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

Last updated on August 29, 2021 at 05:26pm

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 $d(x, y) = 0$ if and only if $x = y$.

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: “$(x, y)$” should be “$(\phi(x), \phi(y))$”

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”

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

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

Chapter 3

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

![Figure 3.23b](image)

Figure 3.23b Varying tree size from 1 to 150.

2. Page 141, In Figure 3.16, “width > 3” should be “breadth > 3”:

![Figure 3.16](image)

Figure 3.16 Decision tree model for web robot detection.
3. Page 164, In Figure 3.32, “width > 3” should be “breadth > 3”:

![Decision Tree Diagram]

**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:

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

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:

\[
\text{Weighted accuracy} = \frac{w_1 TP + w_4 TN}{w_1 TP + w_2 FP + w_3 FN + w_4 TN}. \quad (1)
\]
The relationship between weighted accuracy and other performance measures is summarized in the following table:

<table>
<thead>
<tr>
<th>Measure</th>
<th>$w_1$</th>
<th>$w_2$</th>
<th>$w_3$</th>
<th>$w_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recall</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Precision</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$F_\beta$</td>
<td>$\beta^2 + 1$</td>
<td>$\beta^2$</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Accuracy</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>