

Data Mining And Data Analysis Assignment 1

% problem 1

We define a model to estimate the probability of a certain question in one of the three subjects(Chinese, English, Math) being answered correctly with certain ability level.

3PL model is defined as: $p(i, j) = c_j + \frac{1 - c_j}{1 + e^{-1.7a_j * (\theta_i - b_j)}}$

Note: for each subject, the data has two aspects,

questions in three dimensions(a, b, c)

ability level(θ)

Hint: $p(i, j)$ represents the probability of the j^{th} question in subject a, b or c (i.e., a_j, b_j, c_j) being answered correctly with ability level i (i.e., θ_i).

% Design & Implementations

首先从用户得到输入的 i, j 以及需要计算的subject。为了提升程序的健壮性，我对用户输入进行了输入检查,判断用户的输入是否是一个合法的值然后再进行计算。

根据用户给的 i, j 和subject到给出的data_prob1.mat中导入相对应的 a, b, c ，以及 θ 。

将得到的 a, b, c 和 θ 代入 3PL Model 的计算公式算出 $p(i, j)$ 。

将 p 输出到Console给用户。

% Source code & Result

Assignment1.m :

```

1 %get user's input
2 i=input('Please give the first parament i: ','s');
3 i = isInt(i);
4
5 j=input('Please give the second parament j: ','s');
6 j = isInt(j);
7
8 prompt = 'What is the subject ?\n 1, Chinese 2, Englsih 3, Math\n';
9 subject = input(prompt,'s');
10 subject = isInt(subject);
11 while subject >3 || subject<1 ,
12     subject = input('Invaild input, please choose number from 1,2,3: ','s');
13     subject = isInt(subject);
14 end
15
16 % use i,j subject get a,b,c and theta.
17 load('data_prob1.mat');
18 [~,~] = size('data');
19 a = data(subject).par(j,1);
20 b = data(subject).par(j,2);
21 c = data(subject).par(j,3);
22 theta = data(subject).theta(i);
23 disp('a[j],b[j],c[j],theta[i] equals: ');
24 disp(a);
25 disp(b);
26 disp(c);
27 disp(theta);
28
29 % call the model function:
30 p = ModelFunc(a,b,c,theta);
31
32 %return the results:
33 disp('the result p(i,j) is :');
34 disp(p);

```

isInt.m :

```

1 function retval = isInt(val)
2 [val, status] = str2num(val);
3 while ~status || ~isscalar(val) || val ~= floor(val)
4     val = input('Invalid input. Please input an integer, try again: ', 's');
5     [val, status] = str2num(val);
6 end
7 retval = val;
8 end
9
10

```

ModelFunc.m :

```

1 function [ p ] = ModelFunc( a,b,c,theta)
2
3 p = c + ((1-c) / (1 + (exp(1)^(-1.7*a*(theta-b)))));
4 end
5

```

Result:

(To make the program more stronger, I added the input check.)

```
>> Assignment1
Please give the first parament i:  df
Invalid input. Please input an integer, try again: 5
Please give the second parament j:  mmda
Invalid input. Please input an integer, try again: 6
What is the subject ?
    1, Chinese 2, Englsih 3, Math
asdf
Invalid input. Please input an integer, try again: 5
Invaild input, please choose number from 1,2,3:  2
a[j],b[j],c[j],theta[i] equals:
    2.3389

    -0.5709

    0.1978

    -3.6900

the result p(i,j) is :
    0.1978

fx >>
```

% problem 2

>>Given the .PAR and .PH3 files, please write scripts to read the files and save the parameters a, b, c and θ for each question as .MAT file.

>>Calculate the probabilities for the defined 3PL model function using the parameters from the PAR and PH3 files.

>>Plot the results, please specify the title, x-label, y-label, etc.

>> save the results into an excel file.

Note: for the .PAR file, take the column 3, 5, 9 as the parameter a, b, c. For .PH3 file, extract the POINT values as the parameter θ .

% Design & Implementations

首先要将.PAR 和.PH3文件中的数据导入到程序中。这里使用了相对路径，如果想从其他地方导入数据，为了减少数据路径的次数以便降低输入路径出错的概率，可以直接将修改loadData-path的值，然后取消下面的注释即可。

观察这两个文件发现，数据并不是非常的干净，有很多的多余数据数据需要进行删除，只保留需要的数据。通过比对字符串删除无关的数据最终得到需要的数据，然后把这些数据导出为标准的.mat文件。

然后用两层for loop 将上一步得到的数据代入3PL Model Func 计算得到所有的possibility，并用矩阵保存这些possibility。

将possibility作为y轴， theta作为x轴，利用上一步保存的possibility矩阵和读入的theta数据画出图形。为了使得到的图形更加美观一些，插入了x-label， y-label， title等标签。

将得到的possibility矩阵用 xlswrite函数输出到excel文件。

% Source code:

Assignment2.m:

Data Mining & Data Analysis

```

1 % get data from the data path
2 % -----
3 % Notice here:If you wanna load data from other places
4 % you can just add the data path to loadDataPath,
5 % and cancel the following comments ,
6 % then you will run the script perfectly.
7 % -----
8 %loadDataPath = '*****';
9 %
10 realChineseParData = strcat('','Chinese/C951C.PAR');
11 realEnglishParData = strcat(loadDataPath,'English/C951E.PAR');
12 realMathParData = strcat(loadDataPath,'Math/C951M.PAR');
13 realChinesePH3Data = strcat(loadDataPath,'Chinese/C951C.PH3');
14 realEnglishPH3Data = strcat(loadDataPath,'English/C951E.PH3');
15 realMathPH3Data = strcat(loadDataPath,'Math/C951M.PH3');
16 %}
17
18 % read .par data and change it to .mat
19 readPARData('Chinese/C951C.PAR','CS','ChinesePar.mat');
20 readPARData('English/C951E.PAR','ES','EnglishPar.mat');
21 readPARData('Math/C951M.PAR','MS','MathPar.mat');
22
23 % read .ph3 data and get the theta
24 readPH3Data('Chinese/C951C.PH3','ChineseTheta.mat');
25 readPH3Data('English/C951E.PH3','EnglishTheta.mat');
26 readPH3Data('Math/C951M.PH3','MathTheta.mat');
27
28 % calculate every param
29 [ChineseX,ChineseY] = CalculateData('ChinesePar.mat','ChineseTheta.mat');
30 [EnglishX,EnglishY] = CalculateData('EnglishPar.mat','EnglishTheta.mat');
31 [MathX,MathY] = CalculateData('MathPar.mat','MathTheta.mat');
32
33 % Plot
34 hFig = figure('numbertitle','off','name','3PL Model for Chinese English Math','color','white');
35 set(hFig,'Position',[100,100,1600,1000]);
36
37 subplot(1,3,1);
38 plot(ChineseX,ChineseY);
39 PlotGood('Chinese 3PL Model');
40
41 subplot(1,3,2);
42 plot(EnglishX,EnglishY);
43 PlotGood('English 3PL Model');
44
45 subplot(1,3,3);
46 plot(MathX,MathY);
47 PlotGood('Math 3PL Model');
48
49 saveas(hFig, '3PLModelFigure', 'png');
50
51 %export to excel
52 xlswrite('MathResultData.xls',MathY);
53 xlswrite('EnglishResultData.xls',EnglishY);
54 xlswrite('ChineseResultData.xls',ChineseY);

```

readPARData.m:

```

1  function readPARData( dataPath,filterName,matName )
2  -   txtName = 'temp.txt';
3  -   fid = fopen(dataPath);
4  -   fileID = fopen(txtName,'w');
5  -   while 1
6  -       tempLine = fgetl(fid);
7  -       if ~ischar(tempLine), break, end
8  -       if strfind(tempLine,filterName),
9  -           fprintf(fileID,'%s\n',tempLine),
10 -           end
11 -       end
12 -       fclose('all');
13 -       temp = importdata(txtName);
14 -       A = temp.data;
15 -       a = A(:,3);
16 -       b = A(:,5);
17 -       c = A(:,9);
18 -       save(matName,'a','b','c');
19 -       delete(txtName);
20 -   end

```

PlotGood.m:

```

1  function PlotGood(titleName)
2  -   %PLOTGOOD Summary of this function goes here
3  -   % Detailed explanation goes here
4  -   xlabel('Theta','FontSize',15,'FontWeight','bold');
5  -   ylabel('Possibility','FontSize',15,'FontWeight','bold');
6  -   title(titleName,'FontSize',17,'FontWeight','bold');
7  -   end
8

```

readPH3Data.m:

```

1  function readPH3Data(loadPath,thetaDataPath )
2  -   fid = fopen(loadPath,'rt');
3  -   fileId = fopen('tempTxtName.txt','w');
4  -   tempX = [];
5  -   while 1
6  -       tempLine = fgetl(fid);
7  -       if ~ischar(tempLine), break, end
8  -       if strfind(tempLine,'POINT'),
9  -           if strfind(tempLine,'TEST'),
10 -               else
11 -                   tempLine = tempLine(8:end),
12 -                   fprintf(fileID,'%s\n',tempLine),
13 -               end
14 -           end
15 -       end
16 -       fclose('all');
17 -       temp = importdata('tempTxtName.txt');
18 -       for i = 1:8
19 -           for j = 1:5
20 -               tempX = [tempX temp(i,j)];
21 -           end
22 -       end
23 -       disp(tempX);
24 -       delete('tempTxtName.txt');
25 -       save(thetaDataPath,'tempX');
26 -   end
27

```

CalculateData.m:

```

1  function [X,Y]= CalculateData(ParDataPath,ThetaDataPath)
2  -   %PLOTIMG Summary of this function goes here
3  -   % Detailed explanation goes here
4  -   load(ParDataPath);
5  -   load(ThetaDataPath);
6  -   [rlen,~] = size(a);
7  -   [~,thetaLength] = size(tempX);
8  -   testX = zeros(1,40);
9  -   testY = [];
10 -   disp(rlen);
11 -   disp(thetaLength);
12 -   for i = 1:rlen
13 -       for x = 1:thetaLength
14 -           p = ModelFunc(a(i),b(i),c(i),tempX(x));
15 -           testY(i,x) = p;
16 -       end
17 -   end
18 -   for i = 1:40
19 -       testX(1,i) = tempX(i);
20 -   end
21 -   X = testX;
22 -   Y = testY;
23 -   end
24

```

% Result:

