^2 = n(\hat{\mu} - \mu)^2 + \sum_{i=1}^{n}(x_i - \hat{\mu})^2$$

Chapter 3: Categorical Attributes

- p81, Table 3.6, Attribute value for X_2: "$\texttt{Short}(a_{23})$" should be "$\texttt{Long}(a_{23})$"

Chapter 4: Graph Data

- p103, 2 lines above Eq (4.3): "$\gamma_{jk} = 0$" should be "$\gamma_{jk}(v_i) = 0$"
- p103, Eq (4.3): "γ_{jk}" should be "$\gamma_{jk}(v_i)$"
- p103, Example 4.5, last line: "$\gamma_{jk} > 0$" should be "$\gamma_{jk}(v_5) > 0$"
- p104, Example 4.5:

  $c(v_5) = \gamma_{18} + \gamma_{24} + \gamma_{27} + \gamma_{28} + \gamma_{38} + \gamma_{46} + \gamma_{48} + \gamma_{67} + \gamma_{68}$

should be

$$c(v_5) = \gamma_{18}(v_5) + \gamma_{24}(v_5) + \gamma_{27}(v_5) + \gamma_{28}(v_5) + \gamma_{38}(v_5) + \gamma_{46}(v_5) + \gamma_{48}(v_5) + \gamma_{67}(v_5) + \gamma_{68}(v_5)$$

- p107: $\mathbf{p}_1 = \frac{1}{2}\begin{pmatrix} 1 \\ 1 \\ 2 \\ 1 \\ 2 \end{pmatrix}$ should be $\mathbf{p}_1 = \frac{1}{2}\begin{pmatrix} 1 \\ 2 \\ 2 \\ 1 \\ 2 \end{pmatrix}$
- p127, 4th Line after Eq (4.22): "initial n_0 edges" should be "initial n_0 nodes"

Chapter 5: Kernel Methods

- p138, Example 5.4:

$$\mu_\phi = \sum_{i=1}^{5} \phi(\mathbf{x}_i) = \sum_{i=1}^{5} \mathbf{x}_i$$

should be

$$\mu_\phi = \frac{1}{5}\sum_{i=1}^{5} \phi(\mathbf{x}_i) = \frac{1}{5}\sum_{i=1}^{5} \mathbf{x}_i$$

- p140, 7th Line after Eq (5.3): "$\sum_{i=1}^{m_a} \sum_{j=1}^{m_a} \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x})$" should be "$\sum_{i=1}^{m_a} \sum_{j=1}^{m_a} \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j)$"
- p141, 3rd line and 10th Line before Sec 5.1.2: There is an extra left bracket in definition of $\phi(\mathbf{x})$, that is,

  "$\big((K(\mathbf{x}_1, \mathbf{x}), \ldots$" should be "$\big(K(\mathbf{x}_1, \mathbf{x}), \ldots$"
- p144, 2nd line: "$\int a(\mathbf{x})^2 \, d\mathbf{x} < 0$" should be "$\int a(\mathbf{x})^2 \, d\mathbf{x} < \infty$"
- p144, last line: "$\sum_{k=1}^{q}$" should be "$\sum_{k=0}^{q}$"
- p156, Section 5.4.2: all occurrences of "path/paths" should be "walk/walks"

Chapter 6: High-dimensional Data

- p164: In the definitions of the hyperball and and hypersphere

  "$\mathbf{x} = (x_1, x_2, \ldots, x_d)$" should be "$\mathbf{x} = (x_1, x_2, \ldots, x_d)^T$"
- p171: "$\mathbf{0}_d = (0_1, 0_2, \ldots, 0_d)$" should be "$\mathbf{0}_d = (0_1, 0_2, \ldots, 0_d)^T$"
- p172, Section 6.6, 1st Line after Eq. (6.11):

  μ in equation "$\mu = \mathbf{0}_d$" should be in bold.
- p178, section "Volume in d dimensions":

  "$x_1 = r\cos\theta_1 \cos\theta_2 \cos\theta_3 = rc_2c_2c_3$" should be "$x_1 = r\cos\theta_1 \cos\theta_2 \cos\theta_3 = rc_1c_2c_3$"

  "$x_3 = r\cos\theta_1 \sin\theta_2 = rc_1s_1$" should be "$x_3 = r\cos\theta_1 \sin\theta_2 = rc_1s_2$"
- p178, Equation for $J(\theta_1, \theta_2, \theta_3)$, Entry in first row, fourth column: "$rc_1c_2s_3$"should be "$-rc_1c_2s_3$"
- p207, line 3, Alg 7.2: "$\eta_1, \eta_2, \ldots, \eta_d$" should be "$\eta_1, \eta_2, \ldots, \eta_n$"

Chapter 7: Dimensionality Reduction

- p186, line 1: "$\mathbf{a}_r$ is vector" should be "$\mathbf{a}_r$ is a vector"
- p207, line 3, Alg 7.2: "$\eta_1, \eta_2, \ldots, \eta_d$" should be "$\eta_1, \eta_2, \ldots, \eta_n$"

Chapter 8: Itemset Mining

- p235, Example 8.13, 2nd last line: "$\ldots, AB(3), AD(4), \ldots$" should be "$\ldots, AB(4), AD(3), \ldots$"
- p236, 5th line: "$\ldots, AD(4), \ldots$" should be "$\ldots, AD(3), \ldots$"

Chapter 9: Summarizing Itemsets

- p250, 2nd line under **Generalized Itemsets**: "k-tidsets" should be "k tidsets"
- p250, 4th line from bottom: "$Z = Y \setminus X$" should be "$Z = X \setminus Y$"
- p252, Eq. (9.3) and Eq. (9.4): "$\big|X \setminus Y\big|$" should be "$\big|X \setminus W\big|$" on the right hand side in both equations, so that they read

$$\textbf{Upper Bounds}\big(|X \setminus Y| \text{is odd}\big) : sup(X) \leq \sum_{Y \subseteq W \subset X} -1^{\big(|X \setminus W|+1\big)} sup(W)$$

$$\textbf{Lower Bounds}\big(|X \setminus Y| \text{is even}\big) : sup(X) \geq \sum_{Y \subseteq W \subset X} -1^{\big(|X \setminus W|+1\big)} sup(W)$$

- p254, Section **Nonderivable Itemsets**, 1st Equation after line 1: "$\big|X \setminus Y\big|$" should be "$\big|X \setminus W\big|$" , so that it reads

$$IE(Y) = \sum_{Y \subseteq W \subset X} -1^{\big(|X \setminus W|+1\big)} \cdot sup(W)$$

Chapter 10: Sequence Mining

- p264, alg 10.2, line 9: "$\mathbf{P}$" should be "P_a"

Chapter 11: Graph Pattern Mining

- p288, sec 11.3, 2nd paragraph, line 6: "$sup(C) = sup(t)$" should be "$sup(C') = sup(t)$"
- p290, Figure 11.8: The last tuple in the DFS-code for graph C_{19} should be "$\langle 2, 0, a, a \rangle$" and not "$\langle 2, 0, a, b \rangle$"
- p292, Algorithm 11.2, Line 14: "$b = \langle u_r, v, L(u_r), L(v), L(u_r, v) \rangle$" should be "$b = \langle u_r, v, L(\phi(u_r)), L(\phi(v)), L(\phi(u_r), \phi(v)) \rangle$"
- p293, Figure 11.9 (c): There there should be one more extension for ϕ_5, namely $\langle 0, 3, a, b \rangle$
- p294, Algorithm 11.3, Line 12: "N_{G_j}" should be "N_G"
- p295, Algorithm 11.4, Line 0: "C" should be "$C = \{t_1, t_2, \ldots, t_k\}$"

Chapter 12: Pattern and Rule Assessment

- p322 (Alg 12.1) and p326 (Alg 12.2): replace "=" with "$\leftarrow$"

Chapter 13: Representative-based Clustering

- p343, in 3rd equation: "$P(C_i)$" should be "$P(C_1)$"
- p335, Algorithm 13.1, line 7: "μ_i^t" should be "μ_i^{t-1}"

Chapter 14: Hierarchical Clustering

- p366, Fig 14.2: "(a) $m = 1$", "(b) $m = 2$", and "(c) $m = 3$" should be "(a) $n = 1$", "(b) $n = 2$", and "(c) $n = 3$", respectively.
- p373, sec 14.4: "EXERCISES AND PROJECTS" should be "EXERCISES"
- p373, Q1, "$SMC(X_i, X_j)$, $JC(X_i, X_j)$, $RC(X_i, X_j)$" should be "$SMC(\mathbf{x}_i, \mathbf{x}_j)$, $JC(\mathbf{x}_i, \mathbf{x}_j)$, $RC(\mathbf{x}_i, \mathbf{x}_j)$", respectively.

Chapter 15: Density-based Clustering

- p385, line after Eq. (15.6): "... having two parts. A vector ... " should be "... having two parts: a vector ..."

Chapter 16: Spectral and Graph Clustering

- p411, 2nd last equation: "$\frac{1}{2}\, p_{rs}$" should be "p_{rs}" so that it reads

$$p_{rs} = \frac{d_r}{2m}\,\frac{d_s}{2m} = \frac{d_r d_s}{4m^2}$$

- p413, Line 5: "$\sum_{j=1}^{n} \mathbf{d}^T\mathbf{c}_i$" should be "$\mathbf{d}^T\mathbf{c}_i$"
- p413, Line 10: "$\left(\mathbf{d}_i^T\mathbf{c}_i\right)^2$" should be "$\left(\mathbf{d}^T\mathbf{c}_i\right)^2$"
- p424, Q5: "$\mathbf{c}_n = \frac{1}{\sqrt{n}}\,\mathbf{1}$" should be "$\mathbf{c}_n = \frac{1}{\sqrt{\sum_{i=1}^{n} d_i}}\,\mathbf{\Delta}^{1/2}\mathbf{1}$"
- p424, Q6 (b): "$\mathbf{K} = \mathbf{M}$" should "$\mathbf{K} = \mathbf{M} + \mathbf{I}$"

Chaper 17: Clustering Validation

- p428, Example 17.1, Table below 2nd para: "$n = 100$" should be "$n = 150$" for the total count
- p463, Q10: Add the sentence "Assume that the clusters are:
$C_1 = \{a, b, c, d, e\}, C_2 = \{g, i\}, C_3 = \{f, h, j\}, C_4 = \{k\}$."

Chapter 18: Probabilistic Classification

- p472, Table 18.2: "13/50" should be "11/50"
- p472, Example 18.2, 2nd Para, lines 6 and 7: "$P(c_1|\mathbf{x})$" and "$P(c_2|\mathbf{x})$" should be "$\hat{P}(c_1|\mathbf{x})$" and "$\hat{P}(c_2|\mathbf{x})$", respectively.

Chapter 20: Linear Discriminant Analysis

- p503: Example 20.2: There should be no transpose operator "T" on the mean vectors, i.e.,

$$\mu_1 = \begin{pmatrix} 5.01 \\ 3.42 \end{pmatrix}^T \qquad \mu_2 = \begin{pmatrix} 6.26 \\ 2.87 \end{pmatrix}^T \qquad \mu_1 - \mu_2 = \begin{pmatrix} -1.256 \\ 0.546 \end{pmatrix}^T$$

  should be

$$\mu_1 = \begin{pmatrix} 5.01 \\ 3.42 \end{pmatrix} \qquad \mu_2 = \begin{pmatrix} 6.26 \\ 2.87 \end{pmatrix} \qquad \mu_1 - \mu_2 = \begin{pmatrix} -1.256 \\ 0.546 \end{pmatrix}$$

- p509, Example 20.4, line 4: "*iris-virginica*" should be "`Iris-versicolor`"
- p512, Q1: In part (a) "$\mathbf{S}_B$" should be "$\mathbf{B}$", and in (b) "$\mathbf{S}_W$" should be "$\mathbf{S}$"

Chapter 21: Support Vector Machines

- p526, 7th line, in L_{dual}: "$(C - \alpha_i + \beta_i)$" should be "$(C - \alpha_i - \beta_i)$"
- p536, Algorithm 21.1, line 15: "$\alpha_{t+1} = \alpha$" should be "$\alpha_{t+1} \leftarrow \alpha$"
- p538, Example 21.8, line 5: "homogeneous quadratic kernel $K(\mathbf{x}_i, \mathbf{x}_j) = \left(\mathbf{x}_i^T\mathbf{x}_j\right)^2$" should be "inhomogeneous quadratic kernel $K(\mathbf{x}_i, \mathbf{x}_j) = \left(1 + \mathbf{x}_i^T\mathbf{x}_j\right)^2$"