

Nguyễn Khắc Sơn - MSSV_21085691

Bài tập về nhà môn học máy – tuần 1

Bài 1: Dựa trên định nghĩa về học máy của Tom Mitchell, hãy xác định các thành phần PET (P: Performance measure, E: Experience, T: Task) cho các bài toán học sau:

- a/ Một hệ thống nhận dạng người đi bộ qua đường
- b/ Hệ thống điều chỉnh nhiệt độ trong một tòa nhà thông minh
- c/ Robot hút bụi tự động di chuyển trong nhà

Ngày: 13-8

BTVN 1

Bài 1:

Dựa trên định nghĩa về học máy của Tom Mitchell, hãy xác định các thành phần P-E-T (P - Performance, E: Experience, T: Task) của các bài toán sau:

a) Hệ thống nhận dạng người đi bộ qua đường:

- Task (T): Nhận dạng người đi bộ qua đường.
- Experience (E): Dữ liệu bao gồm hình ảnh và video của người đi bộ qua đường, người không qua đường, phương tiện khác (xe máy, xe đạp...).
- Performance measure (P): Tỷ lệ chính xác của việc nhận dạng người đi bộ, bao gồm các chỉ số như độ chính xác, độ nhạy, độ đặc hiệu, F1-score.

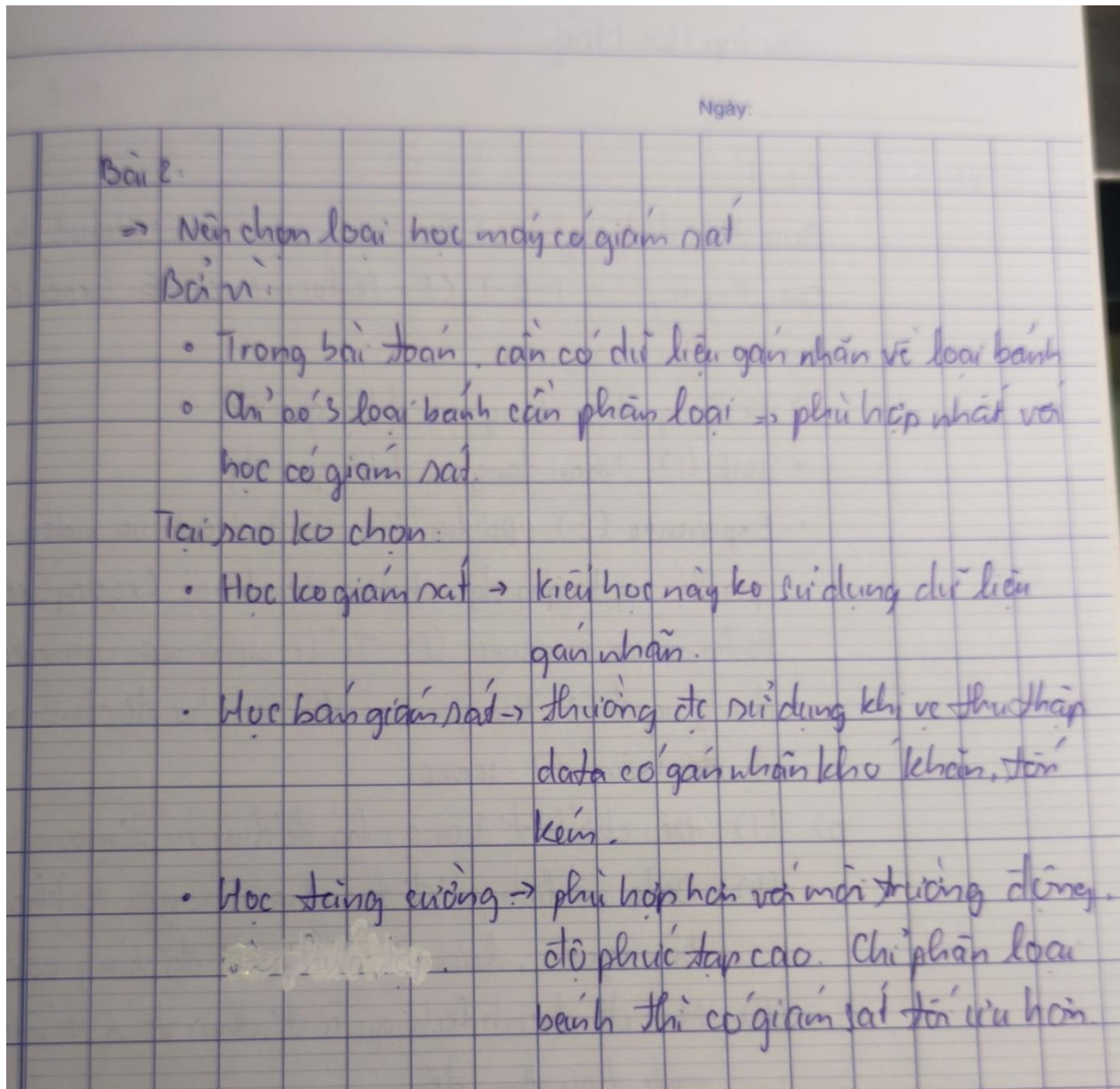
b) (T): Điều chỉnh nhiệt độ trong nhà để duy trì nhiệt độ mong muốn
(E): Dữ liệu về lịch sử nhiệt độ trong nhà, các yếu tố khác như nhiệt độ bên ngoài, lượng khách, nhiệt độ đã từng cài đặt.
(P): Mức độ chênh lệch giữa nhiệt độ thực tế và nhiệt độ mong muốn, mức tiêu thụ điện, ...

c) (T): Làm sạch bãi bẩn trong nhà
(E): Dữ liệu từ cảm biến về bãi bẩn trong nhà, vị trí bãi bẩn cần làm sạch.
(P): Diện tích nhà cần làm sạch, thời gian hoàn thành.

Bài 2

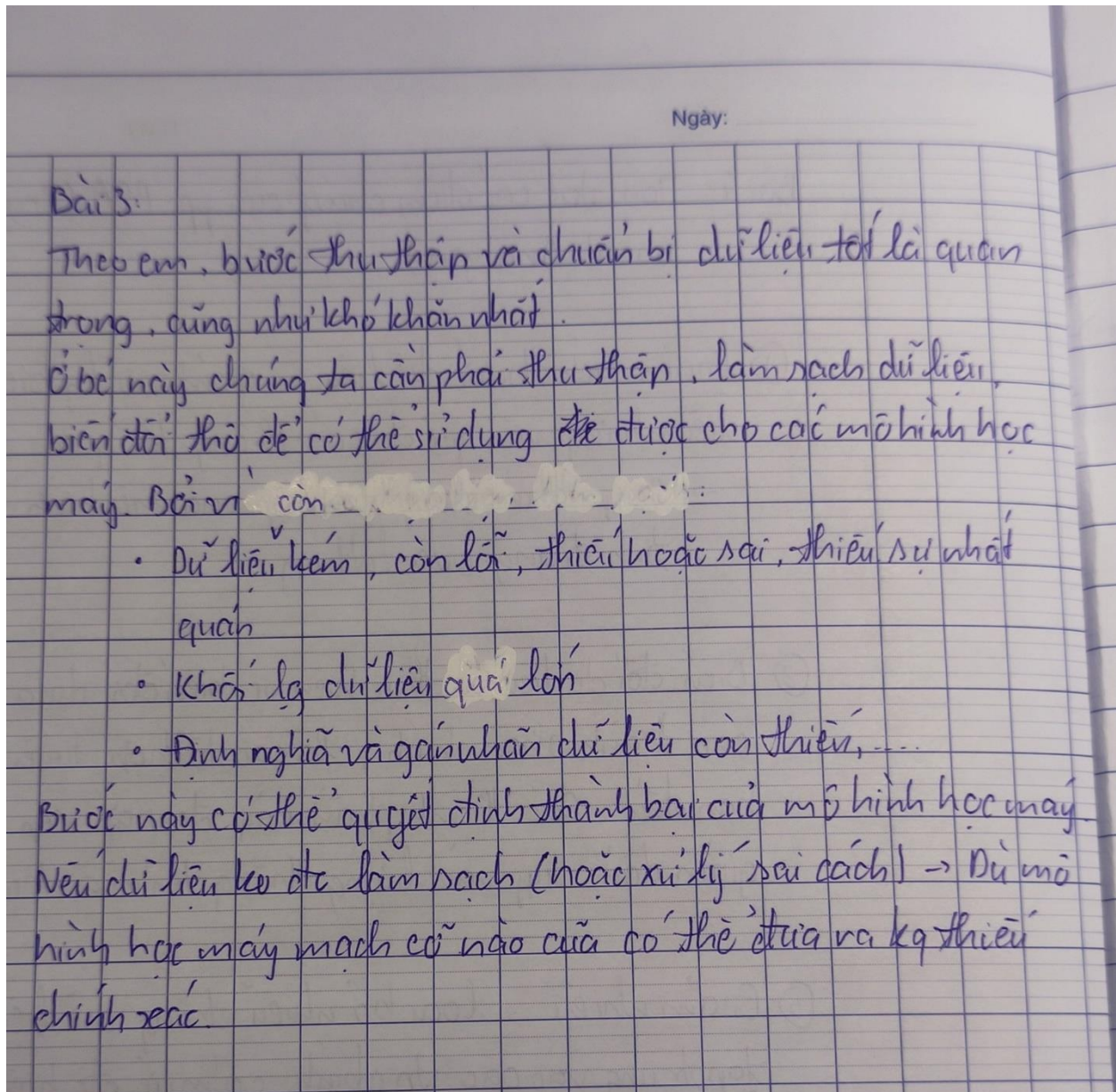
Trong một nhà máy sản xuất bánh, trên băng chuyền có ba loại bánh với hình dạng khác nhau lần lượt xuất hiện. Để thực hiện việc gấp bánh và xếp riêng vào ba thùng, một sinh viên được giao nhiệm vụ thiết kế một tay máy có ứng dụng học máy.

Áp dụng cách phân loại khi xem xét đến số lượng và loại giám sát trong quá trình huấn luyện, sinh viên đó sẽ thực hiện nhiệm vụ được giao này dựa trên loại học máy nào? Giải thích lý do.



Bài 3:

Sau bước đầu làm quen với việc xây dựng hệ thống học máy, theo quan điểm của em bước khó khăn và thách thức nhất là bước nào? Vì sao?



Bài 4: Phương pháp trích chọn đặc trưng của dữ liệu đóng vai trò quan trọng trong việc xây dựng và giải quyết vấn đề học máy. Phân tích thành phần chính, hay PCA (Principal

component analysis), là phương pháp giảm chiều thường được sử dụng để giảm chiều của các tập dữ liệu lớn, bằng cách chuyển đổi một tập hợp biến lớn thành tập hợp nhỏ hơn nhưng vẫn chứa hầu hết thông tin trong tập hợp lớn đó. Thực hiện việc tìm kiếm và tra cứu tài liệu chuyên môn. Sau đó ghi tóm tắt từ 3 đến 5 điểm đáng chú ý của phương pháp này.

Ngày:

Bài 4: Tóm tắt các điểm chính của pp PCA (Principal Component Analysis)

- ① Giảm kích thước dữ liệu: giảm số lg biến trong dữ liệu nhưng vẫn giữ được các thông tin quan trọng
→ Giúp giảm độ phức tạp của mô hình và tránh tình trạng overfitting
- ② Biến đổi dữ liệu: thực hiện vt biến đổi data gốc → tập hợp các thành phần chính, là các tổ hợp tuyến tính của biến ban đầu. Các thành phần này được sắp xếp theo thứ tự giảm dần, tùy vào lượng thông tin mà chúng mang lại
- ③ Giảm nhiễu: loại bỏ nhiễu trong dữ liệu bằng cách tập trung vào các tp chính, có mức độ biến động cao, loại bỏ các tp biến thiên thấp.
- ④ Tiền xử lý dữ liệu: để PCA hoạt động hiệu quả, dữ liệu cần được chuẩn hóa hoặc chuẩn chính, tách biệt với các biến có độ đo lường khác nhau.

Từ khóa:

• Dimensionality Reduction	• Noise Reduction
• Data Transformation	• Data Preprocessing