Bài tập 3 – Machine learning (Decision tree)

Môn học: Cơ sở trí tuệ nhân tạo

*Lecturer: Nguyễn Hải Minh*

*TA: Nguyễn Thái Vũ*

1. **Giới thiệu**

* Mục tiêu của bài tập: sinh viên làm quen với các thuật toán về học máy, cụ thể là cây quyết định (decision tree).
* Thời gian: 2 tuần (chi tiết xem trên moodle)
* Ngôn ngữ lập trình: Python
* Nền tảng: jupyter notebook.
* Mô tả cụ thể, được cho trong 2 file notebook tương ứng “Lab3 – mnist - SV.ipynb” và “Lab3 – scratch – SV.ipynb”.

1. **Lab3 – scratch (50%)**

Bài tập này, rèn luyện khả năng hiểu và cài đặt thuật toán decision tree từ đầu mà không dùng thư viện hỗ trợ nào.

* Gini Index (4đ)
  + Cài đặt hàm gini (2đ)
  + Cài đặt hàm reduction\_gini\_impurity (2đ)
* Split a dataset (4đ)
  + Cài đặt hàm partition (2đ)
  + Hoàn thành hàm find\_best\_split (2đ)
* Kiểm tra trên tập kiểm thử (2đ)

1. **Lab 3 – mnist (50%)**

Bài tập, rèn luyện khả năng phân tích dữ liệu và sử dụng thuật toán decision tree để phân loại bộ dữ liệu mnist. (Do tính chất của môn học về phân tích và khai thác dữ liệu, những phần liên quan đến biểu diễn ảnh,… mình đã hỗ trợ các bạn).

* Đọc dữ liệu (2đ)
* Khám phá dữ liệu
  + Biểu diễn dữ liệu
  + Kiểm tra các lớp balance (1đ)
* Chuẩn bị dữ liệu
  + Tiền xử lý [0, 1] (1đ)
  + One hot vector (1đ)
* Cài đặt và huấn luyện Decision Tree (1đ)
* Kiểm tra độ chính xác
  + Tiền xử lý dữ liệu test (1đ)
  + Kiểm tra độ đo accuracy (1đ)
  + Confusion matrix (2đ)

1. Quy định

* Các bạn có thể tham khảo, tuy nhiên code phải là của các bạn. Nếu vi phạm sẽ bị 0đ toàn bộ môn học.
* Về quy định nộp bài: bạn tạo thư mục nộp bài theo cấu trúc sau:
  + Thư mục `MSSV` (vd, nếu bạn có MSSV là 123 thì bạn đặt tên thư mục là `123`)
    - File Lab3 – scratch - SV.ipynb
    - File Lab3 – mnist - SV.ipynb
  + Các bạn nén lại MSSV.zip (\*.zip chứ không phải .rar hay gì khác)
* Mọi thắc mắc các bạn có thể gửi vào group zalo của lớp (không gửi Zalo cá nhân) hoặc email cho mình vunguyenthai73@gmail.com