THE UNIVERSITY OF HONG KONG
DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE

Topics for STAT4798 Statistics and Actuarial Science Project (6 credits)
(Offered in both 1st and 2nd semesters of 2022 - 2023 for STAT4798)

1. Generalizable training algorithms for deep learning based image classification

Optimization algorithms such as stochastic gradient descent (SGD) and Adam play a key role in machine learning methods. Recently, it has been observed that different optimization problems, although can all successfully optimize the training objective function, may have drastically different performances on unseen test data points. For example, many recent works have pointed out that for deep learning based image classification, Adam can minimize the training loss function faster than SGD, but the obtained deep learning classifier has a much larger test error. This motivates us to study generalizable training algorithms for deep learning based image classification.

The target students are undergraduate students with a strong background in deep learning and python (PyTorch/TensorFlow) programming. The students are expected to compare the performances of various existing optimization algorithms in training neural networks, and, ideally, design new generalizable optimization algorithms for deep learning based image classification.

References:

- Wilson, A. C., Roelofs, R., Stern, M., Srebro, N., & Recht, B. (2017). The marginal value of adaptive gradient methods in machine learning. Advances in neural information processing systems, 30.
- Chen, J., Zhou, D., Tang, Y., Yang, Z., Cao, Y., & Gu, Q. (2020, January). Closing the Generalization Gap of Adaptive Gradient Methods in Training Deep Neural Networks. In IJCAI.
- Zou, D., Cao, Y., Li, Y., & Gu, Q. (2021). Understanding the generalization of adam in learning neural networks with proper regularization. arXiv preprint arXiv:2108.11371.

Supervisor: **Dr. Y. Cao,** yuancao@hku.hk, Dept of Statistics and Actuarial Science

2. Dependence Structures in Multiple Life Insurances and Annuities

The price of a multiple-life insurance/annuity product depends not only on the marginal distributions of the underlying future lifetimes, but also on their dependence structure. In this project, the effect of dependence structure on the actuarial present values will be studied. In the course of the research, the student will learn some basic theory of dependence structures.

Supervisor: **Prof. K.C. Cheung**, kccg@hku.hk, Dept of Statistics and Actuarial Science

3. Brain Imaging Analysis with Statistical Learning

Brain imaging analysis has played a central role in understanding the functions of human brain. From Electroencephalography (EEG) to magnetic resonance imaging (MRI) and from MRI to functional MRI (fMRI), the advancement of brain imaging technologies has benefitted tremendously to the diagnosis and treatment of brain disease. In this project, students will learn to develop statistical machine learning models to analyze brain imaging data and make predictions for brain diseases. As imaging data is usually represented as tensors (or multidimensional arrays), tensor decomposition and tensor regression methods would also be studied. Students are expected to have good knowledge in programming languages such as Python or R.

Supervisor: **Dr. L. Feng**, ug_enquiry@saas.hku.hk, Dept of Statistics and Actuarial Science

4. Chronological age prediction based on DNA methylation

(This project will be offered in Semester 1 only.)

Over the years, the correlation between DNA methylation levels and chronological age has been discovered in different species. In this project, we are going to use the 450K Human Methylation Beadchip data for human age prediction. High dimensional variable selection methods and machine learning models will be attempted. Students are expected to have good knowledge in programming languages such as R or Python.

Supervisor: **Prof. Tony W.K. Fung**, wingfung@hku.hk, Dept of Statistics and Actuarial Science

5. Open-world object discovery with deep learning

Deep learning has achieved remarkable success in many tasks, even surpassing humans, for example in image classification. However, the success comes at the cost of intensively labeled data, e.g., ImageNet which contains over 1.2 million manually annotated images. When a trained classification model meets an image from an unseen class, it often mistakenly predicts the image as one of the seen classes with high confidence. In other words, current learning models struggle to handle open-world problems where there are unseen or unfamiliar objects. In this project, the students will study the open-world object discovery problem with deep learning and develop solutions to enable the model to deal with unseen or unfamiliar objects.

Requirement: Knowledge and hands-on experience in computer vision and deep learning; familiar with Python; preferably also familiar with PyTorch/TensorFlow/JAX.

Supervisor: **Dr. K. Han**, kaihanx@hku.hk, Dept of Statistics and Actuarial Science

6. Effective self-supervised learning with large-scale unlabeled data

The success of modern machine learning techniques is driven by large-scale datasets with human annotations. However, it is not possible to annotate a large-scale dataset for all possible tasks. Some tasks may require domain-specific expertise and there is no large-scale data available, for example, medical images for a rare disease. Self-supervised learning, which requires no human annotations, appears to be an intriguing direction. It aims at learning useful representations in an unsupervised manner, which can be effectively used for various downstream tasks like object recognition, detection, and segmentation in visual data. In this project, the students will study various self-supervised deep representation learning techniques and develop solutions for effective self-supervised learning with large-scale real-world unlabeled data.

Requirement: Knowledge and hands-on experience in computer vision and deep learning; familiar with Python; preferably also familiar with PyTorch/TensorFlow/JAX.

Supervisor: **Dr. K. Han**, kaihanx@hku.hk, Dept of Statistics and Actuarial Science

7. Risk factors of distress among Chinese caregivers

Informal caregivers are invaluable partners of the health care system. However, their caring responsibilities often affect their psychological wellbeing and ability to continue in their role. This study examines the effects of various risk factors on caregivers' psychological distress who were the primary caregivers of frail older adults living in the community in Hong Kong. Traditional logistic regression model and deep learning method are considered.

Requirement: Knowledge in Python.
Knowledge in logistic regression and ANN.

Supervisor: **Dr. C.W. Kwan**, cwkwan@hku.hk, Dept of Statistics and Actuarial Science

8. Analysis of correlated zero-inflated count data

In many medical and public health investigations, the count data encountered often exhibit an excess of zeros, and very frequently this type of data are collected on clusters of subjects or by repeated measurements on each subject. For example, in the analysis of medical expenditure, members in the same family may exhibit some correlation possibly due to housing locality, genetic predisposition, similar dietary and living habit. Ignoring such correlation may lead to misleading statistical inference. This project will survey the models and methods in the literature and apply them to a real data set.

Requirement: Knowledge in R or Python.

Supervisor: **Dr. Eddy K.F. Lam**, hrntlkf@hku.hk, Dept of Statistics and Actuarial Science

9. Building a mixed reality e-learning platform and content (Local schools, SAAS and Data Science Lab - Mixed Reality and AI in E-learning)

This project aims to develop a platform for implementing the mixed reality e-learning content. The e-learning content include mathematics/statistics/AI, English, and/or chemistry, etc. Students will implement some mixed reality e-learning content and platform, and will implement a taxonomy for classifying, grouping and relating the e-learning goals with quality matter metrics. Students will learn how to implement quality matter metrics, a mixed-reality e-learning content, different VR/AR/MR tools and software programming. Students who have basic knowledge in statistics, AI, machine learning, text analysis are preferred, and willing to learning new mixed reality e-learning tools.

Supervisor: **Dr. Adela S.M. Lau**, adelalau@hku.hk, Dept of Statistics and Actuarial Science

10. Data Visualization in Metaverse (NSC and Data Science Lab –Mixed Reality and Data Visualisation)

Metaverse is extremely hot and people is trying to do everything/anything is Metaverse. The aim of this project is to provide data visualization, manipulation, reporting, and to enable people to "touch" the data, and "seeing" the data visualization in stereo mode. Students will develop the stereo visualization of data such as charts (bar chart, line chart, bubble cloud, etc), tables, graphs, maps, infographics, dashboard, to implement an interactive UI to manipulate the data visualization such as selecting data to visual, overlaying difference visualization, and combining data (structure & unstructured), to implement a dashboard that focuses on the element through touching it, and the flow of usage. Students will learn the mixed reality technology and some popular reporting tools (e.g. QlikView, Tableau, etc.), and learn how to reuse/import the visualization/dashboard, and have ability for users collaborating the visualization in metaverse. Students who have basic knowledge in statistics and willing to learn data visualization tools are preferred.

Supervisor: **Dr. Adela S.M. Lau**, adelalau@hku.hk, Dept of Statistics and Actuarial Science

11. Building a social media analytics platform for social science, healthcare, criminal justice, financial analytics, and marketing research (Company and Data Science Lab – AI in Social Media Analytics)

This project aims to develop different social media crawler and social media analytics for social science, healthcare, criminal justice, and insurance marketing research. Students will learn how to implement different types of crawlers and its challenges, learn how to implement an ontology (a taxonomy of words with semantic meaning), and perform text analysis and network analysis. Students will learn how to crawl data from internet and different open source AI tools for NLP and text analysis, and learn how to develop a knowledge base. Students who have basic knowledge in statistics, AI, machine learning, text analysis are preferred, and have a minor in computer science are taken an advantage.

Supervisor: **Dr. Adela S.M. Lau**, adelalau@hku.hk, Dept of Statistics and Actuarial Science

12. ESG Analytics (NSC and Data Science Lab – Big Data Analytics for Financial Intelligence)

Environmental, Social and Governance (ESG) has been a growing focus for businesses and investors around the world. This project aims to develop new ESG-focused solutions through ESG data and report analysis to help businesses navigate the evolving landscape. Students will develop a crawler to crawl the structured and unstructured ESG data (i.e. ratings, reports, standards, policy, etc), develop a data model for ESG product/services taxonomy database (i.e. a set of keywords), use NLP/ ML technique to analyze the unstructured data in ESG reports, use NLP/ Graph analytics to perform ESG verification on vendors' data/ reports/information. Students will learn how to crawl data from internet and different open source AI tools for NLP and text analysis. Students who have basic knowledge in statistics, AI, and machine learning are preferred.

Supervisor: **Dr. Adela S.M. Lau**, adelalau@hku.hk, Dept of Statistics and Actuarial Science

13. Enterprise Instant Messaging Platform in Banking/Insurance Setting (NSC and Data Science Lab – Text and Emotion Analytics for Sales and Marketing)

This research aims to develop add-on functions for Enterprise Instant Messaging Platform to get actionable insight to assist the business users in their daily business in a Banking/Insurance setting. Students will develop an algorithm to extract key sales information in conversation based on users' conversation (i.e. text/emoticon), develop an algorithm to provide best action recommendation (e.g. recommend suitable product/ service), develop an algorithm of sentiment analysis and churn prediction based on conversation log, develop an algorithm to match the business user/agent to customer based on personality analysis, usage pattern, products interests, etc., and/or combine and analyze customer data and external data to create customer analytical record . Students will learn some text analysis, emotion analysis and/or video/image analysis skills and open source tools in this project. Students who have basic knowledge in statistics, AI, machine learning, text analysis are preferred. Image/video analytics is optional to added in the study scope.

Supervisor: **Dr. Adela S.M. Lau**, adelalau@hku.hk, Dept of Statistics and Actuarial Science

14. Human behaviours Analytics for Online Interview and Business Video Communication (NSC and Data Science Lab – AI in Speech/Voice/Image/Video Analytics)

The research aims to develop AI add-on on the common video communication platform (Microsoft Teams, Zoom, Google Meet, etc). The primary use case target for online interviews, but the developed assets can be adopted in other business scenarios such as staff training, customer services, etc. Students will develop a model with the capability of ingest/digest the video, picture, sound file, and text file, develop a model to detect emotion, personality, attitude, develop the model to identify lying, develop a model to identify which part of the conversation/video/text representing which emotion, personality, attitude and lies telling. Students will learn some open source tools for audio/speech/image/text analysis. Students who have basic knowledge in statistics, AI, and machine learning are preferred, have knowledge in image analysis will take an advantage.

Supervisor: **Dr. Adela S.M. Lau**, adelalau@hku.hk, Dept of Statistics and Actuarial Science

15. Video Analytics for Smart Retail Intelligence (NSC and Data Science Lab –Video/Image Analytics for Smart Retail Stores)

AI on video (or VA, video analytics) has been applied in many areas like public safety, transport, healthcare, and retails. Whereas some of the use case in smart retail, the goal is clear but the result is yet promising. This research aims to increase the accuracy of use case by leveraging VA and IOT technology. Students will develop the models to measure how often the same customer revisit the store, how long individual customer stays in store, at which location, and the demographic including age and gender, to identify the way to measure precisely on face recognition by camera, and people location tracking by 3D sensor. Students will learn some open source tools for object identification and classification. Students who have basic knowledge in statistics, AI, and machine learning are preferred, have knowledge in image analysis will take an advantage.

Supervisor: **Dr. Adela S.M. Lau**, adelalau@hku.hk, Dept of Statistics and Actuarial Science

16. Contract document conformity checking (NSC and Data Science Lab – Image Analytics for FinTech)

Contract document conformity checking is a critical process in business world, both for Banks and the commercial clients. There are numerous types of forms, document and cheques. The current processes are performed manually with little automation. This project aims to develop an AI solution that perform necessary checks automatically in with latest AI algorithms. Students will develop an algorithm to detect different kinds of hand-made edits of the document (i.e. overwrites, modification, cancellations or addition), develop an algorithm to detect and recognize stamps, logo and flagged and unflagged checkbox of documents, develop an algorithm to detect and verify the signature are signed by the same person or more different people of documents and cheques, and develop an algorithm to verify the cheque signature are consistent within document and specimen. Students will learn some open source tools for object identification and classification. Students who have basic knowledge in statistics, AI, and machine learning are preferred, have knowledge in image analysis will take an advantage.

Supervisor: **Dr. Adela S.M. Lau**, adelalau@hku.hk, Dept of Statistics and Actuarial Science

17. Applications of Extreme Value Models

Extreme value theory concerns the behaviour of maxima or minima, and has been used extensively in areas such as finance, hydrology, engineering and meteorology where the occurrence of extremes may have catastrophic consequences. In this project, the student will learn the basic modelling techniques for data of extremes and will apply such models to data sets of practical interest. The emphasis is on conceptual understanding of the underlying theory and interpretation of the fitted models.

Requirement: The student should be competent in computer programming. Knowledge in or willingness to learn the R programming language is essential.

Supervisor: **Dr. David Lee**, leedav@hku.hk, Dept of Statistics and Actuarial Science

18. Resampling Methods for Regression

Recent years have found increasing use of resampling methods in regression studies. Examples include the paired bootstrap, the residual bootstrap, the wild bootstrap, random perturbation, bagging, etc. In this project we explore their potential applications in contemporary regression settings where statistical inference remains prohibitively difficult.

Supervisor: **Prof. Stephen M.S. Lee**, smslee@hku.hk, Dept of Statistics and Actuarial Science

19. Applications of Secure Blockchain Solution

In this project we begin with a review of the basic architecture for blockchain in Python. This includes state transition rules, method for creating blocks, mechanisms for checking the validity of transactions, blocks, and the full chain. Next, we will create new blocks from data, validate the new blocks and add them to the existing blockchain.

Security is of the utmost importance in any blockchain architecture, in this project we will discuss 3 popular verification methods: public key cryptography, digital signature algorithm and trusted time-stamping. Finally, we will construct practical blockchain solutions to current fintech problems.

Supervisor: **Dr. Eric A.L. Li**, ericli11@hku.hk, Dept of Statistics and Actuarial Science

20. Introduction to Quantum Computing Algorithms

First we begin with a basic understanding of quantum computing (QC). Then we move on to some popular QC algorithms, written in Javascript and Python. In addition to constructing these QC codes, we will also provide the meanings, purposes and theoretical bases of these QC codes.

The QC algorithms we will cover include: Deutsch-Jozsa Algorithm, Simon's Algorithm, Super Dense Coding, Period Finding, and Shor's Factoring Algorithm. The last one is particularly important in modern cryptography: given an integer which is a product of two distinct prime numbers, this algorithm finds one of its prime factors.

Supervisor: **Dr. Eric A.L. Li**, ericli11@hku.hk, Dept of Statistics and Actuarial Science

21. Statistical Inference for Tensor Data

Tensors have been used in many fields and have provided powerful applications in various practical domains. They generalize vectors and matrices and have been studied from different viewpoints. The study of tensor methods has a long history in statistics. In the era of big data, tensor data appear frequently in the forms of video data, spatio-temporal expression data, relationship data in recommending and mining, and latent variable models, from a vast range of statistical applications. However, the extension of methods for dealing with matrices to tensors is much more difficult than those from vectors to matrices. This project targets to several tensor-based statistical methods.

Supervisor: **Prof. G. Li**, gdli@hku.hk, Dept of Statistics and Actuarial Science

22. Cointegration in Financial Analysis

The goal of this project is to test cointegration in financial time series. Students are required to have basic understanding of cointegration and some knowledge of computer programming.

Supervisor: **Dr. C. Wang**, stacw@hku.hk, Dept of Statistics and Actuarial Science

23. A Simulation Study on Risk Process with Dependence Modelling

In the recent paper by Cossette, et al. (2019), the traditional collective risk model in actuarial science was extended to include elements of dependence structure. Three different dependence models were investigated: (i) collective risk model with multivariate mixed Erlang distributions; (ii) collective risk model with mixing; and (iii) collective risk models with multivariate copulas.

This project aims to explore the modelling of different dependence structures in the aggregate claim process and to study the effects of the dependence on the aggregate loss through different risk measures.

Students taking this project are expected to have fundamental knowledge in risk theory and programming skills for extensive simulation study.

This project is only offered in the First Semester.

References:

- Cossette, H., Marceau, E., and Mtalai, I. (2019). Collective Risk Models with Dependence. *Insurance: Mathematics and Economics*, 87, 153-168. (full paper accessible online from HKUL by HKU students)
- Klugman, S. A., Panjer, H. H. and Willmot, G. E. (2019). *Loss Models: From Data to Decisions (5th Edition)*. Wiley.
- Kaas, R., Goovaerts, M., Dhaene, J., and Denuit, M. (2008). *Modern Actuarial Risk Theory: Using R (2nd Edition)*. Springer.
- Joe, H. (2014). *Dependence Modeling with Copulas*. CRC Press.

Supervisor: **Dr. K.P. Wat**, watkp@hku.hk, Dept of Statistics and Actuarial Science

24. A Simulation Study on an Extended Optimal Approach to Bühlmann Credibility Theory

In the recent paper by Yan and Song (2022), the classical Bühlmann credibility theory was extended to include non-linear Bayesian credibility estimators. Various numerical examples and simulation studies were performed in their work.

This project aims to study the benefits and effectiveness of this framework by establishing further examples and numerical studies. Monte Carlo experiments are to be performed to assess the performance of the proposed method in finite sample cases. Emphasis can be put on the examples for heavy-tailed excess claims distributions. Students taking this project are expected to have fundamental knowledge in credibility theory and programming skills for extensive simulation study.

This project is only offered in the Second Semester.

Requirement: Pass in STAT3908

References:

- Yan, Y. and Song, K.-S. (2022). A General Optimal Approach to Bühlmann Credibility Theory. *Insurance: Mathematics and Economics*, 104, 262-282. (full paper accessible online from HKUL by HKU students)
- Bühlmann, H. and Gisler, A. (2005). *A Course in Credibility Theory and its Applications*. Springer.

Supervisor: **Dr. K.P. Wat**, watkp@hku.hk, Dept of Statistics and Actuarial Science

25. Reserve Calculation under Stochastic Interest Rate Models

Insurance companies need to have enough reserve all the time to pay the claims. Net Premium Valuation is the preliminary approach to calibrate actuarial liabilities, but it is increasingly replaced by Gross Premium Valuation due to some major disadvantages. Also the reserve depends on the mortality risk and the investment risk. Due to the long term nature of insurance products, the mortality and the interest rate are changing when the economy and/or other environments change. In the life contingencies courses, you always assume the interest is a constant. In this project, we consider the stochastic interest effect. The students should pass the life contingencies courses when they take this project.

Supervisor: **Prof. H. Yang**, hlyang@hku.hk, Dept of Statistics & Actuarial Science

26. Deep Learning with Application in Artificial Intelligence

This project will focus on extracting useful information from structured and unstructured data and formulating statistical models for inference and prediction. In particular, we will develop deep learning, including deep neural networks for imaging analysis and computer vision and natural language processing for text data analysis. Extensive computation will be needed and real data will be used for analysis and illustration.

Requirement: The student needs to have experience with Python and R programming.

Supervisor: **Prof. G.S. Yin**, gyin@hku.hk, Dept of Statistics and Actuarial Science

27. Novel SIR modeling for Omicron transmission

The COVID-19 new variant Omicron is spreading out in Hong Kong at an unprecedented rate. The traditional SIR (susceptible, infected, removed) model cannot fit the Omicron data adequately, which results in poor prediction of numbers of infection and asymptomatic cases. The key is to estimate the number of asymptomatic cases who are still transmissible to others, for while they are not aware of. The SIR model was developed about 100 years ago, and now it is the time to reformulate the problem and develop a better model to accommodate complex scenarios, e.g., the population is not closed and fixed, infected individuals can be re-infected, immunized individuals can also be infected, as well as strict social distance and restrictions.

Supervisor: **Prof. G.S. Yin**, gyin@hku.hk, Dept of Statistics and Actuarial Science

28. Meta analysis on COVID-19 vaccines: effectiveness and side-effects

A number of different types of COVID-19 vaccines have been developed and in use worldwide. These include the traditional inactivated virus type of vaccine and mRNA vaccine. The COVID-19 related deaths have been surging in Hong Kong, particularly in the elderly people. A thorough investigation on the vaccine effectiveness would provide a guide on what type of vaccine should be given. Comparisons with other countries or districts should be also included for a comprehensive picture and devising a better strategy for fighting against Omicron.

Supervisor: **Prof. G.S. Yin**, gyin@hku.hk, Dept of Statistics and Actuarial Science

29. Random survey on COVID-19 in Hong Kong

To better understand the views of Hong Kong people on COVID-19, random survey provides an effective tool for statistical analysis. Critical opinions include use of vaccine, lock down of the city, strict rules on quarantine, estimation of infections, travel bans, strategies on better handling the surge of the death rate etc. A comprehensive survey can be designed to answer many of such questions. The survey could be web-based, telephone-based or face-to-face interviews for collecting the data. AI-based NLP techniques can also help to collect information from the internet for sentiment analysis and psychological stress and mental anxiety due to COVID-19.

Supervisor: **Prof. G.S. Yin**, gyin@hku.hk, Dept of Statistics and Actuarial Science

30. Blockchain and cryptocurrency analysis and their future

More cryptocurrencies have been created on the daily basis. The underlying technology, blockchain, has many other potentials and applications, including smart contract, NFT etc. Unlike stocks which has a fundamental company as the basis, cryptocurrencies have no earning report (ER), no underlying product, etc. Although many countries are banning cryptocurrencies, they will stay and live with us. Rather than avoid them which you cannot, it is better to understand the trend, applications, and future of this sector. Comprehensive analysis of cryptocurrencies and its derivatives are needed, including forecasting, novel application of blockchain, better strategies for trading.

Supervisor: **Prof. G.S. Yin**, gyin@hku.hk, Dept of Statistics and Actuarial Science

31. Reliable machine learning methods with application in healthcare

High-stakes decision-making in areas like healthcare, finance and governance requires accountability for decisions and for how data is used in making decisions. Many concerns have been raised about whether machine learning (ML) models can meet these expectations. ML models are often complex black-boxes and thus have varying, unknown failure modes that are revealed only after deployment: models fail to achieve the reported high accuracies, lead to unfair decisions, and sometimes provide predictions that are plain unacceptable given basic domain knowledge.

This project will study and explore reliable machine learning technology with regard of model stability, fairness and explanation. The application of such technology in healthcare domain (such as medical image analysis) will be analysis and illustration.

Requirement: The student needs to have experience with Python programming and be familiar with basic machine learning/deep learning technique.

Supervisor: **Dr. L. Yu**, lqyu@hku.hk, Dept of Statistics and Actuarial Science

32. Multi-modal machine learning methods with application in healthcare

Multi-modal biomedical data, such as medical images, EHRs and genomic data, are often jointly used for clinical decision marking for the sake of complementary characteristics of different data modality. Utilizing the relationship between different modality can alleviate the data insufficiency in single modality and improve the analysis performance.

This project will study and explore machine learning/deep learning-based multi-modal techniques and demonstrate its applications in healthcare domain by analysing image, text, or even genomic data.

Requirement: The student needs to have experience with Python programming and be familiar with basic machine learning/deep learning.

Supervisor: **Dr. L. Yu**, lqyu@hku.hk, Dept of Statistics and Actuarial Science

33. Applications of Graph Neural Networks

Graphs are all around us; real world objects are often defined in terms of their connections to other things. A set of objects, and the connections between them, are naturally expressed as a graph. Graph Neural Networks (GNNs) are a class of deep learning methods designed to perform inference on data described by graphs. It can be directly applied to graphs and provide an easy way to do node-level, edge-level, and graph-level prediction tasks.

This project will study and explore GNN methods and demonstrate its applications in biomedical data analysis, drug discovery, or natural language processing.

Requirement: The student needs to have experience with Python programming and be familiar with basic machine learning/deep learning.

Supervisor: **Dr. L. Yu**, lqyu@hku.hk, Dept of Statistics and Actuarial Science

34. Optimality Studies with Dependent Risks

Due to the complexity of modern insurance and financial products, contemporary insurance risk models have taken many realistic features into consideration. In the actuarial literature, the incorporation of realistic features such as dividends, investment and reinsurance into the basic insurance risk process has generated a lot of interesting research on optimality in the past two decades. This project aims at studying optimal dividends, investment and/or reinsurance for an insurance risk models with dependent risks.

Supervisor: **Prof. K.C. Yuen**, kcyuen@hku.hk, Dept of Statistics and Actuarial Science

35. Statistical Modelling for Biological/Medical Data

In this project, the students will implement statistical methods to analyse real biological/medical data set to understand/interpret biology/disease etiology. Statical methods include Bayesian methods, variable selection, network analysis, etc.

Requirement: Students need to know at least one programming language (such as R, Python, etc) and basic data analysis skills.

Supervisor: **Dr. Dora Y. Zhang**, doraz@hku.hk, Dept of Statistics and Actuarial Science\

36. Multiple Output Online Non-stationary GPs

The goal of this project is to implement an online algorithm for multiple output Gaussian processes. The student will extend a Sequential Monte Carlo sampler for online Gaussian processes by writing a linear co-regionalization kernel to model multiple time series signals. Possible applications include medical settings or financial settings. Strong programming ability in Python and prior experience in Bayesian inference is required.

Supervisor: **Dr. Michael M.Y. Zhang,** mzhang18@hku.hk, Dept of Statistics and Actuarial Science

37. Online Spectral Mixture Kernel

The goal of this project is to implement a method to estimate the parameters in the flexible "Spectral Mixture Kernel" in an online setting using a Sequential Monte Carlo algorithm. Applications of this method include medical or financial settings. Strong programming ability in Python and prior experience in Bayesian inference is required.

Supervisor: **Dr. Michael M.Y. Zhang,** mzhang18@hku.hk, Dept of Statistics and Actuarial Science

38. Online Student-t Process Algorithm

The goal of this project is to implement an online inference algorithm to learn a heavy tailed Student-t process for time series analysis. Strong programming ability in Python and prior experience in Bayesian inference is required.

Supervisor: **Dr. Michael M.Y. Zhang,** mzhang18@hku.hk, Dept of Statistics and Actuarial Science

39. Non-linear Network Embedding

The goal of this project is to model relational data as a non-linear decomposition of a lower dimensional representation of the relations between observations. Strong programming ability in Python and prior experience in Bayesian inference is required.

Supervisor: **Dr. Michael M.Y. Zhang,** mzhang18@hku.hk, Dept of Statistics and Actuarial Science

40. A Bayesian Hypothesis Testing Approach for Generative Adversarial Networks

This project involves combining the popular Generative Adversarial Network with various forms of Bayesian hypothesis testing. If successful, the Bayesian hypothesis testing GAN could have stronger classification abilities and could possible reduce the risk of mode collapse. Prior knowledge of deep learning and strong programming ability in Python and deep learning packages like PyTorch, Tensorflow or Keras are required.

Supervisor: **Dr. Michael M.Y. Zhang,** mzhang18@hku.hk, Dept of Statistics and Actuarial Science

41. Forecasting Time Series: with Application to Stocks Trading

This project aims to forecast forward behavior of stock prices using neural networks. Simulated trading strategies based on the forecast results are also required.

Requirement: Knowledge of course STAT3612 or STAT8017, AI/machine learning/deep learning, and skills in statistical programming using either SAS, R, or C++.

Supervisor: **Dr. Z. Zhang**, zhangz08@hku.hk, Dept of Statistics and Actuarial Science

42. Financial data analysis

This project aims to analyze the financial data by using the time series models, causal semantics, or machine learning techniques. Students are expected to use these methodologies to analyze real data sets, and develop useful trading algorithms.

Requirement: At least one programming language and knowledge about financial time series analysis

Supervisor: **Dr. K. Zhu**, mazhuke@hku.hk, Dept of Statistics and Actuarial Science

********** END **********