

Tasklist 1
Due Date: 04.11.2013

Note: In the course of this semester you will receive several Tasklists, consisting of a part that will help you to improve the **Basics** of your programming skills and one that tells you what to **Implement**. The amount of time you need to work through the basics will highly depend on your previous knowledge. If these contents are really new to you, you should spend enough time working through these tasks in order to catch up with your group members whose skills might be more advanced. On the other hand, if you already know most of these contents, feel free to skip the respective lessons or use them as an opportunity to recap and deepen your knowledge of the respective techniques.

Basics

1. **Java:**

Work through Oracle's *Learning the Java Language* tutorial:

<http://docs.oracle.com/javase/tutorial/java/index.html>

2. **Eclipse:**

Lessons 1 – 8 of Mark Dexter's *Eclipse and Java for Total Beginners* tutorial:

<http://eclipsutorial.sourceforge.net/totalbeginner.html>

Lessons 1 – 6 of Mark Dexter's *Using the Eclipse Workbench* tutorial:

<http://eclipsutorial.sourceforge.net/workbench.html>

3. **svn:**

Understand the basic concepts of svn, e.g. by reading through this:

<http://www.eclipse.org/subversive/documentation/index.php>

Implement

1. Install Java, Eclipse and a plugin for Eclipse that supports working with svn (e.g. subclipse or subversive).

2. Create a Java project and check it in to your svn.

3. Look for appropriate data on the *UCI Machine Learning Repository*.

<http://archive.ics.uci.edu/ml/datasets.html>

Think about properties of “good” data for our purpose: What kind of attributes would you prefer (categorical / integer / real)? How many instances and dimensions should it contain? Are there other properties that you could base your choice on?

Ideally you should pick 5 to 10 different datasets.

4. Implement and test functions that read / write datasets in Java. Think about dealing with missing or categorical values.
5. Implement the standard version of **k-Means** aka **Lloyds Method**. Figure out what values you want to keep (e.g. number of steps, final error, movement of centers, running time in ms) and in what format you want to save the output.

Website of the lecture:

<http://www.i1.informatik.uni-bonn.de/de/content/ws-1314-lab-efficient-algorithms-selected-problems-design-analysis-and-implementation>