Robert Krall

SEIS 763

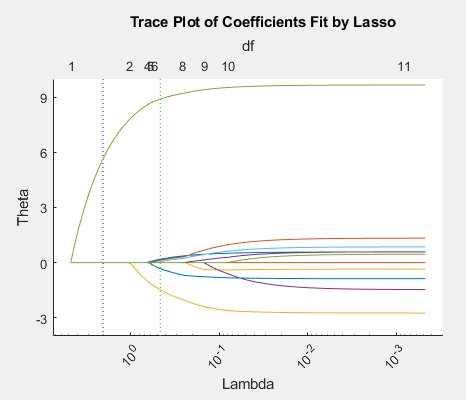
Assignment 3

Use **\*\*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.



Lambda value: 0.97

|  |  |
| --- | --- |
| Predictor | Theta (Θ) |
| Smoker | 7.9109 |
| Health Status Fair | -0.0998 |

Program:

%{

Robert Krall

6/13/2018

SEIS 763 - Assignment 3

%}

%load data

load patients

opt = statset('UseParallel',true);

%% ================ Part 1: Dummy Variables ================%

gender\_group = nominal(Gender);

dummy\_gender = dummyvar(gender\_group);

location\_group = nominal(Location);

dummy\_location = dummyvar(location\_group);

health\_group = nominal(SelfAssessedHealthStatus);

dummy\_health = dummyvar(health\_group);

%C = unique(SelfAssessedHealthStatus);

%% ================ Part 2: Normalize Data ================%

%only going to normalize none cat fields

X = [Age, Height, Weight];

y = Systolic;

%normalize/standardize (use built zscore funcation or my own)

[X,mu,sigma] = zscore(X);

%Adding dummy observations (or columns)

X = [X,dummy\_gender(:,end), Smoker,dummy\_location, dummy\_health];

%% ================ Part 3: Lasso Model ================%

[B,FitInfo] = lasso(X,y,'CV',10,'Options',opt,'Alpha',1,'PredictorNames',{'Age','Height','Weight','Gender\_Male','Smoker','Location\_CountyGeneralHositpal','Location\_StMarysMedicalCenter','Location\_VaHospital','HealthStatus\_Excellent','HealthStatus\_Fair','HealthStatus\_Good','HealthStatus\_Poor'});

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

%graph settings

xlabel('Lambda')

xtickangle(45)

ylabel('Theta')

yticks(-6:3:10)

[Lb, Lfitinfo] = lasso(X, y, 'Alpha', 1, 'Lambda', 0.97);

Lb

Lfitinfo