Data science bootcamp:

Web lessons:

Lesson 1: Introduction to Big Data, Data Science and Predictive Analytics

* We won’t learn much of data engineering, more focus on data science
* More focus on best practice then covering as many topics as possible
* You can make Hadoop cluster in 2-3 hours!
* Mahout is easy if you know R and Hive

Comments:

To do:

1. Learn R! Practice it on different examples.
2. Find a course R – machine learning = similar to the one of Python I did via Youtube.
3. Go through GIT course
4. Make list of questions for the teachers!
5. NoSQL vs SQL, sta je razlika?
6. Google: Hive!

Questions for teachers:

1. Advantage disadvantage of Azure vs. AWS vs. Google
2. Can you put R or Python code in Azure or AWS?
3. What would be the proposal for my problem? Multiclass classification all vs all, or one vs all? And then what with the rest?
4. How do I make 100% True positive classifier?

Lesson 2: Data mining fundamentals

* Interesting names for datasets, wonder if they are useful to know in the real life setting.
* Not sure about interval and ratio data. Also do not understand Graph data sets.

Comments:

To do:

Questions for teachers:

1. How do we overcome missing data problem?
2. Can I exclude outliers first? How do I clean noise?
3. Is there a rule for number of attributes for the number of objects or classes?
4. What does it mean then with the curse of dimensionality,

Lesson 3: Introduction to R

* No string its all character type

Comments:

To do:

1. Read more on factor data type

Questions for teachers: