Seminar Series in Statistics and Data Science

14.01.2020, 14:15 @ Erling Sverdrups plass, Niels H. Abels hus, 8th floor

Kåre Bævre: Using Machine Learning to establish a Norwegian historical population register

Abstract: An ongoing project seeks to link together historical censuses, church books and vital statistics in order to create a Historical Population Register (HBR) for Norway. The goal is to extend the modern register from its start in 1964 backwards towards 1801, thereby covering all persons that has resided in Norway 1801-1964. This truly unique register will provide detailed micro-level information about central socio-demographic conditions, very rich information on intergenerational processes and full genealogies. The register will open up a lot of new research possibilities in multiple disciplines, such as all social sciences, medicine, genetics, history.

Since all the sources are hand-written, digitalization is a major challenge. In this talk, I will show how we have successfully improved the efficiency of this process by a factor of 30-50 by extensive use of Convolutional Neural Nets and other machine learning techniques. I will emphasize the importance of a well thought out framework for the complete work flow. A major issue has for example been how to establish large collections of training data as cheaply as possible. I suggest an informal boosting-like approach that is very powerful. We have also applied several tricks where we combine approaches to recast a classification problem as a verification problem.

This example of a practical large scale application of machine learning should provide quite a few lessons relevant for the use of such techniques in general, and might suggest some new avenues of research.

Kåre Bævre
Centre for Fertility an Health (CeFH), Norwegian Institute of Public Health

Kåre Bævre is senior researcher at Centre for Fertility an Health (CeFH) at the Norwegian Institute of Public Health. With a background in economics and econometrics his main field of research is historical demography and social epidemiology. For many years he has worked extensively on a project working to establish a Historical Population Register for Norway 1801-1964 and an accompanying project establishing a register of all land properties in Norway 1660-2020.

Next seminar

28.01.2020 @ 14:15 **Kristoffer H. Hellton**
Norwegian Computing Center (NOR)