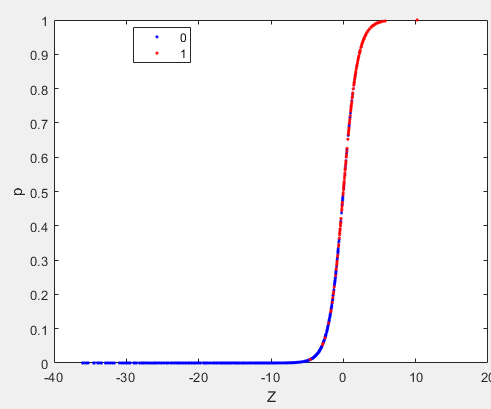
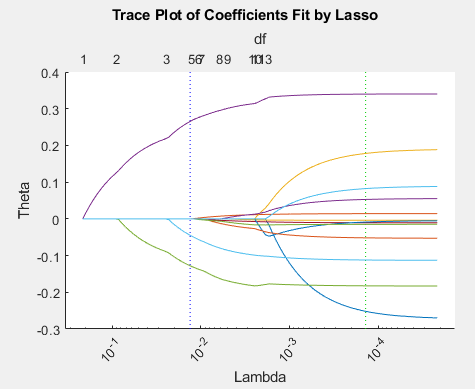
Robert Krall

SEIS 763 – Machine Learning

Assignment 4

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

identify useful predictors.

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

is an “interesting” candidates for further study?

B4, B5, B6

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

you identified in the last question?

0.015

1. What are the **Θ** values for the 3 selected predictors at the lambda (l) value you

identified in the last question?

0.2578

-0.1216

-0.0372

%{

Robert Krall

SEIS 763 - Machine Learning

Assignment 4

%}

%load CellDNA

opt = statset('UseParallel',true);

%Step 1: convert table to array and grab X and y values

CellDNA\_tbl = table2array(CellDNA);

%Step 2. Getting X and y values

X = CellDNA\_tbl(:,1:13);

%convert values to 1 or 0s

y = CellDNA\_tbl(:,14);

y(y>0)=1;

%Step 3: Standardize the data

[ZX, me, signma] = zscore(X);

%Step 4: Logistic Regression Algo

lrm = fitglm(ZX, y, 'distr', 'binomial', 'link', 'logit');

p = lrm.Fitted.Response;

Z = lrm.Fitted.LinearPredictor;

figure, gscatter(Z,p, y, 'br');

%plotSlice(lrm)

[B, FitInfo] = lasso(ZX, y, 'CV', 10,'Options', opt, 'Alpha',1);

lassoPlot(B, FitInfo, 'PlotType','Lambda','XScale','log');

%graph settings

xlabel('Lambda')

xtickangle(45)

ylabel('Theta')

[Lb, Lfitinfo] = lasso(ZX, y, 'Alpha', 1, 'Lambda', 0.015);

Lb