

Trường Đại học Bách Khoa -
 ĐHQG Tp.HCM
 Khoa: Khoa Khoa học và Kỹ
 thuật Máy tính
 Khoa/Bộ môn quản lý MH: Khoa
 học Máy tính

Tp.HCM, ngày tháng
 năm

Đề cương môn học Sau đại học

HỌC MÁY VÀ ỨNG DỤNG (MACHINE LEARNING AND APPLICATIONS)

Mã số MH: CO5241

Số tín chỉ:		Tc (LT.BT&TH.Tự Học): 3					TCHP:					
Số tiết	-Tổng:	69	LT:	30	BT:	12	TH:	0	ĐA:		BTL/TL:	27
Đánh giá:	Thi:				40%	Trắc nghiệm, điền chỗ trống, tính toán, và lập trình						
	Bài tập:				20%	Tính toán và Lập trình						
	Bài tập lớn:				40%	Thực hiện dự án nhỏ theo nhóm 2-3 thành viên						
- Môn tiên quyết:												
- Môn học trước:												
- Môn song hành:												
- CTĐT ngành (Mã ngành):		Khoa Học Máy Tính (8480101)										
- Ghi chú khác:												

1. Mục tiêu môn học:

Môn học cung cấp cho người học những kỹ thuật quan trọng và có tính ứng dụng cao trong lĩnh vực học máy hiện đại. Với những nội dung trong môn học, học viên được kỳ vọng có thể hiểu, vận dụng, và đánh giá được những cách tiếp cận hiện đại trong phân tích dữ liệu văn bản, tiếng nói, hình ảnh/video và các dạng dữ liệu khác.

Aims:

This course is to provide to students important and applied techniques in modern machine learning. With the content given in the course syllabus, students are expected to understand, use and evaluate modern approaches for processing text, speech, image/video and for other kinds of data.

2. Nội dung tóm tắt môn học:

Môn học bao gồm những khối kiến thức và kỹ năng sau đây:

1. các khái niệm cơ bản trong học máy;
2. một số kỹ thuật học máy truyền thống có chọn lọc như cây quyết định và rừng ngẫu nhiên, máy vector hỗ trợ, và mô hình Markov ẩn;
3. kỹ thuật học sâu.

Để hỗ trợ những nội dung ở trên, một số kỹ thuật toán học cũng được bổ sung vào trước một số nội dung khi cần thiết.

Course outline:

This course provides the following contents:

1. basis concepts in machine learning;
2. selected techniques in the traditional approach, for examples, decision tree and random forest, support vector machine, and hidden markov model;
3. deep learning.

To support the above-mentioned contents, selected and necessary techniques in mathematics are also presented before each of the content.

3. Tài liệu học tập:

[1] Stephen Marsland, **Machine Learning: An Algorithmic Perspective**, 2nd Eds, CRC Press, 2015

[2] Lior Rokach and Oded Maimon, **Data Mining with Decision Trees: Theory and Applications**, 2nd Eds, World Scientific Publishing, 2015

[3] Leo Breiman, **Classification and Regression Trees**, 1st Eds, CRC Press, 2017

[4] Nello Cristianini and John Shawe Taylor, **An Introduction to Support Vector Machines and Other Kernel-based Learning Methods**, 1st Eds, Cambridge University Press, 2000.

[5] Lawrence R. Rabiner, **A tutorial on hidden Markov models and selected applications in speech recognition**, in Proceedings of IEEE, Vol. 77, No. 2, pp. 257 - 286, 1989

[6] Ian Goodfellow and Yoshua Bengio and Aaron Courville, **Deep learning**, MIT Press, 2016

[7] FRANÇOIS CHOLLET, **Deep Learning with Python**, Manning Publications, 2018

4. Các hiểu biết, các kỹ năng cần đạt được sau khi học môn học:

STT	Chuẩn đầu ra môn học (CĐRMH)	Công cụ đánh giá CĐRMH	Đóng góp CĐR Chương trình (CĐRCT)		
			Ứng dụng	Nghiên cứu	
CĐRMH.1	Có khả năng giải thích một cách rõ ràng và chi tiết các khái niệm trong học máy; giải thích được nguyên lý xây dựng và hoạt động của từng kỹ thuật đã học; và giải thích được những mặt tích cực và điểm hạn chế của từng kỹ thuật đã học.		a, c, d, e	c, d, e	1.1, 1.2, 1.3, 2.3
CĐRMH.2	Có khả năng cài đặt và thực hành được những kỹ thuật đã học thông qua các thư viện hỗ trợ và ngôn ngữ lập trình.		c, d, e	c, d, e	2.1, 2.2, 2.3, 2.5
CĐRMH.3	Có khả năng sử dụng được các kỹ thuật đã học vào phân tích dữ liệu trong thực tiễn.		c, d, e	c, d, e	2.1, 2.2, 2.3, 2.5, 3.1

Learning outcomes:

No.	Course learning outcomes (CLO)	CLO assessment	Matching with PLO		
			Coursework	Research	
L.O.1	Be able to explain thoroughly important concepts in machine learning; to explain the principle for constructing and for operating the techniques given in the course; and to explain the advantage and the limitation of the learnt techniques.		a, c, d, e	c, d, e	1.1, 1.2, 1.3, 2.3
L.O.2	Be able to implement and to practise the techniques provided in the course by using programming language and using useful frameworks/libraries for machine learning.		c, d, e	c, d, e	2.1, 2.2, 2.3, 2.5
L.O.3	Be able to use the learnt techniques for analyzing data in real life.		c, d, e	c, d, e	2.1, 2.2, 2.3, 2.5, 3.1

Bảng ánh xạ chuẩn đầu ra môn học và chuẩn đầu ra chương trình ứng dụng:

Chuẩn đầu ra môn học (CĐRMH)	Chuẩn đầu ra của chương trình (CĐRCT)										
	a	b	c	d	e	f	g	h	i	j	k
CĐRMH.1	✓		✓	✓	✓						
CĐRMH.2			✓	✓	✓						
CĐRMH.3			✓	✓	✓						

Bảng ánh xạ chuẩn đầu ra môn học và chuẩn đầu ra chương trình nghiên cứu:

Chuẩn đầu ra môn học (CĐRMH)	Chuẩn đầu ra của chương trình (CĐRCT)										
	a	b	c	d	e	f	g	h	i	j	k
CĐRMH.1			✓	✓	✓						
CĐRMH.2			✓	✓	✓						
CĐRMH.3			✓	✓	✓						

5. Hướng dẫn cách học - chi tiết cách đánh giá môn học:

1. Đây là môn học hướng đến cả lý thuyết và thực hành. Học viên cần chuẩn bị các tư liệu sau trong đầu cho môn học:

a. Về lý thuyết:

- Học viên cần trang bị tài liệu tham khảo đầy đủ cho mỗi chương.
- Đọc trước tài liệu cho mỗi buổi lên lớp. Thời gian trên lớp chủ yếu dành cho mở rộng khái niệm, liên hệ thực tiễn, và thảo luận.

b. Về bài tập

- Học viên cần tổng hợp các tài liệu và các ví dụ lập trình liên quan.
- Hiện thực bài tập bằng ngôn ngữ Python với Google Colab.

c. Về bài tập lớn (tiểu luận):

- Học viên cần có khả năng tổng hợp tài liệu liên quan đến đề tài bài tập lớn.
- Học viên cần trang bị các phương tiện lập trình cho môn học, cụ thể là Python với Google Colab
- Hiện thực nội dung bài tập lớn.

Đánh giá sẽ dựa trên:

- Bài tập lớn: 40%
- Bài tập: 20%
- Thi cuối kỳ: 40%

Learning strategies & Assessment Scheme:

1. This course is designed to help students to strongly acquire both of the theoretical and the practical aspect of image and video processing. Student attending the course should prepare the following materials at the beginning of the course.

a. Theoretical parts:

- Student should prepare reference materials for each chapter before the lecture.
- Read reference materials for each session before attending class. Lecture hours are reserved for extending concepts, for relating to applications, and for attending discussion.

b. Exercise parts

- Students should identify and collect related materials and programming examples.

- Implement exercises using Python with Google Colab.

Assignments:

Student should prepare the ability to obtain/search related materials for assignments.

Students should prepare fundamental library/software for practice, especially Python and Google Colab.

Implementing the assignments.

Grading is based on:

- Exercises: 20%

- Assignments: 40%

- Final Examination: 40%

6. Nội dung chi tiết:

Tuần/ Buổi	Chủ đề (chương)	Nội dung	Chuẩn đầu ra môn học	Tài liệu
1	Giới thiệu	1. Giới thiệu về môn học 2. Giới thiệu về học máy 2.1 Các khái niệm cơ bản của học máy 2.2 Các ví dụ về học máy đơn giản 3. Google Colab và các thư viện hỗ trợ	CĐRMH.1	[1], [6], [7]
2	Cây quyết định	1. Giới thiệu 2. Sử dụng cây quyết định cho phân loại và hồi quy 3. Phương pháp xây dựng cây quyết định 3.1 Giải thuật tổng quát 3.2 Tiêu chuẩn dừng 3.3 Phương pháp tách nhánh 3.4 Phương pháp tỉa nhánh 3.5 Các giải thuật truyền thống 4. Rừng ngẫu nhiên 5. Hiện thực và sử dụng cây quyết định và rừng ngẫu nhiên với Python và scikit-learn	CĐRMH.1, CĐRMH.2, CĐRMH.3	[1], [2], [3], [9]
3-4	Máy véc-tơ hỗ trợ	1. Giới thiệu 2. Cơ sở toán học liên quan 2.1 Hình học giải tích 2.3 Nhân tử Lagrange 2.4 Tối ưu lồi có ràng buộc 3. Trường hợp khả tách tuyến tính hai lớp 3.1 Bài toán tối ưu cho SVM 3.2 Giải bài toán bằng lập trình 4. Trường hợp tổng quát 4.1 Biên mềm 4.2 Phương pháp kernel 4.3 Giải bài toán bằng lập trình 4.4 Trường hợp nhiều lớp 5. SVM cho hồi quy 6. Hiện thực và sử dụng SVM với Python và scikit-learn	CĐRMH.1, CĐRMH.2, CĐRMH.3	[1], [4], [8], [9]

Tuần/ Buổi	Chủ đề (chương)	Nội dung	Chuẩn đầu ra môn học	Tài liệu
5-6	Mô hình Markov ẩn	1. Giới thiệu 2. Kiến thức toán liên quan 2.1 Xác suất, xác suất có điều kiện, và độc lập xác suất 2.2 Phân phối xác suất, sinh số ngẫu nhiên 3. Quá trình Markov 4. Mô hình Markov ẩn 4.1 Bài toán xác định xác suất của một chuỗi các quan sát 4.2 Bài toán xác định chuỗi các trạng thái 4.3 Bài toán học các tham số của mô hình từ dữ liệu đầu vào 5. Hiện thực và sử dụng mô hình Markov ẩn với Python và scikit-learn	CĐRMH.1, CĐRMH.2, CĐRMH.3	[1],[5],[9]
7	Học sâu: Chuẩn bị dữ liệu	1. Dữ liệu và nhãn 2. Array, Tensor và GPU 3. Đọc và hiển thị trực quan dữ liệu 4. Biến đổi và chuẩn hóa dữ liệu 5. Thống kê và làm giàu dữ liệu 6. Các tập dữ liệu phổ biến 7. Chuẩn bị dữ liệu với Google Colab, Tensorflow, và Pytorch (bài tập)	CĐRMH.1, CĐRMH.2, CĐRMH.3	[6][7][10][11]
8	Học sâu: Thiết kế mạng I	1. Các lớp tính toán cơ bản: 1.1 Lớp kết nối đầy đủ; 1.2 Lớp truyền; 1.3 Lớp gộp (pooling); 1.4 Lớp softmax. 2. Hàm MSE và bài toán hồi quy 3. Hàm Cross-Entropy và bài toán phân loại 4. Giải thuật SGD cơ bản 5. Các phương pháp đánh giá cho bài toán hồi quy và phân loại 6. Huấn luyện, kiểm thử, và kiểm tra 7. Hiện thực các lớp với Python 8. Tạo và sử dụng mạng với Tensorflow/Keras và Pytorch	CĐRMH.1, CĐRMH.2, CĐRMH.3	[6][7][10][11]
9	Học sâu: Cài đặt I	1. Giới thiệu về Google Colab, Tensorflow, Pytorch và GPU 2. Bài tập về chuẩn bị dữ liệu 3. Bài tập về thiết kế mạng nơron 1	CĐRMH.1, CĐRMH.2, CĐRMH.3	[6][7][10][11]
10	Học sâu: Thiết kế mạng nơron II	1. Các lớp tính toán: 1.1 Lớp tích chập (1-,2-, và 3-chiều); 1.2 Lớp Dropout; 1.3 Lớp Batchnorm và chuẩn hóa khác; 1.4 Khối thịnh dư; 1.5 Khối kết nối dày. 2. Kỹ thuật chống quá khớp 3. Triplet loss và bài toán nhận dạng 4. SGD: các biến thể 5. Thu thập thông tin thống kê và hiển thị trực quan quá trình huấn luyện mạng nơron 6. Tạo và sử dụng mạng nơron học sâu với Tensorflow/Keras và Pytorch	CĐRMH.1, CĐRMH.2, CĐRMH.3	[6][7][10][11]

Tuần/ Buổi	Chủ đề (chương)	Nội dung	Chuẩn đầu ra môn học	Tài liệu
11	Học sâu: Cài đặt II	1. Bài tập về thiết kế mạng nơron tích chập 2. Bài tập về xây dựng các biểu đồ thống kê để hiển thị trực quan quá trình huấn luyện mạng nơron. 3. Huấn luyện, kiểm tra và kiểm thử cho mạng nơron với các biến thể của SGD.	CĐRMH.1, CĐRMH.2, CĐRMH.3	[6][7][10][11]
12	Mạng nơron hồi quy	1. Giới thiệu về mạng nơron hồi quy (RNN) 2. Các phiên bản hiện thực phổ biến: LSTM và GRU 3. Nguyên tắc lan truyền ngược theo thời gian 4. Các ứng dụng của RNN 5. Tạo và sử dụng mạng nơron với Tensorflow/Keras và Pytorch	CĐRMH.1, CĐRMH.2, CĐRMH.3	[6][7][10][11]
13	Học sâu: Cài đặt III	1. Luyện tập tính toán LSTM và GRU dùng Numpy, Tensorflow và Pytorch 2. Bài tập về sử dụng LSTM và GRU cho phân tích dữ liệu văn bản và dữ liệu có yếu tố thời gian.	CĐRMH.1, CĐRMH.2, CĐRMH.3	[6][7][10][11]
14	Báo cáo và ôn tập	1. Báo cáo tiểu luận 2. Ôn tập cuối chương trình		[1-11]

7. Giảng viên tham gia giảng dạy:

CBGD
chính:

TS. Lê
Thành
Sách

CBGD
tham
gia:

TS.
Trần
Tuấn
Anh
TS.
Nguyễn
Hồ
Mẫn
Rạng

**XÁC NHẬN
CỦA HỘI
ĐỒNG XÂY
DỰNG
CHƯƠNG
TRÌNH ĐÀO
TẠO VÀ KHOA**

*Tp. Hồ Chí
Minh, ngày
..... tháng
..... năm*

.....
**GIẢNG
VIÊN
LẬP ĐỀ
CƯƠNG**

**TS. Phạm
Hoàng
Anh**