

APPLIED LINEAR ALGEBRA FOR IT

ĐẠI SỐ TUYẾN TÍNH ỨNG DỤNG CHO CNTT

501032

Anh H. Vo

1 Giới thiệu

Chọn nghề và chọn ngành học phù hợp là một trong những vấn đề đang được xã hội quan tâm trong thời điểm hiện tại. Hàng năm, tỉ lệ sinh viên bị xử lý học vụ ở các trường đại học cho thấy một con số đáng báo động, nhiều sinh viên phát hiện mình không phù hợp ngành đang học khi đã quá muộn. Việc chọn ngành không phù hợp gây lãng phí về thời gian, công sức và tài chính của sinh viên và gia đình. Có nhiều nguyên nhân dẫn đến tỉ lệ sinh viên bị xử lý học vụ, ngoài các yếu tố chủ quan còn những yếu tố khách quan, mà một trong số đó là việc sinh viên không được tư vấn kịp thời để đưa ra những quyết định đúng đắn trong quá trình học. Một trong những giải pháp cấp thiết là xây dựng một *hệ thống tư vấn học tập thông minh* có thể gợi ý giúp sinh viên lựa chọn môn học phù hợp với năng lực. Hệ thống tư vấn học tập thông minh bao gồm các chức năng cơ bản được trình bày trong phần 3 và dữ liệu được mô tả thành các ma trận như trong phần 2

2 Mô tả dữ liệu

Ma trận R thể hiện kết quả học tập của m sinh viên tương ứng với 6 môn học, nếu sinh viên i chưa học môn học j thì kết quả $R(i, j) = \infty$. Trong đó, ma trận R của mỗi sinh viên nhận được trong tập tin *ques_MSSV.dat* sẽ có giá trị khác nhau, phụ thuộc vào chiều dài tên gọi của từng sinh viên.

$$R = \left(\begin{array}{c|cccccc} & C_1 & C_2 & C_3 & C_4 & C_5 & C_6 \\ \hline S_1 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ S_i & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ S_m & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{array} \right)$$

Tương tự, ma trận M thể hiện chuyên ngành mà mỗi sinh viên đăng ký ban đầu, với m thể hiện số lượng sinh viên và số lượng cột thể hiện số lượng chuyên ngành trong chương trình đào tạo. Nếu sinh viên i đăng ký chuyên ngành M_1 thì dòng $M(i, 1)$ có giá trị là 1 và $M(i, 2)$ có giá trị là 0. Mỗi sinh viên được cung cấp

ma trận M trong tập tin *ques_MSSV.dat*.

$$M = \left(\begin{array}{c|cc} & M_1 & M_2 \\ S_1 & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ S_i & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ S_m & \vdots & \vdots \end{array} \right)$$

Ma trận C thể hiện môn học i thuộc chuyên ngành nào và số tín chỉ tương ứng với môn học đó (trong đề bài này, ma trận C cố định).

$$C = \left(\begin{array}{c|ccc} & M_1 & M_2 & \text{credit} \\ C_1 & 0 & 1 & 3 \\ C_2 & 0 & 1 & 4 \\ C_3 & 1 & 0 & 3 \\ C_4 & 1 & 0 & 3 \\ C_5 & 0 & 1 & 3 \\ C_6 & 1 & 0 & 4 \end{array} \right)$$

3 Yêu cầu

Ví dụ giả sử rằng hệ thống cần quản lý 4 sinh viên đã có kết quả học tập được thể hiện qua ma trận R như bên dưới:

$$R = \left(\begin{array}{c|cccccc} & C_1 & C_2 & C_3 & C_4 & C_5 & C_6 \\ S_1 & \infty & \infty & 3 & \infty & 5 & 0 \\ S_2 & 2 & \infty & 5 & \infty & 7 & 0 \\ S_3 & 5 & 7 & 0 & \infty & 8 & 9 \\ S_4 & \infty & \infty & \infty & 7 & \infty & 8 \end{array} \right)$$

Với các chuyên ngành được sinh viên đăng ký ban đầu thể hiện trong ma trận M như sau:

$$M = \left(\begin{array}{c|cc} & M_1 & M_2 \\ S_1 & 1 & 0 \\ S_2 & 0 & 1 \\ S_3 & 1 & 0 \\ S_4 & 1 & 0 \end{array} \right)$$

Yêu cầu 1: Viết lệnh tính điểm trung bình học tập của từng sinh viên trong ma trận R , công thức tính được thể hiện như bên dưới.

$$GPA_k = \frac{\sum_i^l credit_i \times point_i}{\sum_i^l credit_i} \quad (1)$$

Trong đó GPA_k là điểm trung bình học tập của sinh viên k ($1 \leq k \leq m$), m và l lần lượt là số lượng sinh viên cần phải quản lý và số lượng môn học mà sinh viên k đã học, $credit_i$ là số lượng tín chỉ tương ứng với môn học i , $point_i$ là điểm mà sinh viên đạt được khi tham gia môn học thứ i . Kết quả trả về là một vec-tơ cột có m phần tử. Lưu kết quả tính được vào biến **a.Req1** (1 điểm)

Ví dụ:

Kết quả trả về:

$$G\vec{P}A = \begin{pmatrix} 2.4 \\ 3.2 \\ 6.1 \\ 7.5 \end{pmatrix}$$

Yêu cầu 2: Viết lệnh tìm tất cả các chuyên ngành có nhiều sinh viên bị cảnh báo học vụ nhất trả về một vec-tơ dòng có k thành phần, trong đó (k tương ứng với số lượng chuyên ngành tìm được), nếu không có bất kỳ chuyên ngành nào có sinh viên bị cảnh báo học vụ thì kết quả trả về 0.

Gợi ý: Sinh viên bị cảnh báo học vụ khi có điểm trung bình học tập dưới 5.0

Lưu kết quả tính được vào biến **a.Req2** (1 điểm)

Ví dụ:

Kết quả trả về:

$$\vec{F} = \begin{pmatrix} 1 & 2 \end{pmatrix}$$

Yêu cầu 3: Viết lệnh dự đoán kết quả học tập cho các môn sinh viên chưa học trong ma trận \mathbf{R} , kết quả trả về là một ma trận có kích thước $m \times n$ cùng kích thước với ma trận \mathbf{R} . Lưu kết quả tính được vào biến **a.Req3** (2 điểm)

Gợi ý: Sử dụng phương pháp phân tích ma trận thành phần (Singular Value Decomposition)

$$\mathbf{R}_{m \times n} = \mathbf{U}_{m \times m} \mathbf{\Sigma}_{m \times n} \mathbf{V}_{n \times n}^T$$

Trong đó \mathbf{V}^T (right-singular value) là ma trận chứa các vec-tơ trị riêng của ma trận $\mathbf{R}^T \mathbf{R}$. $\mathbf{\Sigma}$ là ma trận đường chéo chứa các giá trị riêng (singular value), \mathbf{U} (left-singular value) là ma trận chứa các vec-tơ trị riêng của ma trận $\mathbf{R} \mathbf{R}^T$.

Sinh viên thực hiện bước tiền xử lý ma trận \mathbf{R} (sinh viên tự tìm hiểu) trước khi thực hiện chuẩn hóa

Chuẩn hóa: $\hat{\mathbf{R}}_{m \times n} = \mathbf{R}_{m \times n} - \bar{\mathbf{R}}_{m \times n}$ ($\bar{\mathbf{R}}_{m \times n}$ là ma trận trung bình các môn học sinh viên đã học)

$\hat{\mathbf{R}}_{m \times n} \approx \mathbf{U}_{m \times k} \sqrt{\mathbf{\Sigma}_{k \times k}} \sqrt{\mathbf{\Sigma}_{k \times k}} \mathbf{V}_{k \times n}^T$ (Sinh viên lựa chọn $k = 3$)

$$\hat{\mathbf{R}}_{m \times n} = \hat{\mathbf{R}}_{m \times n} + \bar{\mathbf{R}}_{m \times n}$$

(Lưu ý rằng, điểm môn học chỉ nằm trong khoảng từ 0-10 và điểm các môn học i sinh viên k đã học không thay đổi)

Ví dụ:

Kết quả trả về:

$$\hat{\mathbf{R}} = \begin{pmatrix} \begin{matrix} S_1 \\ S_2 \\ S_3 \\ S_4 \end{matrix} & \begin{matrix} C_1 & C_2 & C_3 & C_4 & C_5 & C_6 \end{matrix} \\ \begin{matrix} 2.5 & 4.5 & 3 & 5.2 & 5 & 0 \\ 2 & 5.1 & 5 & 7.9 & 7 & 0 \\ 5 & 7 & 0 & 6.5 & 8 & 9 \\ 3.8 & 8.0 & 2.8 & 7 & 8.7 & 8 \end{matrix} \end{pmatrix}$$

Yêu cầu 4: Từ kết quả học tập dự đoán ở **a.Req3**, hãy viết lệnh cho biết khả năng những sinh viên nào có thể bị cảnh báo học vụ, trả về một vec-tơ cột chứa danh sách sinh viên. Kết quả tìm được lưu vào biến **a.Req4** (1 điểm)

Ví dụ:

Kết quả trả về:

$$\vec{F}^{pred} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

Yêu cầu 5: Tương tự, dựa trên kết quả học tập đã được dự đoán ở **a.Req3**, viết lệnh để gợi ý cho sinh viên có mã số **id** (sinh viên xem trong tập tin `ques_MSSV.dat` được cung cấp riêng để tìm thông tin `id_query` phù hợp với yêu cầu của mình, nếu sinh viên có `id_query` là 0 thì chỉnh lại là 1):

- Những môn học sinh viên không nên đăng ký trong kế hoạch học tập của mình. Nếu sinh viên có khả năng học được tất cả các môn hoặc sinh viên đã học hết tất cả các môn, kết quả trong trường hợp này trả về 0, ngược lại trả về một vec-tơ dòng chứa danh sách các môn sinh viên không nên đăng ký trong kế hoạch. Kết quả tìm được lưu vào biến **a.Req5a** (1 điểm)
- Những môn học sinh viên nên đăng ký trong kế hoạch học tập của mình. Nếu sinh viên không có khả năng học được bất kỳ môn nào hoặc sinh viên đã học hết tất cả các môn, kết quả trong trường hợp này trả về 0, ngược lại trả về một vec-tơ dòng chứa danh sách các môn sinh viên nên đăng ký trong kế hoạch. Kết quả tìm được lưu vào biến **a.Req5b** (1 điểm)

Gợi ý:

- Sinh viên nên chọn những môn mình có khả năng học từ 5 điểm trở lên.
- Sinh viên không nên chọn những môn mình có khả năng học dưới 5 điểm.

Ví dụ:

Kết quả trả về: `id = 3`

$$\vec{C}^{pred} = 0$$
$$\vec{C}^{pred} = \begin{pmatrix} 4 \end{pmatrix}$$

Yêu cầu 6: Viết lệnh để giúp sinh viên lựa chọn lại chuyên ngành dựa trên điểm trung bình học tập theo từng chuyên ngành của mỗi sinh viên trong ma trận **R**, kết quả trả về là một ma trận có kích thước bằng với ma trận **M**. Kết quả tìm được lưu vào biến **a.Req6** (1 điểm)

Gợi ý: Nếu điểm trung bình học tập của sinh viên *k* theo chuyên ngành *i* cao hơn điểm trung bình học tập theo chuyên ngành *j* thì hệ thống đưa ra kết quả dự đoán sinh viên chọn học chuyên ngành *i*. Ngược lại, nếu điểm trung bình học tập của sinh viên *k* theo chuyên ngành *j* cao hơn chuyên ngành *i* thì hệ thống dự đoán sinh viên chọn học chuyên ngành *j*. Nếu xảy ra trường hợp sinh viên *k* có kết quả trung bình học tập theo cả hai chuyên ngành bằng nhau thì giữ nguyên theo chuyên ngành cũ mà sinh viên đã chọn trước đó. Điểm trung bình học tập theo chuyên ngành *i* được tính như công thức (1), tuy nhiên *l* thể hiện số lượng môn học theo chuyên ngành *i* mà sinh viên đã học, trường hợp sinh viên *k* chưa học bất kỳ môn học nào thuộc chuyên ngành *i* thì điểm trung bình học tập theo chuyên ngành *i* là 0.

Ví dụ:

Kết quả trả về:

$$\mathbf{M}^{pred} = \left(\begin{array}{c|cc} & M_1 & M_2 \\ S_1 & 0 & 1 \\ S_2 & 0 & 1 \\ S_3 & 0 & 1 \\ S_4 & 1 & 0 \end{array} \right)$$

Yêu cầu 7: Viết lệnh thống kê (%) số lượng sinh viên lựa chọn chuyên ngành không đúng với kết quả học tập hiện tại của sinh viên thể hiện thông qua ma trận **R**. Kết quả tìm được lưu vào biến **a.Req7** (1 điểm)

Ví dụ:

Kết quả trả về: $w = 50$

Yêu cầu 8: Viết lệnh trả về danh sách những sinh viên nào hiện đang lựa chọn chuyên ngành không đúng với kết quả học tập hiện tại đang có trong ma trận **R**, kết quả trả về một vec-tơ cột. Kết quả tìm được lưu vào biến **a.Req8** (1 điểm)

Ví dụ:

Kết quả trả về: $\vec{w} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$

Chú ý:

- Trong nội dung đề bài này, sinh viên chỉ được phép sử dụng một số hàm có sẵn trong MATLAB như sau:

- `sum(.)`
- `min(.)`, `max(.)`
- `find(.)`, `round(.)`, `size(.)`, `length(.)`, `zeros(.)`, `repmat(.)`
- `svd(.)`

Trường hợp sinh viên sử dụng các hàm ngoài danh sách các hàm được cho phép, sinh viên phải tự chịu trách nhiệm nếu nó ảnh hưởng đến kết quả bài làm của mình (những hàm không thuộc danh mục được cho phép bên trên sẽ được đóng lại trước khi chạy chấm bài).

- Sinh viên phải sử dụng hàm `round(z,1)` để làm tròn số thực z trước khi gán vào các biến được yêu cầu.
- Sinh viên dùng lệnh `load(tên_tập_tin.mat tương ứng)` để lấy dữ liệu được cung cấp trong tập tin `ques_MSSV.dat`, sinh viên không đổi tên biến sau khi load (biến **A** được hiểu như ma trận **R**).
- Sinh viên phải đảm bảo mã nguồn viết chuẩn có thể chạy được trên tất cả phiên bản Matlab, kết quả các ví dụ minh họa được thực hiện trên phiên bản Matlab ở phòng thực hành của trường.

4 Nội bài

- Tạo thư mục tên `Nx_SSSStudentID` và lưu trữ tập tin bài làm của mình có tên `SSStudentID.m` vào thư mục này, trong đó `Nx` là số thứ tự nhóm sinh viên đang học.
- Thư mục `Nx_SSSStudentID` được nén thành `Nx_SSSStudentID.zip` trước khi nộp.
- Sinh viên phải tự làm bài của mình. Tất cả những trường hợp tham khảo mã nguồn từ sinh viên khác hoặc từ các nguồn khác đã được công bố trước đó thì sinh viên sẽ nhận 0 điểm quá trình 2.

- Sinh viên mặc định sẽ nhận điểm 0 nếu chương trình không chạy được.
- Hạn nộp: **23/5/2021**