TRƯỜNG ĐẠI HỌC THỦY LỢI

**KHOA CÔNG NGHỆ THÔNG TIN**

A blue and white logo

Description automatically generated with low confidence

**BÀI TẬP LỚN**

HỌC PHẦN: HỌC MÁY

**ĐỀ TÀI: XXXXXXXXXXXXXXXXXXXXXXX**

Giáo viên hướng dẫn: xxxxx

Sinh viên/nhóm sinh viên thực hiện:

1. xxxx, lớp x1

2. yyyy, lớp y1

**Hà Nội, năm 2022**

**MỤC LỤC**

**Phần 1: Tổng quan**

***1. Giới thiệu về học máy***

* Lịch sử và vai trò của machine learning
* Ưu điểm và hạn chế của Học có giám sát và không giám sát

***2. Trình bày phương pháp học máy được sử dụng trong bài tập lớn***

* Mục đích của phương pháp.
* Input:
* Output:
* Cách thực hiện: Cách xây dựng hàm mất mát, cách giải bài toán tối ưu, nghiệm của bài toán tối ưu, ...
* Đánh giá phương pháp.

***3.Trình bày bài toán***

* Mô tả bài toán
* Input:
* Ouput:

**Phần 2: Thực nghiệm**

***1.Mô tả tập dữ liệu***

* Mô tả tập dữ liệu của bài toán, dữ liệu gồm những chiều thông tin gì (mỗi mẫu (vertor) dữ liệu có những thông tin gì), có bao nhiêu mẫu dữ liệu (ít nhất là 100 vector dữ liệu). Mô tả nhãn lớp của dữ liệu.
* Mô tả ma trận dữ liệu (X), nhãn lớp (Y).
* Chia tập dữ liệu thành 2 phần: training data dùng để huấn luyện mô hình và test data dùng để kiểm tra sự phù hợp của mô hình.

***2. Mô tả cách giải bài toán bằng phương pháp học máy***

**Trường hợp 1:** Nếu trong nhóm phương pháp học máy lựa chọn có phương pháp thành phần chính (PCA)

Bước 1: Dùng phương pháp Phân tích thành phần chính để lựa chọn tập các thuộc tính tốt nhất cho bài toán. Từ tập training data và test data ban đầu, sử dụng các thành phần chính tốt nhất đã chọn để tạo ra tập training data và test data mới.

Bước 2: Dùng phương pháp học máy còn lại trong nhóm để xây dựng mô hình và đánh giá mô hình trên tập dữ liệu mới.

**Trường hợp 2:** Nếu trong nhóm phương pháp học máy lựa chọn có phương pháp CROSS-VALIDATION

Bước 1: Chia toàn bộ tập dữ liệu thành k phần (phương pháp k-fold cross validation).

Bước 2: Chọn ngẫu nhiên k-1 phần làm training data, 1 phần còn lại làm test data. Sử dụng phương pháp học máy đã lựa chọn trên tập training data và test data để xây dựng và đánh giá mô hình. Bước 2 này được làm k lần.

Bước 3: Chọn mô hình có (train error + validation error) là nhỏ nhất.

Ví dụ:

Bước 1: Chia tập dữ liệu thành 3 tập A, B, C.

Bước 2: Huấn luyện và đánh giá mô hình trên tập training data và test data.

* Lần 1: training data: A, B; test data: C.
* Lần 2: training data: A, C; test data: B.
* Lần 3: training data: B, C; test data: A.

Trong 3 lần trên, lần nào có (train error + validation error) nhỏ nhất thì mô hình huấn luyện của lần đó được chọn làm mô hình dự đoán cho dữ liệu mới.

***3. Đánh giá mô hình***

Dùng tập test data để đánh giá chất lượng của mô hình được lựa chọn.

***4. Mô tả các chức năng của chương trình***

-Chương trình bao gồm các giao diện đồ họa

**Kết luận**

- Tóm lược các nội dung chính mà bài tập lớn làm được

**Tài liệu tham khảo**

-Các tài liệu được tham khảo trong báo cáo