COURSE SYLLABUS
CSC15004 – Statistical Machine Learning

1. GENERAL INFORMATION

Course name:	Statistical Machine learning
Course name (in Vietnamese):	Học thống kê
Course ID:	CSC15004
Knowledge block:	Major
Number of credits:	4
Credit hours for theory:	45
Credit hours for practice:	30
Credit hours for self-study:	90
Prerequisite:	Discrete mathematics, Data structure and algorithm, Programming
Prior-course:	Introduction to Machine learning
Instructors:	

2. COURSE DESCRIPTION

The course is designed to introduce basic methods in machine learning models using probability and statistics. This course focuses students on applying knowledge in probability and statistics in learning machine learning models like linear regression, neural network, hidden Markov, support vector machine and decision tree. Students will learn these models, implement and apply them to solving some simple problems in pattern recognition and data analysis. This course will also require students to practice programming skills in order to build demonstration program, skills in collecting and standardize data and analyzing outcome models.

3. COURSE GOALS

At the end of the course, students are able to

ID	Description	Program LOs
G1	Employ and apply reading, writing and presentation skills	2.1, 2.2
G2	Understand concepts in statistics and machine learning in data analysis	1.1.1, 1.2.2, 1.3.6, 2.1, 2.2
G3	Apply statistical models (single and multivariate linear regression) to data analysis	1.1.1, 1.3.6, 2.1, 4.1, 4.2
G4	Apply machine models (neural network, hidden Markov, support vector machine, decision tree) to data modeling	1.1.1, 1.3.6, 2.1, 4.1, 4.2
G5	Solve practical problems in data analysis	1.1.1, 1.3.6, 2.1, 4.1, 4.2

4. COURSE OUTCOMES

CO	Description	I/T/U
G1.1	Able to apply reading, writing and presentation skills	U
G1.2	Able to build, organize and manage team	U
G1.3	Able to present ideas, discuss and debate in course topics	T
G2.1	Understand concepts in statistics and machine learning in data analysis	T
G3.1	Deeply understand multivariate linear regression model	I
G3.2	Able to build multivariate linear regression model from a given dataset	TU
G3.3	Able to explain and apply multivariate linear regression model to data inference	TU
G4.1	Able to explain and apply neural network model to learning patterns from simple dataset	TU
G4.2	Able to explain and apply hidden Markov model to learning time series data	TU

G4.3	Able to explain and apply support vector machine to simple classification problems	TU
G4.4	Able to explain and apply decision tree model to simple rule-based problems	TU
G5.1	Able to solve simple problems in data analysis	U

5. TEACHING PLAN

THEORY

ID	Topic	Course outcomes	Teaching/Learning Activities (samples)	Assessments
1	Introduction to Machine learning	G1.2, G1.3	Lecturing	
2	Review of statistics	G1.2, G1.3, G2.1	Lecturing, discussion	
3	Review of statistics	G1.3, G2.1	Lecturing, discussion	
4	Decision tree	G1.2, G1.3, G4.4	Lecturing, discussion	
5	Linear regression	G1.2, G1.3, G3.1, G5.1	Lecturing, discussion	
6	Multivariate linear regression	G1.2, G1.3, G3.1, G3.2, G3.3, G5.1	Lecturing, discussion	
7	Dimensionality reduction	G1.2, G1.3, G5.1	Lecturing, discussion	
8	Hidden Markov model	G1.1, G1.2, G1.3, G4.2, G5.1	Lecturing, discussion	

9	Support vector machine	G1.1, G1.2, G1.3, G4.2, G5.1	Lecturing, discussion	
10	Neural network	G1.2, G1.3, G4.1, G5.1	Lecturing, discussion	
11	Review	G1.1, G1.2, G1.3, G4, G5.1	Q&A, discussion	Final test

LABORATORY

ID	Topic	Course outcomes	Teaching/Learning Activities (samples)	Assessments
1	Review of statistics	G1.2, G1.3, G2.1	Discussion, Q&A	HW#1, HW#2
2				
3	Multivariate linear regression	G1.2, G1.3, G3.1, G3.2, G3.3, G5.1	Discussion, Q&A	
4				
5	PCA	G1.2, G1.3, G5.1	Discussion, Q&A	HW#3, HW#4
6				
7				
8	HMM	G1.1, G1.2, G1.3, G4.2, G5.1	Discussion, Q&A	HW#3, HW#4
9	SVM	G1.2, G1.3, G4.1, G5.1	Discussion, Q&A	HW#3, HW#4
10				

6. ASSESSMENTS

ID	Topic	Description	Course outcomes	Ratio (%)
	Assignments			**70%**
HW#1	Matlab		G1.2, G1.3, G2.1	10%
HW#2	Gaussian distribution		G1.2, G1.3, G2.1	10%
HW#3	Image classification	Literature review	G1.1, G1.2, G1.3, G2.1, G3.1, G3.2, G3.3, G4.1, G4.2, G4.3, G4.4, G5.1	20%
HW#4	Scene categorization		G1.1, G1.2, G1.3, G2.1, G3.1, G3.2, G3.3, G4.1, G4.2, G4.3, G4.4, G5.1	30%
	Final test	**Q&A**		**30%**

7. RESOURCES

Textbooks

- **Pattern Recognition and Machine Learning**, Chris Bishop, *Springer*, UK, 2006
- **Pattern Classification**, 2nd edition, Richard O. Duda, Peter E. Hart, David G. Stork, *Wiley Interscience*, USA, 2001
- **Machine Learning**, Tom Mitchell, *McGraw Hill*, USA, 19997

Others

- Matlab

8. GENERAL REGULATIONS & POLICIES

- All students are responsible for reading and following strictly the regulations and policies of the school and university.
- Students who are absent for more than 3 theory sessions are not allowed to take the exams.
- For any kind of cheating and plagiarism, students will be graded 0 for the course. The incident is then submitted to the school and university for further review.
- Students are encouraged to form study groups to discuss on the topics. However, individual work must be done and submitted on your own.