1. Use **\*\*logistic + lasso regression\*\*** with **\*\*10-fold cross-validation\*\*** to

identify useful predictors. Plot a lasso plot with readable tick labels on the X and

Y coordinates in your plot for easy visualization and verification. Missing clear

and readable tick labels in your plot will cost you significant points for this

assignment.

%Read the file

[X] = xlsread('CellDNA.xls');

%Read the predictors and response variables

Y = X(:,14);

X(:,14) = [];

%Convert the response variables into non-interesting = 0 and interesting=1 value

for i = 1:length(Y)

if Y(i,1) > 0

Y(i,1) = 1;

else

Y(i,1) = 0;

end

end;

% Standardize

ZX = [zscore(X)];

[b ,fitinfo] = lassoglm(X, Y, 'poisson', 'CV', 10,'Alpha',1);

%Plot graph

lassoPlot(b,fitinfo,'PlotType', 'Lambda' ,'XScale','log');

lassoPlot(b,fitinfo,'PlotType','CV');

lassoPlot(b,fitinfo,'PlotType','L1');

minpts = find(b(:,fitinfo.IndexMinDeviance))

bestValue = find(fitinfo.Lambda == fitinfo.LambdaMinDeviance);

fitinfo.Lambda

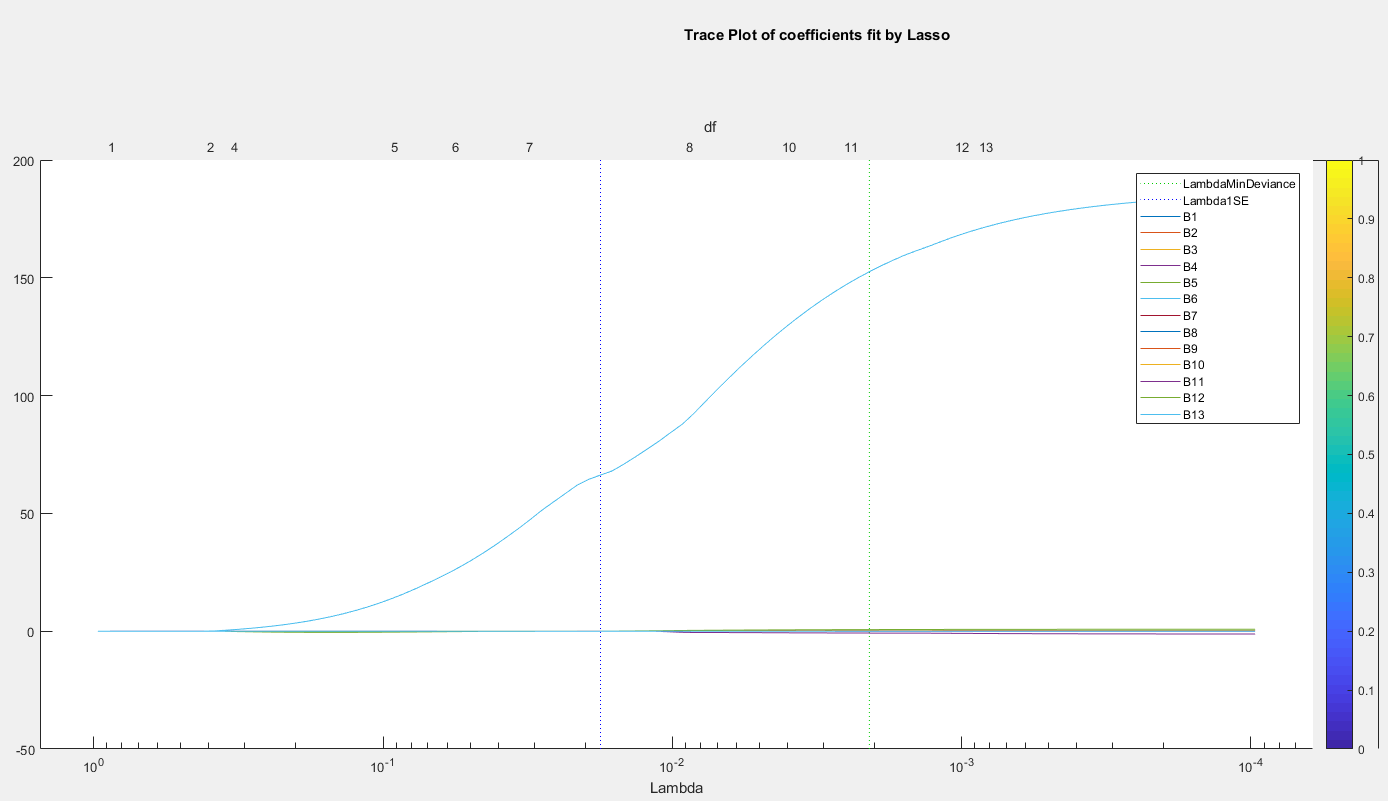
pcolor(b)

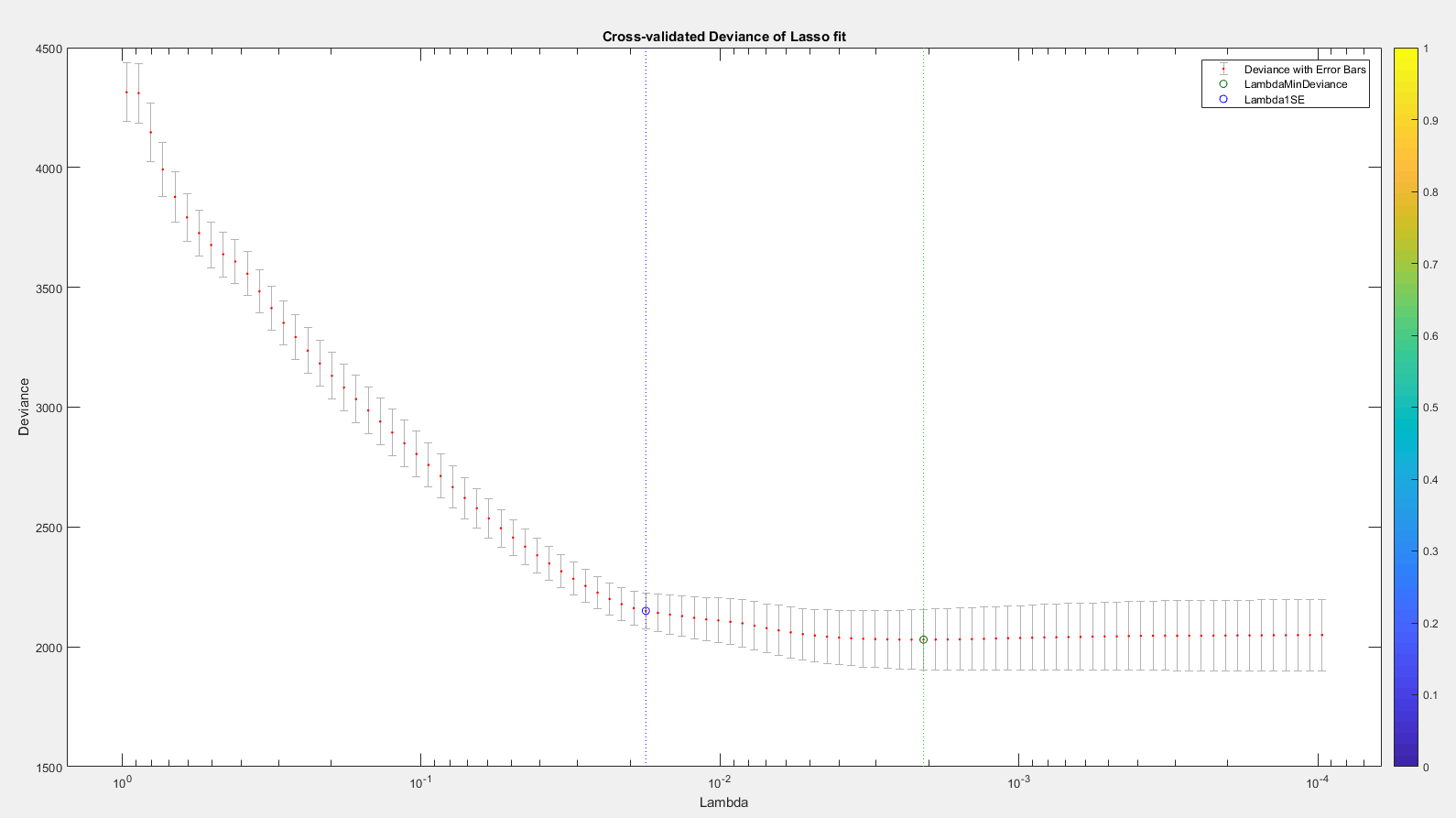
mesh(b)

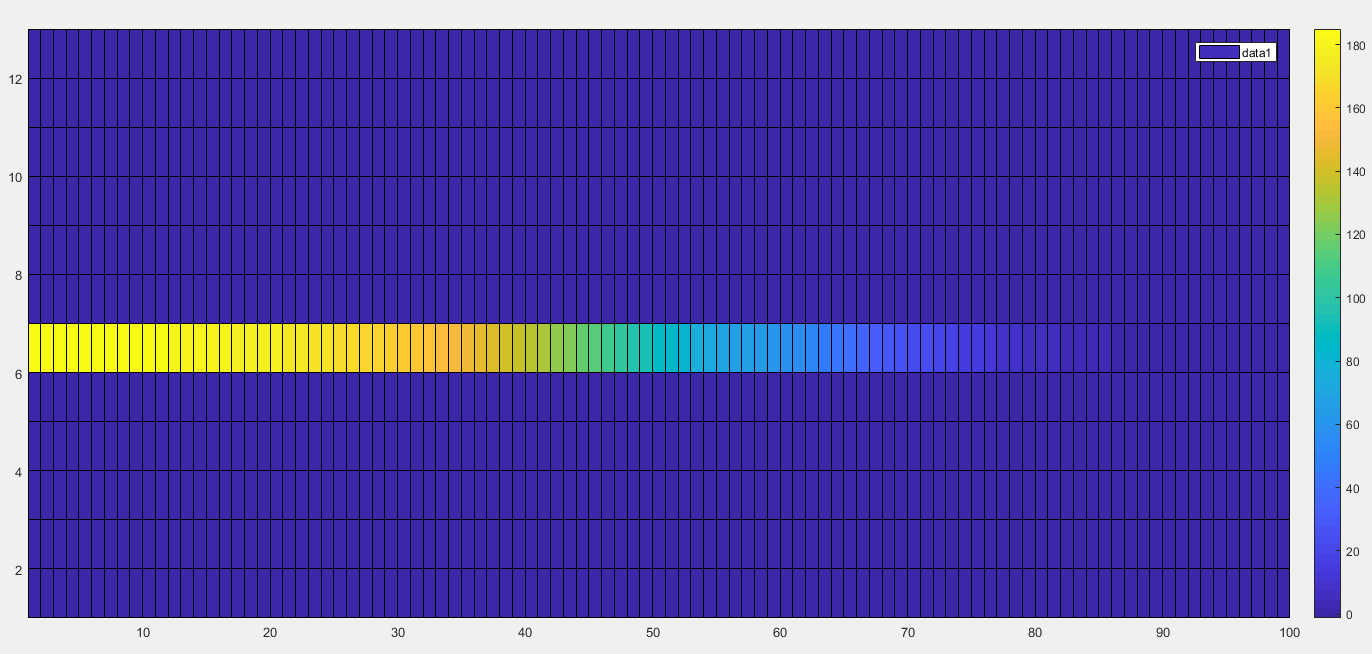
pcolor(b), colorbar

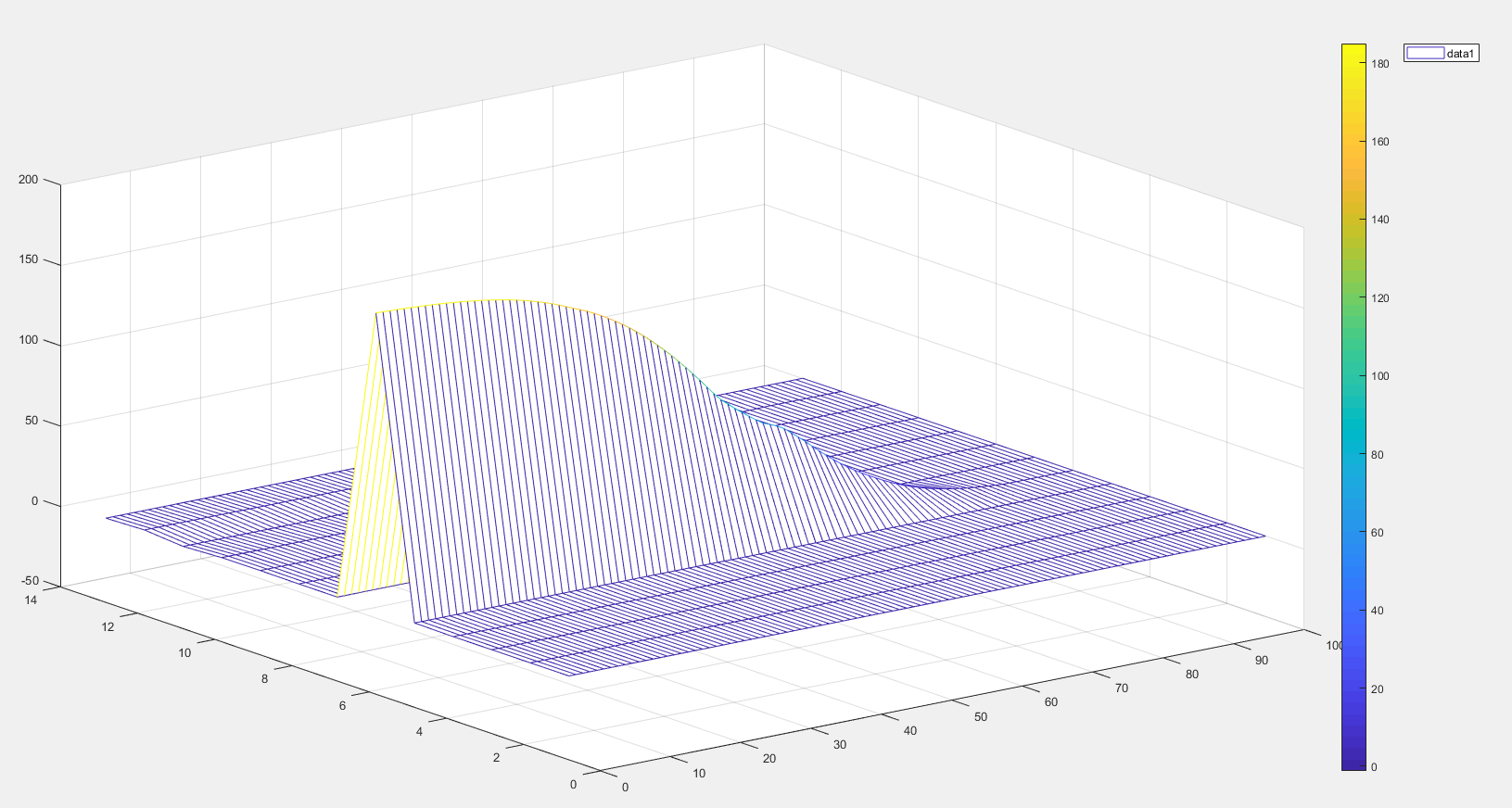
[b fitinfo] = lasso(ZX, Y, 'CV',10, 'Alpha', 1);

lassoPlot(b,fitinfo,'PlotType', 'Lambda','XScale','log');







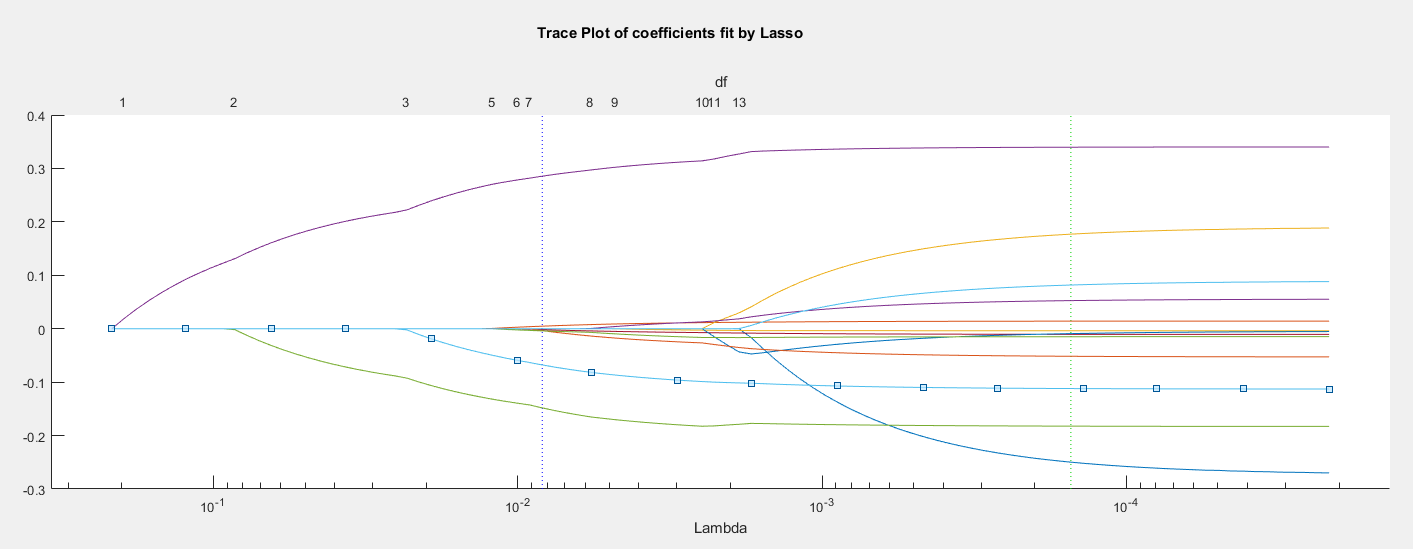


1. **Which top \*\*Three\*\* predictors are you going to select to explain why a bacteria**

**is an “interesting” candidates for further study?**

**Solution:**

Column 6, 5 and 4

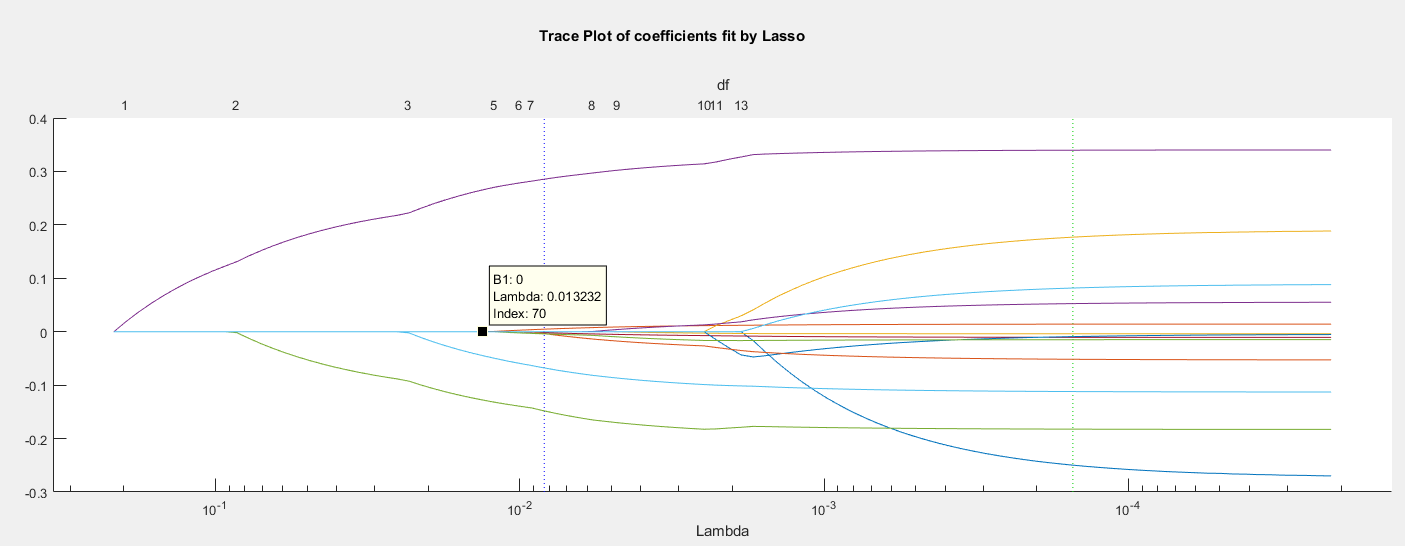


1. **What is the lambda (l) value you choose in order to select the top 3 predictors**

**you identified in the last question?**

**Solution:**

Lambda = 0.013232



1. **What are the q values for the 3**

**Solution:**

Column 4 = 0.2655

Column 5 = -0.1279

Column 6 = -0.0448

