ĐỀ CƯƠNG HỌC PHẦN

**HỌC MÁY**

**(Machine Learning)**

**1. Thông tin tổng quát** *(general information)*

*(thông tin tổng quát và điều kiện đăng ký môn học)*

|  |  |
| --- | --- |
| * Tên môn học: | Học máy |
| * Mã số môn học: |  |
| * Thuộc khối kiến thức/ kỹ năng:   🞎 Kiến thức cơ bản ⌧ Kiến thức cơ sở ngành  🞎 Kiến thức chuyên ngành 🞎 Kiến thức khác  🞎 Môn học chuyên về kỹ năng chung 🞎 Môn học đồ án/ luận văn tốt nghiệp | |
| * Số tín chỉ: | 3 TC |
| + Số tiết lý thuyết/số buổi: | 39/13 |
| + Số tiết thực hành/số buổi: | 18/6 |
| + Số tiết thực tập tại cơ sở hoặc đồ án/số buổi |  |
| * Môn học tiên quyết: |  |
| * Môn học song hành: |  |
| * Môn học tiếp theo: |  |

**2. Mô tả học phần***(course descriptions)*

Học phần Machine learning (học máy) cung cấp cho sinh viên các kiến thức, kỹ năng về quy trình xây dựng một hệ thống machine learning. Học phần sẽ giới thiệu một số bài toán điển hình trong machine learning (hồi quy, phân lớp, gôm cụm) và một số thuật toán machine learning kinh điển (hồi quy tuyến tính, k- Nearest Neighbors, cây quyết định, Support Vector Machines, K-Means, …) cũng như hiện đại (phương pháp tối ưu Gradient Descent, mạng neuron nhân tạo) để giải quyết các bài toán đó. Thông qua học phần, sinh viên sẽ có các kỹ năng xây dựng, tinh chỉnh, thực nghiệm, đánh giá hệ thống machine learning..

**3. Nguồn học liệu** *(learning resources: course books, reference books, and softwares) (Các giáo trình, tài liệu tham khảo, các phần mềm, không quá 5 cuốn)*

***Giáo trình:***

[1] Hands-On Machine Learning with Scikit- Learn, Keras, and Tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems, Aurélien Géron.

[2] Machine Learning Algorithms, Giuseppe Bonaccorso, 2017

***Tài liệu tham khảo:***

[3]. Python Machine Learning Tutorial – Tasks and Applications

[<https://data-flair.training/blogs/python-machine-learning-tutorial/> ]

[4] Machine Learning cơ bản, Vũ Hữu Tiệp, 2020

***Phần mềm:***

[1] Google colab

[2] Visual code

**4. Mục tiêu học phần** *(course goals)*

*(các mục tiêu cụ thể của học phần cần nêu rõ học phần sẽ cung cấp cho sinh viên kiến thức và kỹ năng gì, thể hiện sự liên quan với các chủ đề CĐR (X.x.x) của CTĐT được phân nhiệm cho HP, tối đa 5 mục tiêu)*

|  |  |  |
| --- | --- | --- |
| **Mục tiêu (Gx) [1]** | **Mô tả mục tiêu [2]** | **Các CĐR của CTĐT**  **[3]** |
| **G1** | * Nắm bắt được các thành phần của một hệ thống machine learning. * Nắm rõ những bài toán cơ bản, cũng như các thuật toán machine learning cơ bản. * Xử lý, trích chọn đặc trưng từ dữ liệu thô. * Sử dụng thành thạo scikit-learn để xây dựng, tinh chỉnh mô hình. | 2 |
| **G2** | * Vận dụng thành thạo kỹ năng làm việc nhóm, lập kế hoạch và viết báo cáo kỹ thuật * Có năng lực trình bày giải pháp kỹ thuật. | 3 |

*[1]: Ký hiệu mục tiêu của học phần. [2]: Mô tả mục tiêu bao gồm các động từ chủ động, các chủ đề CĐR (X.x.x) và bối cảnh áp dụng tổng quát. [3]: Ký hiệu CĐR của CTĐT đã được xác định trong mục tiêu đào tạo và chuẩn đầu ra của CTĐT, phần chuẩn đầu ra của CTĐT.*

**5. Chuẩn đầu ra học phần** *(course learning outcomes)*

|  |  |  |  |
| --- | --- | --- | --- |
| **CĐR**  **(LO.x) [1]** | **Mô tả CĐR [2]** | **Các CĐR của CTĐT**  **[3]** | **CĐR theo đề cương CDIO [4]** |
| LO.1 | Hiểu được bài toán machine learning và tầm quan trọng của dữ liệu | 1 | 1 |
| LO.2 | Áp dụng thuật toán kinh điển giải quyết bài toán machine learning | 2.1 | 2.2.1 |
| LO.3 | Áp dụng thuật toán hiện đại vào bài toán machine learning | 2.1 | 2.2.1 |
| LO.4 | Có khả năng giải quyết một số vấn đề thực tế | 2.2 | 2.2.4 |
| LO.5 | Có khả năng thực nghiệm và khám phá tri thức để giải quyết bài toán về học máy. | 2.2 | 2.2.4 |

*[1]: Ký hiệu CĐR của môn học. [2]: Mô tả CĐR, bao gồm các động từ chủ động mô tả năng lực của sinh viên (theo nội dung CĐR) và bối cảnh áp dụng cụ thể. [3]: Ký hiệu CĐR của CTĐT đã được xác định trong mục tiêu đào tạo và chuẩn đầu ra của CTĐT, phần chuẩn đầu ra của CTĐT. [4]: Các CĐR theo đề cương CDIO*

**6. Đánh giá học phần** *(course assessment)*

Kế hoạch các bài đánh giá tương ứng với CĐR

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **STT** | **Bài đánh giá** | **Tuần/Buổi** | **CĐR môn học (LO.x) [1]** | **Tỷ lệ (%) [2]** | **Thành phần đánh giá** | |
| **Quá trình [3]** | **Tổng kết [4]** |
| 1 | BĐG 1: Đánh giá phần  thực hành |  | LO.2; LO.3 | 20% | 20% |  |
| 2 | BĐG 2: Đánh giá bài kiểm tra thường xuyên 1 |  | LO.3 | 15% | 15% |  |
| 3 | BĐG 3: Kiểm tra đánh giá thường xuyên 2 |  | LO.3 | 15% | 15% | 0% |
| 4 | BĐG 4: Bài tập lớn |  | LO.3; LO.4  LO.5 | 20% |  | 20% |
| 5 | BĐG 5: Kiểm tra đánh giá cuối kỳ |  | LO.2 | 30% |  | 30% |
|  | Tổng |  |  | **100%** | **50%** | **50%** |

*[1]: Các CĐR được đánh giá. [2]: Trọng số điểm của BĐG trên điểm đánh giá tổng hợp HP. [3]: Trọng số điểm quá trình của BĐG trên điểm đánh giá tổng hợp của HP [4]: Trọng số điểm tổng kết của BĐG trên điểm đánh giá tổng hợp của HP.*

**7. Kế hoạch giảng dạy** *(lesson plan):*

*(các nội dung giảng dạy theo buổi học, thể hiện sự liên quan với các CĐR của môn học, các hoạt động dạy và học (ở lớp, ở nhà), và các bài đánh giá của môn học)*

***Lý thuyết***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Buổi học (3 tiết)** [1] | **Nội dung**  [2] | **CĐR HP** [3] | **Hoạt động**  **dạy và học** [4] | **Bài đánh giá** [5] |
| 1 | **Chương 1. Giới thiệu về machine learning (1 buổi)**   * 1. Định nghĩa machine learning   2. Quy trình xây dựng hệ thống machine learning   3. Phân loại thuật toán machine learning      1. Phân loại dựa trên phương pháp học      2. Phân loại dựa trên chức năng   4. Một số bài toán   5. Datasets   6. Thư viện lập trình machine learning | LO1; | **Giảng viên**   * Thảo luận với sinh viên về định nghĩa machine learning * Thuyết giảng về quy trình xây dựng hệ thống machine learning. * Thuyết giảng về các loại thuật toán machine learning và đặc trưng từng loại. * Giới thiệu một số bài toán ứng dụng. * Giới thiệu về các Dataset thực hành với machine learning * Giới thiệu các thư viện lập trình machine learning.   **Sinh viên:**  - Học ở lớp: nghe giảng lý thuyết; thảo luận định nghĩa machine learning.  - Học ở nhà: Đọc tài liệu; | BĐG 1, BĐG 2 |
| 2 | **Chương 2. Trích chọn đặc trưng (1buổi)**   * 1. Mô tả dữ liệu   2. Biên tập dữ liệu   3. Tóm tắt dữ liệu   4. Trực quan hoá dữ liệu   5. Khái niệm về đặc trưng (features)   6. Trích đặc trưng trên dữ liệu liên tục   7. Trích đặc trưng trên dữ liệu rời rạc   8. Trích đặc trưng trên dữ liệu văn bản   9. Trích đặc trưng trên dữ liệu đa phương tiện   10. Trích đặc trưng trên dữ liệu tuần tự   11. Chuẩn hóa đặc trưng   12. Chọn đặc trưng | LO2;LO3 | **Giảng viên**  - Thuyết giảng về nhiệm vụ và tầm quan trọng của mô tả dữ liệu và biên tập dữ liệu.  - Thảo luận về vai trò của tóm tắt dữ liệu.  - Thuyết giảng về khái niệm trực quan dữ liệu, giới thiệu về các kỹ thuật trực quan hoá dữ liệu.  - Thuyết giảng về khái niệm đặc trưng trong dữ liệu.  - Thuyết giảng về kỹ thuật trích chọn đặc trưng trên dữ liệu liên tục.  - Thuyết giảng về kỹ thuật trích chọn đặc trưng trên dữ liệu rời rạc.  - Thuyết giảng về kỹ thuật trích chọn đặc trưng trên dữ liệu văn bản.  - Thuyết giảng về kỹ thuật trích chọn đặc trưng trên dữ liệu đa phương tiện như ảnh, âm thanh, video…  - Thuyết giảng về kỹ thuật trích chọn đặc trưng trên dữ liệu tuần tự.  - Thuyết giảng về các kỹ thuật chuẩn hoá đặc trưng và vai trò của chuẩn hoá đặc trưng.  **Sinh viên:**  - Học ở lớp: nghe giảng lý thuyết; thảo luận định nghĩa machine learning.  - Học ở nhà: Đọc tài liệu; tìm hiểu thêm một số kỹ thuật chuẩn hoá đặc trưng. | BĐG 1, BĐG 4 |
| 3 | **Chương 3. Hồi quy (Regression**   * 1. Bài toán hồi quy   2. Hồi quy tuyến tính   3. Hồi quy dựa trên cây | LO2; LO.4 | **Giảng viên**  - Giới thiệu về bài toán hồi quy.  - Đưa ra ví dụ trực quan về bài toán hồi quy tuyến tính.  - Thuyết giảng về thuật toán hồi quy tuyến tính, lý thuyết xây dựng mô hình.  - Thuyết giảng về thuật toán hồi quy dựa trên cây.  - Trình diễn code lập trình xử lý bài toán hồi quy tuyến tính.  **Sinh viên:**  - Học ở lớp: nghe giảng lý thuyết; thảo luận định nghĩa machine learning.  - Học ở nhà: Đọc tài liệu; thực hành lại code lập trình giải quyết bài toán hồi quy tuyến tính | BĐG 1, BĐG 2 |
| 4-6 | **Chương 4. Phân lớp (Classification) (3 buổi)**   * 1. Bài toán phân lớp dữ liệu   2. k-Nearest Neighbors   3. Cây quyết định   4. Naïve Bayes   5. Support Vector Machines | LO2 | **Giảng viên**  - Giới thiệu về bài toán phân lớp dữ liệu.  - Thuyết giảng về thuật toán k-Nearst Neighbors.  - Thuyết giảng về thuật toán cây quyết định  - Lấy ví dụ về bài toán cây quyết định, hướng dẫn áp dụng để thực hiện ví dụ.  - Thuyết giảng về thuật toán Naïve Bayes  - Lấy ví dụ về bài toán Naïve Bayes, hướng dẫn áp dụng để thực hiện ví dụ.  - Thuyết giảng về thuật toán SVM  - Lấy ví dụ về bài toán SVM, hướng dẫn lập trình để thực hiện ví dụ.  **Sinh viên:**  - Học ở lớp: nghe giảng lý thuyết;  - Học ở nhà:  + Đọc tài liệu; thực hành lại code lập trình.  + Làm bài tập về nhà thuật toán cây quyết định và Naïve Bayes | BĐG 1, BĐG 4, BĐG 5 |
| 7-8 | **Chương 5. Phân cụm (Clustering) (2 buổi)**   * 1. Bài toán phân cụm dữ liệu   2. K-Means   3. DBSCAN | LO2; | **Giảng viên**  - Thuyết giảng về bài toán phân cụm dữ liệu.  - Thuyết giảng về thuật toán K-means.  - Thuyết giảng về thuật toán DBSCAN.  - Hướng dẫn code python giải quyết bài toán K-means.  **Sinh viên:**  - Học ở lớp: nghe giảng lý thuyết;  - Học ở nhà: Đọc tài liệu; thực hành lại code lập trình. | BĐG 1, BĐG 3 |
| 9 | **Chương 6. Phương pháp tối ưu Gradient Descent (1 buổi)**   * 1. Gradient descent cho hàm một biến   2. Gradient descent cho hàm nhiều biến   3. Kỹ thuật Momentum   4. Stochastic Gradient Descent.   Mini-batch Gradient Descent | LO2;LO3 | **Giảng viên**  - Thuyết giảng về thuật toán Gradient descent cho hàm một biến.  - Thuyết giảng về thuật toán Gradient descent cho hàm nhiều biến.  - Thuyết giảng về kỹ thuật Momentum khắc phục tình trạng tối ưu cục bộ.  - Thuyết giảng về thuật toán Stochastic Gradient Descent cải tiến của thuật toán Gradient Descent  **Sinh viên:**  - Học ở lớp: nghe giảng lý thuyết;  - Học ở nhà: Đọc tài liệu; làm bài tập nhóm | BĐG1,BĐG4 |
| 10-11 | **Chương 7. Mạng neuron nhân tạo (Artificial Neural Networks - ANNs) (2 buổi)**   * 1. Neuron sinh học   2. Neuron nhân tạo (Perceptron)   3. Mạng neuron đa tầng (Multilayer Perceptrons - MLPs)   4. Thuật toán lan truyền tiến   5. Thuật toán huấn luyện   6. Bốn thành phần của mạng neuron | LO2;  LO3 | **Giảng viên**  - Thuyết giảng về phương thức  - Giải đáp các thắc mắc về bài tập nhóm  **Sinh viên:**  - Học ở lớp: nghe giảng lý thuyết; thảo luận việc thiết kế một bài trình bày và cách trình bày.  - Học ở nhà: Đọc tài liệu; làm bài tập nhóm | BĐG1, BĐG 4 |
| 12 | **Thực hiện BĐG2, BĐG3** | LO1;LO2;  LO3 | **Giảng viên**  - Giao bài kiểm tra 2,3  **Sinh viên:**  - Học ở lớp: Thực hiện bài đánh giá  - Học ở nhà: Đọc tài liệu; làm bài tập; làm bài tập nhóm. | BĐG2,BĐG3 |
| 13 | Ôn tập | LO1;LO2;  LO3 | **Giảng viên**  - Ôn tập kiến thức  **Sinh viên:**  - Ôn tập kiến thức  - Nộp báo cáo bài tập lớn. | BĐG5  BĐG5 |
| *[1]: Thông tin về tuần/ buổi học. [2]: Liệt kê nội dung giảng dạy theo chương, mục. [3]: Liệt kê CĐR liên quan của học phần (ghi ký hiệu Gx.x). [4]: Liệt kê các hoạt động dạy và học (ở lớp, ở nhà), bao gồm đọc trước tài liệu (nếu có yêu cầu). [5]: Liệt kê các bài đánh giá liên quan (ghi ký hiệu Ax.x).* | | | | |

***Thực hành***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tuần/Buổi học** [1] | **Nội dung**  [2] | **CĐR học phần** [3] | **Hoạt động**  **dạy và học** [4] | **Bài đánh giá**  [5] |
| 14 | Bài 1. Sử dụng thư viện   * numpy * pandas * scikit-learn | LO2 | **Giảng viên**   * Thuyết giảng * Đặt câu hỏi * Hướng dẫn sinh viên thực hiện   **Sinh viên**   * Nghe giảng, ghi chú * Trả lời câu hỏi * Thực hành trên máy | BĐG1, BĐG 4 |
| 15 | Bài 2. Trích chọn đặc trưng  Bài 3. Hồi quy | LO2 | **Giảng viên**   * Thuyết giảng * Đặt câu hỏi * Hướng dẫn sinh viên thực hiện   **Sinh viên**   * Nghe giảng, ghi chú * Trả lời câu hỏi * Thực hành trên máy | BĐG1, BĐG 4 |
| 16 | Bài 4. Phân lớp dữ liệu | LO2 | **Giảng viên**   * Thuyết giảng * Đặt câu hỏi * Hướng dẫn sinh viên thực hiện   **Sinh viên**   * Nghe giảng, ghi chú * Trả lời câu hỏi * Thực hành trên máy | BĐG1, BĐG 4 |
| 17 | Bài 6. Phân cụm | LO2 | **Giảng viên**   * Thuyết giảng * Đặt câu hỏi * Hướng dẫn sinh viên thực hiện   **Sinh viên**   * Nghe giảng, ghi chú * Trả lời câu hỏi * Thực hành trên máy | BĐG1, BĐG 4 |
| 18 | Bài 8. Gradient descent | LO2 | **Giảng viên**   * Thuyết giảng * Đặt câu hỏi * Hướng dẫn sinh viên thực hiện   **Sinh viên**   * Trả lời câu hỏi * Thực hành trên máy | BĐG1, BĐG 4 |
| 19 | Bài 9. Mạng neuron | LO2 | **Giảng viên**   * Thuyết giảng * Đặt câu hỏi * Hướng dẫn sinh viên thực hiện   **Sinh viên**   * Trả lời câu hỏi * Thực hành trên máy | BĐG1, BĐG 4 |
| *[1]: Thông tin về tuần/ buổi học. [2]: Liệt kê nội dung thực hành theo bài thực hành. [3]: Liệt kê CĐR liên quan của học phần (ghi ký hiệu Gx.x). [4]: Liệt kê các hoạt động dạy và học (ở lớp, ở nhà), bao gồm đọc trước tài liệu (nếu có yêu cầu). [5]: Liệt kê các bài đánh giá liên quan (ghi ký hiệu Ax.x).* | | | | |

**8. Quy định của học phần** *(course requirements and expectations)*

*(Các quy định của học phần (nếu có), thí dụ: sinh viên không nộp bài tập và báo cáo đúng thời hạn, được coi như không nộp bài; sinh viên vắng 2 buổi thực hành trở lên, không được phép dự thi cuối kỳ…)*

* Sinh viên vắng 2 buổi thực hành không được thi cuối kỳ.

**9. Phụ trách học phần**

- Khoa/ Bộ môn: Bộ môn KHMT, Khoa CNTT, Trường Đại học Xây dựng Hà Nội

- Các giảng viên giảng dạy: Tất cả giảng viên Bộ môn

- Địa chỉ và email liên hệ: Phòng 701 Nhà TN, bm.khmt@huce.edu.vn