統計與機器學習概論一

Introduction of Statistics and Machine Learning (I)

Course Description:

統計與機器學習概論一 (Introduction of Statistics and Machine Learning (I)) is designed for machine learning beginners where many examples (but not all) are made for students with biomedical background. The class lays a sufficient mathematical foundation for its sequel class 統計與機器學習概論二 (Introduction of Statistics and Machine Learning (II)) that we recommend against students to take if MLI or similar has never been taken. These relevant mathematical concept and details include but not limited to relation between statistics and probability, concept of and inference from sampling, distributions, linear algebra, linear transformation, regression and correlation, Goodness-of-fit test, discriminant analysis, training methods, maximum likelihood and Bayesian parameter estimation, and ODEs applied for biomedical research and Machine Learning Basics. Python workshop is provided along with the class where the successful completion of homework in the class requires programming skills.

Teachers: 張筱涵、楊立威、羅中泉老師 (Profs HH Chang, LW Yang, CC Lo)

Textbooks:
1. Biological Sequence Analysis Probabilistic Models of Proteins and Nucleic Acids (@NTHU library)

2. Michael C. Whitlock and Dolph Schluter. 2014. *The Analysis of Biological Data (Second Edition)*. Roberts and Company Publishers, Greenwood Village, Colorado. (ISBN-10: 1936221489) https://www.amazon.com/Analysis-Biological-Data-Michael-Whitlock/dp/0981519407 (@NTHU library)

3. Peter Dayan & L. F. Abbott. Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems. The MIT Press (Physical and electronic copies available @ NTHU library)

Additional resources (額外學習資源之內容):
1. Statistics 110 from Harvard University (https://projects.iq.harvard.edu/stat110/home)

Course Outline:

Date	Topic	Instructor
9/3	Matrix and linear algebra	CC Lo
9/10	Data types, sampling biases and rules, inference from samples I	HH Chang
9/10	Exam – 40 minutes	CC Lo/ LW Yang
9/17	Mid-Autumn Festival (no class)	
9/24	Inference from samples II	HH Chang

10/1	Analyzing categorical data: goodness-of-fit test, contingency analysis	HH Chang
10/8	Correlation and Regression I	HH Chang
10/15	Correlation and Regression II	HH Chang
10/22	From the things you already knew (to some extent) – mean, SD, variance, how is statistics related to probability? Hypothesis testing. Why does statistics/probability constitute the basics of machine learning (examples on drug development etc)?	LW Yang
10/29	Read the math formula – index, difference between probability (discrete function) and probability density (continuous function), Comparing two means (doing it with your MS Excel sheet), learning from playing dynamics programming games to align two biological sequences	LW Yang
11/5	Distributions Rules Probability addition/multiplication, dependency, conditional probability, marginal probability and Bayes' theorem	LW Yang
11/12	Build your first probabilistic model - a classifier, (Relative) Entropy, Information Content, “Distance” between Distributions, Boltzmann Relation	LW Yang
11/19	Normalization – frequency vs proportion, normalization by controls, by ranking, by probability	LW Yang
11/26	Models of neurons and synapses	CC Lo
12/3	Models of synaptic plasticity and memory	CC Lo
12/10	Stability and dynamical system theory	CC Lo
12/17	Learning and decision making in biological neural networks	CC Lo

Grading:
Yang: two computational homework (15% each – Protein Sequence Aligner & CpG island predictor using log-odd values)
Chang: homework (30%)
Lo: homework (30%)
Quiz (5%)
Class participation (5%)

With programming workshop:
Python programming taught in 10 weeks + one session of basic Linux

Course policy for AI usage: Conditionally open; please specify how to utilize generative AI in course output.

Grounded in the principles of transparency and responsibility, this course encourages students to leverage AI for collaboration and mutual learning to enhance the quality of course outputs. In accordance with the published Guidelines for Collaboration, Co-learning, and Cultivation of Artificial Intelligence Competencies in University Education, this course adopts the following policy:

- Conditionally open - Students must briefly explain how generative AI was used for topic

ideation, sentence refinement, or structural reference in the footnotes of the title page or after the reference in their assignments or reports. If usage is discovered without proper disclosure, instructors, the institution, or relevant units (e.g. our class) have the right to reevaluate the assignment or report or withhold scores.

- If the course materials or learning resources have been derived from generative AI, the instructor will also indicate this in the slides or orally.
- Students enrolled in this course agree to the above ethics statement if registering for the class.