

TIỂU LUẬN

NHẬP MÔN HỌC MÁY

QUY ĐỊNH

- Bài làm cần nộp lên hệ thống eLearning của khoa từ ngày **27/11/2025** và ngày **12/12/2025**. **Bài nộp muộn sẽ không được chấp nhận.**
- Thời gian thuyết trình bắt đầu từ ngày **28/11/2025**. Tất cả các thành viên trong nhóm đều phải tham gia thuyết trình.
- Bài làm cần tuân thủ các quy định sau:
 - Đây là bài làm nhóm. Mỗi nhóm không được quá 3 thành viên.
 - Các trường hợp đạo văn sẽ bị 0 điểm và xử lý theo quy định của khoa.
 - Bài nộp phải được trình bày rõ ràng. Nếu vi phạm định dạng hoặc trình bày cầu thả, bài làm có thể bị trừ từ 10% đến 50% tổng điểm.
 - Mọi thắc mắc cần giải đáp liên hệ giảng viên phụ trách lớp.
- **Nộp bài:**
 - Bài báo cáo dưới dạng file Word **và** PDF (.doc/.docx và .pdf).
 - Mã nguồn Python.
 - Slide báo cáo.
 - Tất cả các file được nén và đặt tên file nén theo **Mã số học viên** của nhóm, ví dụ: **52200000_52200001.docx**.

- Bài làm phải tuân theo quy định của khoa, không tính trang bìa, danh mục tài liệu tham khảo và mục lục.

- **Thời hạn nộp bài:**

- **Lần 1:** Nộp báo cáo sơ bộ (docx, pdf) và file trình bày **bài giữa kỳ (ppt)**. Hạn nộp bài trước 23h59' ngày 27/11/2025.
- **Lần 2:** Nộp báo cáo hoàn chỉnh (docx, pdf) và mã nguồn các chương trình. Hạn cuối 23h59' ngày 12/12/2025.

GIỮA KỲ

Mỗi nhóm sẽ được chỉ định một đề tài duy nhất dựa trên chữ số cuối cùng trong mã số sinh viên của các thành viên trong nhóm. Quy trình xác định như sau:

- Mỗi thành viên trong nhóm lấy **01 chữ số cuối cùng** trong mã số sinh viên của mình.
- **Cộng** tất cả các chữ số cuối cùng này lại để thu được một tổng số nguyên.
- Lấy **phần dư của tổng khi chia cho 6**.
- Kết quả phần dư sẽ được dùng để xác định đề tài của nhóm theo bảng sau:

Phần dư (mod 6)	Đề tài tương ứng
0	Support Vector Machine (SVM)
1	Expectation–Maximization (EM) Algorithm
2	Principal Component Analysis (PCA)
3	Linear Discriminant Analysis (LDA)

4	Singular Value Decomposition (SVD)
5	Ensemble Learning

Ví dụ: nhóm có 2 thành viên với MSSV tương ứng là 01234⁵ và 01234⁶. Mã đề tài của nhóm sẽ là: $(5 + 6) \bmod 6 = 5$.

Lưu ý: Đối với mỗi đề tài, các nhóm cần trình bày chi tiết nguyên lý và quy trình của thuật toán, minh họa các bước thực hiện bằng ví dụ cụ thể. Giới thiệu bài toán minh họa và cung cấp đoạn mã nguồn (code) minh họa thể hiện cách triển khai thuật toán đó.

CUỐI KỲ

Câu 1: 4 điểm

1. Trình bày chi tiết giải thuật lan truyền ngược (Backpropagation) để học tham số cho Feed Forward Neural Network đa tầng ẩn giải cho bài toán Regression và sử dụng Mean Square Error cho hàm Loss.
2. Lập trình theo thuật toán trên (bằng ngôn ngữ Python) và ứng dụng cho một dữ liệu thực tế (có thể sử dụng dataset trong UCI).
3. Tìm hiểu và trình bày kỹ thuật Early Stopping để giải quyết vấn đề Overfitting.
4. Cải tiến code trong câu 2 để ứng dụng kỹ thuật Early Stopping để giải quyết vấn đề Overfitting.

Câu 2: 6 điểm (chọn 1 trong 2 câu bên dưới)

Lựa chọn 2a: Convolutional Neural Networks (CNNs)

a) Trình bày lý thuyết

- Mô tả chi tiết kiến trúc của Convolutional Neural Networks (CNNs), bao gồm các thành phần chính: *convolutional layer*, *pooling layer*, *fully connected layer*, *activation function*, và *output layer*.
- Giải thích cơ chế hoạt động của từng thành phần trong mạng.
- Minh họa quy trình huấn luyện và suy luận (inference) của CNN bằng ví dụ cụ thể cho 01 bài toán phân loại.
- So sánh hiệu năng, ưu và nhược điểm của CNN so với Feedforward Neural Network (FNN) truyền thống.

b) Thực nghiệm và phân tích

- Xây dựng hoặc sử dụng sẵn một mô hình CNN cơ bản để giải quyết bài toán phân loại ảnh.
- Trình bày và thực hiện 02 phương pháp giảm overfitting, ví dụ: Dropout, Early Stopping, Data Augmentation, Regularization (L1/L2).
- So sánh kết quả mô hình trước và sau khi áp dụng các phương pháp trên, kèm theo phân tích định tính và định lượng (accuracy, loss, confusion matrix, v.v.).

c) Nghiên cứu mở rộng

- Tìm hiểu và thử nghiệm một hướng cải tiến cho mô hình (tùy chọn), ví dụ: *Feature selection*, *Hyperparameter optimization* (grid search, random search, Bayesian optimization), *Fine-tuning mô hình pre-trained*.
- Trình bày cách thực hiện, kết quả đạt được, và đánh giá hiệu quả.

Lựa chọn 2b: Long Short-Term Memory (LSTM)

a) Trình bày lý thuyết

- Mô tả chi tiết cấu trúc của Long Short-Term Memory (LSTM): *cell state*, *input gate*, *forget gate*, *output gate*.
- Giải thích cơ chế lưu trữ và quên thông tin của LSTM so với Recurrent Neural Network (RNN) thông thường.
- Minh họa quy trình huấn luyện và dự đoán với ví dụ dữ liệu chuỗi thời gian (time series), ví dụ: dự đoán giá cổ phiếu, nhiệt độ, hoặc lượng tiêu thụ năng lượng dựa trên dữ liệu quá khứ.
- So sánh với các mô hình học máy khác, ví dụ: ARIMA, Linear Regression, Random Forest Regression, hoặc vanilla RNN.

b) Thực nghiệm và phân tích

- Triển khai hoặc sử dụng một mô hình LSTM cơ bản để dự đoán giá trị tương lai của một chuỗi thời gian thực tế.
- Áp dụng các kỹ thuật giảm overfitting, ví dụ: *Dropout*, *Regularization*, *Early Stopping*, *tăng kích thước dữ liệu*.
- So sánh kết quả trước và sau khi áp dụng các kỹ thuật này.
- Đánh giá hiệu năng bằng các chỉ số như *MSE*, *RMSE*, *MAE*, hoặc *R² score*.

c) Nghiên cứu mở rộng

- Tìm hiểu và áp dụng một cải tiến cho mô hình, chẳng hạn: *Feature selection* hoặc *dimensionality reduction*, *Hyperparameter tuning*, *Stacked LSTM* hoặc *Bidirectional LSTM*.

- Tiến hành thực nghiệm và phân tích hiệu quả của cải tiến.
