MA8701: Minutes from the first meeting with the reference group

Time: February 28 at 11.00-11.35
Place: 1236, 12. etg, sentralbygg 2, Gløshagen
Present: Ingeborg Hem, Even Moa Myklebust, Tore Mo-Bjørkelund (on Skype), Mette Langaas (taking the minutes).
All participants have approved the minutes.

1) Reading list - and lectures:
Reading list: The information about the five parts of the course and the reading list is easily found on the wiki pages https://wiki.math.ntnu.no/ma8701/2019v/start
The topics are chosen as "interesting topics within statistical learning and inference that is not covered in other statistics courses", and are not very strongly connected. However, the students find that the Kaggle-avito data set can be seen as the glue in the course – the methods we cover can all be used to solve different aspects of data analysis on the Kaggle-avito data set. However, it would have been even better if all the lecturers could use (parts of) the Kaggle-avito data set to give example of the use of the methods. The students are happy that slides/material are available before (and also after) the lectures, and hope that this will continue. The quality of the lectures is preceived to be good.

2) The compulsory exercise:
It has not been a smooth ride with respect to information about the compulsory exercises. The reason for this is that this is the first time we have compulsory exercises in this course, and also the first time we use Kaggle data sets. First – a) here are some suggestions to how this could have been done differently (with focus on how to do this the next time we give the course), and second – b) suggestions for this year.
a) Clearly present the objectives for the work with the compulsory exercises in the first lecture (=apply method from reading list in order to gain increased understanding of the methods), and give concise requirements for the final presentation (=fit three methods from reading list, not necessarily using all of the training data). Then in the lectures refer to the data set and show how the method lectured be used to solve parts of the problem. Inform students that they may start with analysing data on an standard computer. Wait with presentation of how to perform computations in parallel with GPUs until after the lecture on Part 4 on Deep learning, when this is of prime importance. Be clear about that the use of GPU-based analyses is not the main objective, but is an important concept that student would benefit from knowing about. Make clear that it is useful to know how to analyse text and images. Include a short presentation on simple text analysis with bag of words and the concept of term frequency inverse document frequency. Include information on how to analyses images in practice with deep learning methods.
b) How to proceed this semester: Students must be aware that the primary objective of working with the compulsory exercises is to apply the methods on the reading list for this course, in order to gain increased understanding of the methods. The secondary objective is to be informed about the importance of being able to analyse data in the form of text and images, and how it is possible to run analyses in parallel - with CPUs and GPUs.

3) Plan for the oral presentation of the compulsory exercises:

Mette suggested the following plan, and the members of the reference group were in agreement.

Presentations are (as announced from the start of the course) in week 14 and 15 (April 2 and 9). Groups with 1 or 2 members get 10 minutes presentation, with 3 members 15 minutes and the one group with 4 members get 20 minutes. All group members present for approximate 5 minutes each. After the presentation of the group there are 5 minutes with questions from the coordinator and others in the audience. Times for presentation will be available as soon as all students have answered the survey on when they can present. All students are invited to listen to all presentations. One group has requested to present on April 1 and one student has requested to present on Skype.

What should be the focus of the presentation:
- Short intro to which part of the data is used for training, validation and testing (for the three parts of the project), and which data science preprocessing is used.
- Regularized regression: how used, important assumptions and implications, results.
- Deep learning: how used, important assumptions and implications, results.
- Third method (not necessary from the 5 parts of the reading list): how used, important assumptions and implications, results.
- Summing up

4) Plan for the exam:
The exam is oral, and will be held on May 13 and 14, tentatively with the following starting times.
Monday May 13: 8.30, 9.15, 10.00, 10.45, 11.30, 12.15, 13.00, 15.15, 16.00, 16.45.
Tuesday May 14: 8.30, 9.15, 10.00, 10.45, 11.30, 12.15, 13.00, 13.45, 14.30, 15.15, 16.00. 16.45.
Students are assigned randomly to the time slots (given any constraints).
The students will be asked questions from all parts of the course, and 70/100 score gives the pass limit. Before the exam (a week before?) a set of questions will be available to the students and they can choose to select one question and prepare a short presentation to be given in the start of the exam (maximal duration 10 minutes), either writing on the blackboard or showing a few slides (student can bring notes for this first part). Then the noted must be put aside, and the student gets more questions. The student can choose not to do the 10 minutes presentation and just get questions. The duration of the exam will be approximately 30 minutes.

5) Future changes to the course:
Due to changes in other courses (the new TMA4268 Statistical learning), many of the topics on the reading list for MA8701 in 2017 is this year not on the reading list in 2019. In addition the name of the course is very non-informative and does not reflect the content. This means that in 2019 there is a mismatch between the learning outcomes listed and what should have been the learning outcome. This will be discussed in the statistics group and updated before this course is given in 2021.

6) Activities between the last lecture and the exam:
The reference group investigate with the students what is needed and report back to the coordinator. The reference group will meet after Easter for a final meeting.