Errata and Suggestions of 3rd edition of the Book

Please provide suggestions and make any errors you found in the book. Thanks!

Chapters 1-10.
Chapter 11. Advanced Cluster Analysis

The following is from Hsien-Ting (Tim) Cheng:

page 548: Equation (11.31): the subscript for the summation \( x \in I(a) \) --> \( x \in I(u) \)

the subscript for the summation \( y \in I(b) \) --> \( y \in I(v) \)

page 552: 2nd paragraph, line 3

"If a vertex \( v \) is in the \( \epsilon \)-neighborhood of a core \( u \), then \( v \) is assigned to the same cluster as \( v \)"

--> "...... \( v \) is assigned to the same cluster as \( u \)"

This is from Nikita Spirin:

Below are errors I found during the last rereading of chapters 11, 12

chapter 11

page 515 - group truth => ground truth
page 527 - ) => delete it
page 532 - in formulas 11.10 and 11.11 one should add \( \frac{1}{k} \)
because otherwise it is an unnormalized density
for parameter tuning it is enough but mathematically not correct - if we refer to it as probability
page 535 - in formulas 11.14 and 11.15 don't we have to divide by \( N \) number of objects.

in this case update for \( \mu \) is sum over all objects => but we have to normalize it.
May be the notation is slightly different than what I am get used to?!

page 529 - if \( wij > 0 \) => if \( wij > 0 \), fix indecies of a weight
page 533 - maximize the sum of squared error => minimizes
page 533 - same paragraph - or the expected likelihood => maximizes (here it must be inserted before "the")

page 538 right before Subspace Clustering Methods - n choose k (rotate the binomial coefficient elements)
equality isn't absolutely correct, it is better to replace to asymptotic growth

page 552 - web spams => web spam
page 556 - are related to => are connected to
page 560 - wi can be reached from wi1 => directly reached
page 561 - can be directly reachable => can be directly reachED

Chapter 12. Outlier Detection

Nate Dykens:
pg 06: Change "than such behavior" to "then such behavior".
pg 21: Grammar "this means that means that"
pg 32: Change "AllEletronics" to "AllElectronics"
pg 33: Change "objects with lable "normal"" to "objects with label "normal""
pg 33: Change "declare all objets" to "declare all objects"

This is from Nikita Spirin:

Chapter 12

page 583 - \( = 2.29 = 1.51 \) this must be an approxiamte equality

Best, Nikita

On a page 595 on LOF algorithm and in the original paper page 5
The concept of min-reach-dist is defined. In my opinion definition isn't correct. Because we take \( \min(\max(\text{dist})) \) it will be equivalent to just \( \text{reach-dist} = \max(\text{dist in N_k(0)}) \).

May be in the definition of min-reach-dist we need \( \min(\text{dist}) \) - at least it is in agreement with the illustrations for this definition?

Chapter 13. Trends and Research Frontiers in Data Mining