

ĐỀ CƯƠNG MÔN HỌC

CS114.Machine Learning - Máy Học

Giới thiệu môn học

- Máy học đang có nhiều ứng dụng trong thực tế như xe tự lái, chẩn đoán y khoa, dịch máy, trả lời tự động, và nhận dạng giọng nói. Thay vì con người phải đưa ra các lệnh cụ thể để có thể viết chương trình phân loại một ảnh có người đi bộ hay không, đặc điểm của máy học là có thể tìm được cách phân loại bằng việc học từ dữ liệu ảnh huấn luyện trong đó có ảnh chứa người đi bộ và có ảnh không chứa người đi bộ. Bằng cách này, khi yêu cầu chuyển sang phân loại một ảnh có xe hơi hay không, chỉ cần thay đổi bộ dữ liệu ảnh huấn luyện. Nhu cầu học về máy học đang trở nên cấp thiết trong các chương trình đào tạo CNTT trong nước, nhằm trang bị các kiến thức và kỹ năng cần thiết cho thị trường lao động cạnh tranh trong cuộc cách mạng công nghiệp 4.0

Mục tiêu (Learning Objectives)

- Giới thiệu các khái niệm, thuật ngữ, và các thuật toán máy học, đặc biệt tập trung vào các thuật toán supervised learning và unsupervised learning. Các chủ đề sẽ được trình bày bao gồm: regression, classification, regularization, SVM, neural network, clustering, dimensionality reduction.
- Giới thiệu và hiện thực các bước trong xây dựng một ứng dụng của máy học giải quyết một bài toán cụ thể nào đó, tập trung vào các lĩnh vực như computer vision, natural language processing.
- Sử dụng các thư viện như scikit-learn và keras để xây dựng một ứng dụng máy học viết bằng ngôn ngữ lập trình Python.

Kết quả đầu ra (Learning Outcomes)

Sau khi hoàn thành khoá học, sinh viên có thể:

- Hiểu và diễn giải được mô hình toán học, cách hoạt động của các thuật toán máy học
- Biết cách đánh giá được ưu điểm và nhược điểm của từng thuật toán máy học.
- Biết cách đánh giá độ chính xác của các thuật toán máy học và chỉ ra được các yếu tố ảnh hưởng đến độ chính xác này.
- Biết cách lựa chọn thuật toán máy học phù hợp nhất để giải quyết một bài toán thực tế
- Biết cách vận dụng quy trình xây dựng một ứng dụng của máy học
- Cài đặt một ứng dụng máy học giải quyết bài toán trong lĩnh vực computer vision, NLP sử dụng các thư viện như scikit-learn, keras, viết bằng ngôn ngữ lập trình Python.

Đánh giá

- Điểm quá trình: bài thực hành, trả lời các câu hỏi trong quá trình học
- Điểm cuối kỳ thực hành: báo cáo đồ án và chương trình
- Điểm cuối kỳ lý thuyết: bài thi trên giấy gồm trắc nghiệm và tự luận, chỉ được sử dụng tài liệu giấy

Yêu cầu về thực hành

- SV sử dụng Google Colab hoặc các IDE như Spyder, PyCharm, hay Atom.
- SV nộp bài thực hành qua hệ thống chấm bài tự động WeCode của Trường.
- SV cần có tài khoản Github để lưu bài tập và đồ án cuối kỳ.

Tài liệu tham khảo

- Hands-On Machine Learning with Scikit-Learn& TensorFlow, Aurelien Geron, Oreilly, 2017 <https://github.com/ageron/handson-ml>
- Introduction to Machine Learning with Python, Andreas Mueller and Sarah Guido, Oreilly, 2016
https://github.com/amueller/introduction_to_ml_with_python

Tiết	Nội dung
T1	<p>Tổ chức lớp học</p> <ul style="list-style-type: none"> - chia nhóm - khảo sát ý kiến → để xem SV chuẩn bị cho môn học này như thế nào - điểm danh <p>Giới thiệu môn học</p> <ul style="list-style-type: none"> - mục tiêu (learning objectives), kết quả đầu ra (learning outcomes) - nội dung, - cách đánh giá → điểm quá trình, điểm thực hành, điểm đồ án cuối kì, điểm khuyến khích <p>Thực hành:</p> <ul style="list-style-type: none"> - WeCode: https://khmt.uit.edu.vn/wecode/cs114x/ <p>Tài liệu tham khảo</p> <ul style="list-style-type: none"> - Các khoá học trực tuyến - Các sách tham khảo <p>Phương pháp học</p> <ul style="list-style-type: none"> - Tự học như thế nào? - Luyện tập thực hành
T2 - T6	<p>Giới thiệu về máy học</p> <ul style="list-style-type: none"> - Định nghĩa (What is Machine Learning) - Các ứng dụng (Machine Learning Applications) - Máy học và lập trình thông thường (Machine Learning vs Traditional Programming) - Phân loại các thuật toán máy học (Types of Machine Learning Algorithms → Supervised vs Unsupervised)

	<ul style="list-style-type: none"> - Phương pháp học → học thuật vs ứng dụng <p>Quiz01 → kiểm tra phân biệt bài toán thuộc loại nào trong số supervised và unsupervised</p>
T7 - T9	<p>Các bước xây dựng một ứng dụng máy học (Machine Learning Pipeline)</p> <p>Cách thách thức khi xây dựng một ứng dụng máy học (Challenges of ML)</p> <p>Giới thiệu một số AutoML Framework như Auto-Keras, Auto-Sklearn</p> <p>Chế tác đặc trưng (Feature Engineering)</p>
TH-B1	<p>Tham khảo:</p> <ul style="list-style-type: none"> - https://d2l.ai/chapter_preliminaries/ndarray.html - https://jakevdp.github.io/PythonDataScienceHandbook/ - https://github.com/amueller/introduction_to_ml_with_python - https://github.com/ageron/handson-ml <p>Các nội dung:</p> <ul style="list-style-type: none"> - Jupyter Notebook - Google Colab - Python - Các thư viện: scikit-learn, numpy, scipy, pandas, matplotlib <p>Bài tập lập trình</p>
T10 - T18	<p>Lập trình một số ứng dụng minh họa</p> <ul style="list-style-type: none"> - Classification - Computer Vision: nhận dạng chữ số viết tay - MNIST - Classification - NLP: phân loại bình luận - Regression: dự đoán giá nhà <p>Một số lưu ý khi triển khai ứng dụng thực tế</p>

	<ul style="list-style-type: none"> - Thu thập dữ liệu - Khảo sát dữ liệu - Chia train/test - Chọn model - Tune parameter - Evaluate model - Analyze result - Dữ liệu từ Kaggle, tự thu thập
T19 - T24	Regression → minh hoạ bằng Kaggle Challenge <ul style="list-style-type: none"> - Model Representation - Hypothesis - Cost Function - Optimization - Gradient Descent - Linear Algebra Bài tập lập trình
T25 - T28	Logistic Regression - Classification → minh hoạ bằng Kaggle Challenge <ul style="list-style-type: none"> - Model Representation - Hypothesis - Cost Function - Optimization - Gradient Descent - Overfitting vs Underfitting - Regularization - Tuning Hyperparameters
T29 - T30	SVM <ul style="list-style-type: none"> - Model Representation

	<ul style="list-style-type: none"> - Optimization
TH-B2	<p>Hướng dẫn cài đặt và sử dụng các thuật toán</p> <ul style="list-style-type: none"> - Linear Regression - Logistic Regression - SVM
T31 - T33	<p>Unsupervised Learning</p> <ul style="list-style-type: none"> - Clustering - K-Means - Dimensionality Reduction - PCA
T34 - T36	<p>Neural Network & Deep Learning</p> <ul style="list-style-type: none"> - Feedforward neural network - Backpropagation
T37 - T39	<p>Recommendation System</p> <p>Anomaly Detection</p>
TH-B3	<p>Hướng dẫn cài đặt</p> <ul style="list-style-type: none"> - Recommendation System - Anomaly Detection
T40 - T42	<p>Đồ án cuối kì (Capstone Project)</p> <ul style="list-style-type: none"> - Mô tả input/output của các bài toán <ul style="list-style-type: none"> - Computer Vision - NLP - Speech Recognition - Thiết kế pipeline - Chọn component để tập trung vào: Ceiling Analysis: What Part of the Pipeline to Work on (Coursera)
T43 -	<p>Advanced Topics → Seminar</p>

T45	<ul style="list-style-type: none"> - Máy học với dữ liệu lớn (Large Scale ML) - Reinforcement Learning
	<p>Tham khảo Coursera</p> <ul style="list-style-type: none"> - Machine Learning - Stanford: https://www.coursera.org/learn/machine-learning - Machine Learning with Python - IBM: www.coursera.org/learn/machine-learning-with-python/ - Applied Machine Learning in Python - Univ of Michigan: https://www.coursera.org/learn/python-machine-learning - Introduction to Machine Learning - Duke Univ: https://www.coursera.org/learn/machine-learning-duke - Machine Learning Foundations: A Case Study Approach - Univ of Washington: https://www.coursera.org/learn/ml-foundations - Practical Machine Learning - Johns Hopkins Univ: https://www.coursera.org/learn/practical-machine-learning

Các khoá học trực tuyến (MOOC)

- Machine Learning - Stanford:
<https://www.coursera.org/learn/machine-learning>
- Machine Learning with Python - IBM:
www.coursera.org/learn/machine-learning-with-python/
- Applied Machine Learning in Python - Univ of Michigan:
<https://www.coursera.org/learn/python-machine-learning>
- Introduction to Machine Learning - Duke Univ:
<https://www.coursera.org/learn/machine-learning-duke>
- Machine Learning Foundations: A Case Study Approach - Univ of Washington: <https://www.coursera.org/learn/ml-foundations>