Applied Data Science – Course Outline

Prerequisites

- Data Science Fundamentals course or (proven) equivalent
- R programming (candidates should take a MOOC if needed before training – audit these two courses: https://www.edx.org/course/introduction-r-data-science-microsoft-dat204x-1 and https://prod-edx-mktg-edit.edx.org/course/programming-r-data-science-microsoft-dat209x-0)

1. Recap of Data science Fundamentals [2 Hours]
    a. What we have learned so far (review of main concepts)
    b. Refreshing examples (in Python, to be revisited in R)
    c. Overview of material and activities for part II
    d. Big data overview (possible later encounter)

2. Using R for data science [3 Hours]
    a. Python is still valid (when to use which)
    b. Overview of the language, packages and development environment
    c. Earlier examples, now in R (ex. using caret package)

3. Back to machine learning algorithms and techniques [10 Hours]
    a. Non-linear regression
    b. Logistic regression
    c. Random forests and boosting/ensemble algorithms
    d. K-Nearest Neighbors (K-NN)
    e. Naïve Bayes
    f. Model optimization, evaluation and assessment, overfitting and curse of dimensionality
    g. Multi model combination

4. Unstructured data analysis techniques and tools [8 Hours]
    a. Theory of text analysis and natural language processing
    b. Speech recognition and processing
    c. Image analysis and computer vision

5. Visualization II [4 Hours]
    a. Multi variable and multi-dimensional visualization
    b. Algorithm convergence and quality
    c. Network and social graph representations
    d. Mapping and spatial data
    e. Visualization best practices (when to use which type)

6. Neural networks and deep learning] [8 Hours]
    a. Basics of neural networks

   b. Deep and reinforcement learning /AI [see also: http://machinelearningmastery.com/deep-learning-courses/]
   c. Intelligent systems and smart agents
   d. Tools and examples (with parallel computing and GPU support if possible)

7. Data science ethics II [2 Hours]
   a. Data privacy, safety and security revisited
   b. Data governance
   c. Social and economic impact of AI agents
   d. Legal and ethical considerations of smart agents / AI

8. Elaborate case studies and techniques (focus on practical local examples) [8 Hours]
   a. Selection of best tools and algorithms for specific problems (cheat sheets and workflows)
   b. Packaging, publishing and automating your model (interface)
   c. Model maintenance, security and monetization (ex. ML platforms and services)
   d. Revisit of big data (tools and technologies – brief)
   e. Text analysis cases (sentiments, topics classification … etc.)
   f. Image and speech applications
   g. Applications from education, labor, health, utilities and services, government, environment, search, image/speech/CV …
   h. Smart cities (smart services in general and IoT)

9. Projects and presentations [3hours]
   a. Form groups of 3-5 people after 2nd meeting
   b. Help them pick cases and find datasets or provide datasets and ask them to analyze)
   c. Include tasks throughout the course for evaluation
   d. Trainees should create a model and evaluate it then present as a group (choice of language and platform)