Errata for *Introduction to Data Mining, Second Edition* by Tan, Steinbach, Karpatne, and Kumar.

Last updated on November 4, 2021 at 02:22am

Please send all error reports to dmbook@umn.edu

Preface

Page viii, last sentence of Section entitled, **Support Materials**: The email address for reporting errata has been updated to be **dmbook@umn.edu**. However, the old address **dmbook@cs.umn.edu** should still work.

Chapter 2

1. Page 27: The title "What Is an attribute?" should be "What is an Attribute?".

2. Page 77: In the properties of a metric, condition 1(b) should be $d(\mathbf{x}, \mathbf{y}) = 0$ if and only if $\mathbf{x} = \mathbf{y}$.

3. Page 89, The first sentence after equation (2.15): "$I(X, Y) = I(Y)$" should be "$I(X, Y) = I(Y, X)$"

4. : Page 93, the first line: "$\langle \mathbf{x}, \mathbf{y} \rangle$" should be "$\langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle$"

5. Page 93, 2 lines before equation 2.19: "then these two" should be "then these three"

6. Page 93, Example 2.24, First sentence: "presented in the previous section" should be "discussed above"

   The

7. Page 94, Equation 2.24: The inner product should be a sum, not a tuple, so equation 2.24 should be

$$\phi(\mathbf{x}) = x_1^2 + x_2^2 + \sqrt{2}x_1x_2 + \sqrt{2c}x_1 + \sqrt{2c}x_2 + \ c.$$

2 Errata

Chapter 3

1. Page 148,Figure 3.23b should be as follows:

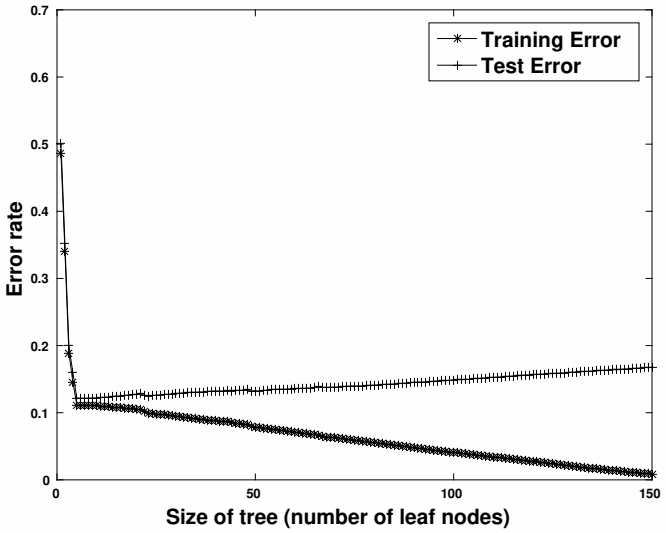

Figure 3.23b Varying tree size from 1 to 150.

2. Page 141, In Figure 3.16, "width $>$ 3" should be "breadth $>$ 3":

Decision Tree:

depth = 1:
| breadth> 7 : **class 1**
| breadth<= 7:
| | breadth <= 3:
| | | ImagePages> 0.375: **class 0**
| | | ImagePages<= 0.375:
| | | | totalPages<= 6: **class 1**
| | | | totalPages> 6:
| | | | | breadth <= 1: **class 1**
| | | | | breadth > 1: **class 0**
| | breadth > 3:
| | | MultiIP = 0:
| | | | ImagePages<= 0.1333: **class 1**
| | | | ImagePages> 0.1333:
| | | | breadth <= 6: **class 0**
| | | | breadth > 6: **class 1**
| | | MultiIP = 1:
| | | | TotalTime <= 361: **class 0**
| | | | TotalTime > 361: **class 1**
depth> 1:
| MultiAgent = 0:
| | depth > 2: **class 0**
| | depth < 2:
| | | MultiIP = 1: **class 0**
| | | MultiIP = 0:
| | | | breadth <= 6: **class 0**
| | | | breadth > 6:
| | | | | RepeatedAccess <= 0.322: **class 0**
| | | | | RepeatedAccess > 0.322: **class 1**
| MultiAgent = 1:
| | totalPages <= 81: **class 0**
| | totalPages > 81: **class 1**

Figure 3.16 Decision tree model for web robot detection.

3. Page 164, In Figure 3.32, "width > 3" should be "breadth > 3":

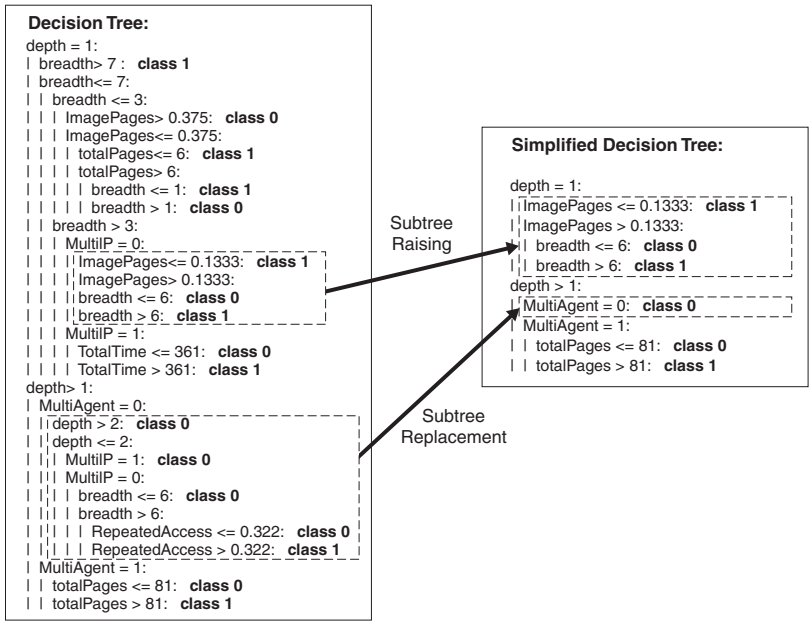

Figure 3.32 Post-pruning of the decision tree for web robot detection.

Chapter 4

1. Page 251, Equation 4.48 should be the following:

$$\hat{y} = \begin{cases} 1, & \text{if } \mathbf{w}^T\mathbf{x} + b > 0. \\ -1, & \text{otherwise.} \end{cases}$$

2. Page 312, second paragraph, the out of bag sample is 37% of the base classifiers, not 27%.

3. Page 322, Table just above Section 4.11.3 should be the following:

$$\text{Weighted accuracy} = \frac{w_1\text{TP} + w_4TN}{w_1\text{TP} + w_2\text{FP} + w_3\text{FN} + w_4\text{TN}}. \qquad (1)$$

The relationship between weighted accuracy and other performance measures is summarized in the following table:

Measure	w_1	w_2	w_3	w_4
Recall	1	0	1	0
Precision	1	1	0	0
F_β	$\beta^2 + 1$	β^2	1	0
Accuracy	1	1	1	1

Chapter 5

1. Page 382, line 3 of algorithm 5.3 should be **if** $k > m$.

Chapter 7

1. Page 586, the second sentence of Example 7.11, which is in parentheses, should be "(The data for this figure consists of the six two-dimensional points given in Table 7.3.)"

2. Page 587, the caption for Table 7.7 should be "Cophenetic distance matrix for single link and data in Table 7.3 on page 557."

3. Page 610, Exercise 29 should be as follows:

   Prove that $\sum_{i=1}^{K} \sum_{x \in C_i} (x - c_i)(c - c_i) = 0$. This fact was used in the proof that TSS = SSE + SSB on page 578 in Section 7.5.2.