[ĐẠI HỌC KHOA HỌC TỰ NHIÊN, ĐHQG-HCM

KHOA CÔNG NGHỆ THÔNG TIN

Logo

Description automatically generated

THỐNG KÊ MÁY TÍNH VÀ ỨNG DỤNG - CQ2018-22

**BÁO CÁO ĐỒ ÁN THỰC HÀNH**

**ĐỒ ÁN 2: SỬ DỤNG LINEAR REGESSION TRÊN TẬP DỮ LIỆU MNIST**

**Thông tin nhóm**

|  |  |  |  |
| --- | --- | --- | --- |
| STT | MSSV | Họ tên | Email |
| 1 | 18120009 | Vương Gia Bảo | 18120009@student.hcmus.edu.vn |
| 2 | 18120374 | Nguyễn Minh Hiếu | 18120374@student.hcmus.edu.vn |

**Phân công công việc**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| STT | MSSV | Họ tên | Nội dung công việc | Hoàn thành |
| 1 | 18120009 | Vương Gia Bảo | Tìm hiểu cách ứng HMM cho bài toán đặt ra, đọc hiểu code và viết báo cáo mục 4,5 | 100% |
| 2 | 18120374 | Nguyễn Minh Hiếu | Tìm hiểu lý thuyết HMM, mô hình HMM cho bài toán đặt ra, viết báo cáo mục 1,2,3 | 100% |

# Tập dữ liệu

MNIST là tập dữ liệu bao gồm các chữ số từ 0-9 viết tay.

Calendar

Description automatically generated

* Tập dữ liệu bao gồm tập Train (60000 ảnh) và Test (10000 ảnh).
* Mỗi ảnh là ảnh xám (1 channel) và kích thước 28x28.
* Giá trị mỗi pixels trong ản nằm trong đoạn [0:255], càng nhỏ thì càng tối và ngược lại.
* Dữ liệu có dáng chuẩn, không bị xoay ngang dọc.

# Áp dụng Linear Regession

## Ứng dụng

Trong mỗi ảnh ta có 28x28 = 784 pixels, mỗi ảnh là 1 vector feature V.

## Kết quả

### Không có tiền xử lý.

Tại đây, chúng ta để mọi thứ tự nhiên, không tác động gì:

* Kết quả:
* Nhận xét:

### Chuẩn hóa giá trị pixel [0:1]

Tại đây, chúng ta chuẩn hóa giá trị pixel về [0:1] để làm giảm sự quá nổi bật của các pixel này:

* Kết quả:
* Nhận xét: