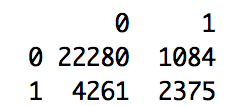
**Problem 2**

After reading all data into a data frame, we first use a function called **factor** to transform some columns of the data frame, such as SEX and EDUCATION, from numerical data into categorical data. Then we built a logistic model using **glm** function in R. As for the family type parameter in the glm function, we used "binomial" here. And we used a threshold of 0.5 here because that seemed to give us a better result after several trials. And using a threshold of 0.5 here means that any predicted value greater than 0.5 will be classified as 1, otherwise it will be classified as 0.

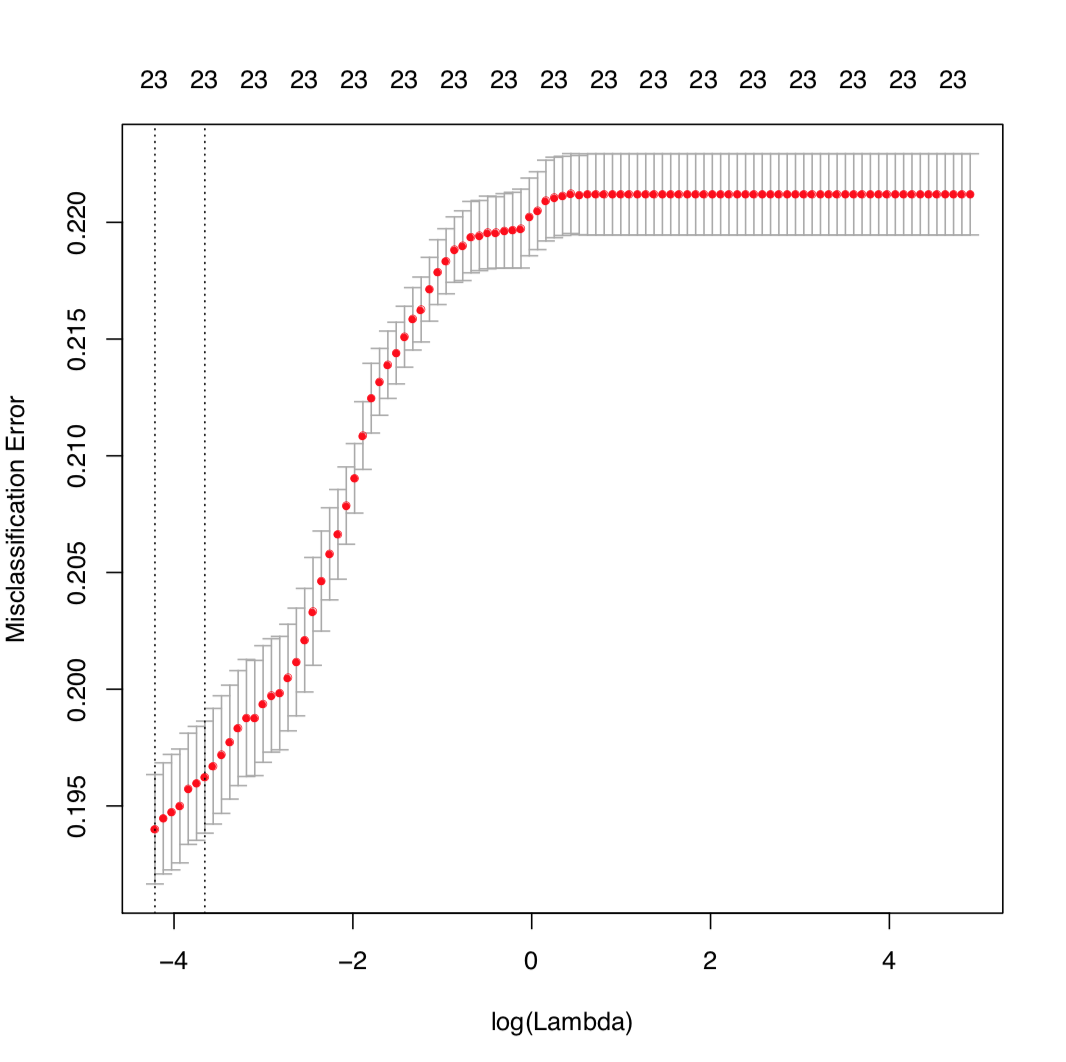
The accuracy of the logistic model we built was **82.18333%**. And the confusion matrix looks like the following:



We then tried the various regularization schemes discussed in lecture. Here we tried ridge, lasso and elastic net regularization method by using different values of alpha. And we used the **cv.glmnet** function in R to build models with different regularization schemes by passing different values into the alpha parameter. In addition, we assigned the nfolds parameter in the cv.glmnet function to be 10.

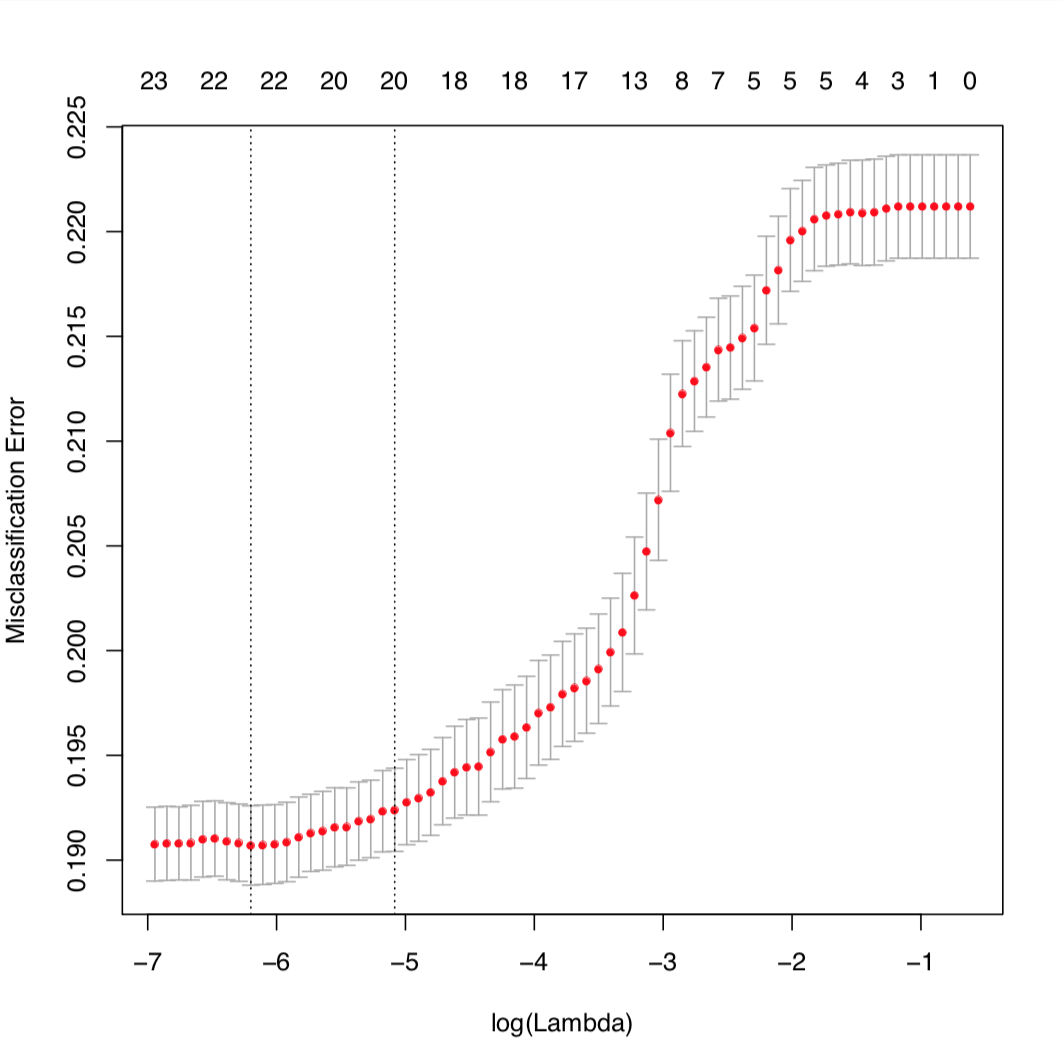
**Ridge Regression (alpha = 0)**

We got an accuracy of **80.6%** in the ridge regression model. And the Misclassification Error vs Log(Lambda) plot is showed in the following figure:



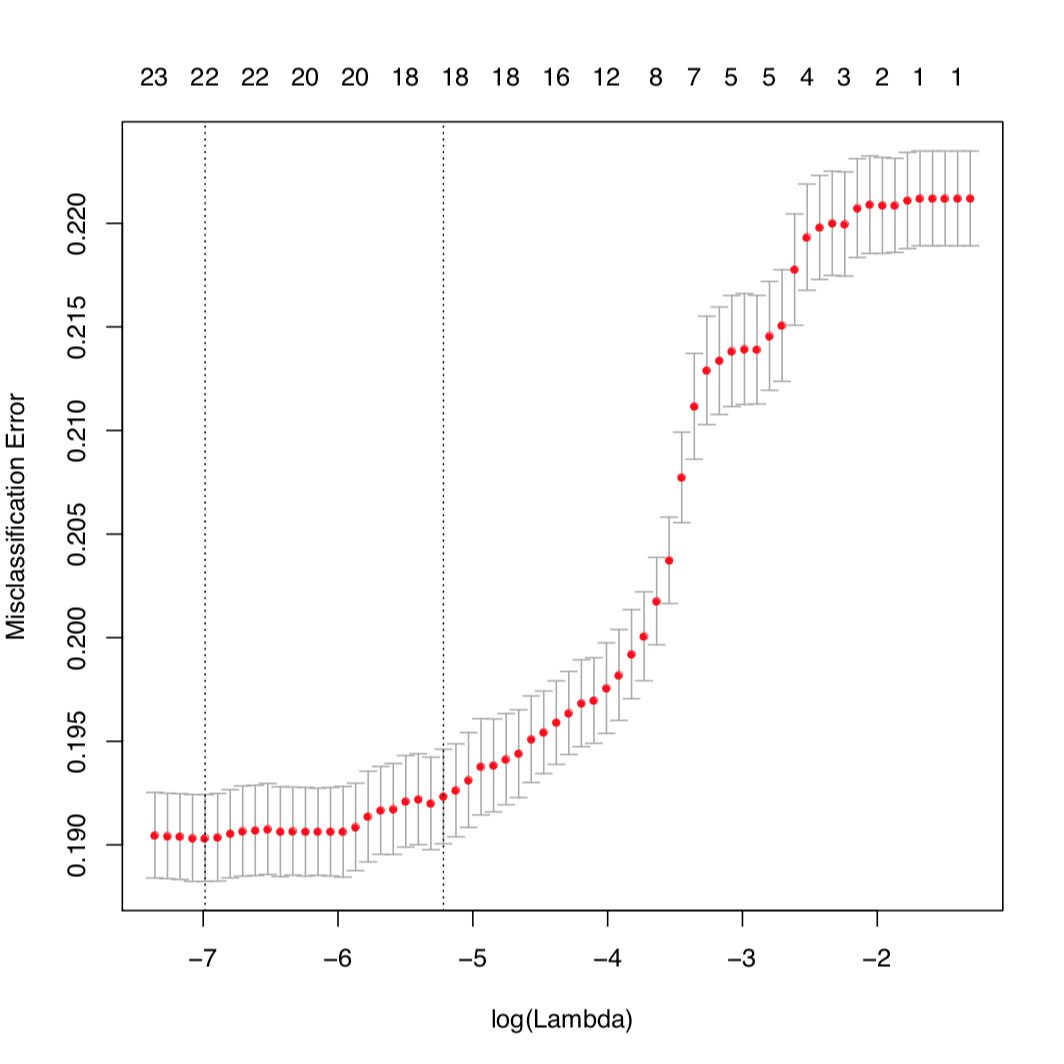
**Elastic Net Regression (alpha = 0.25)**

We got an accuracy of **80.93%** in the elastic net model, with alpha=0.25. And the Misclassification Error vs Log(Lambda) plot is showed in the following figure:

****

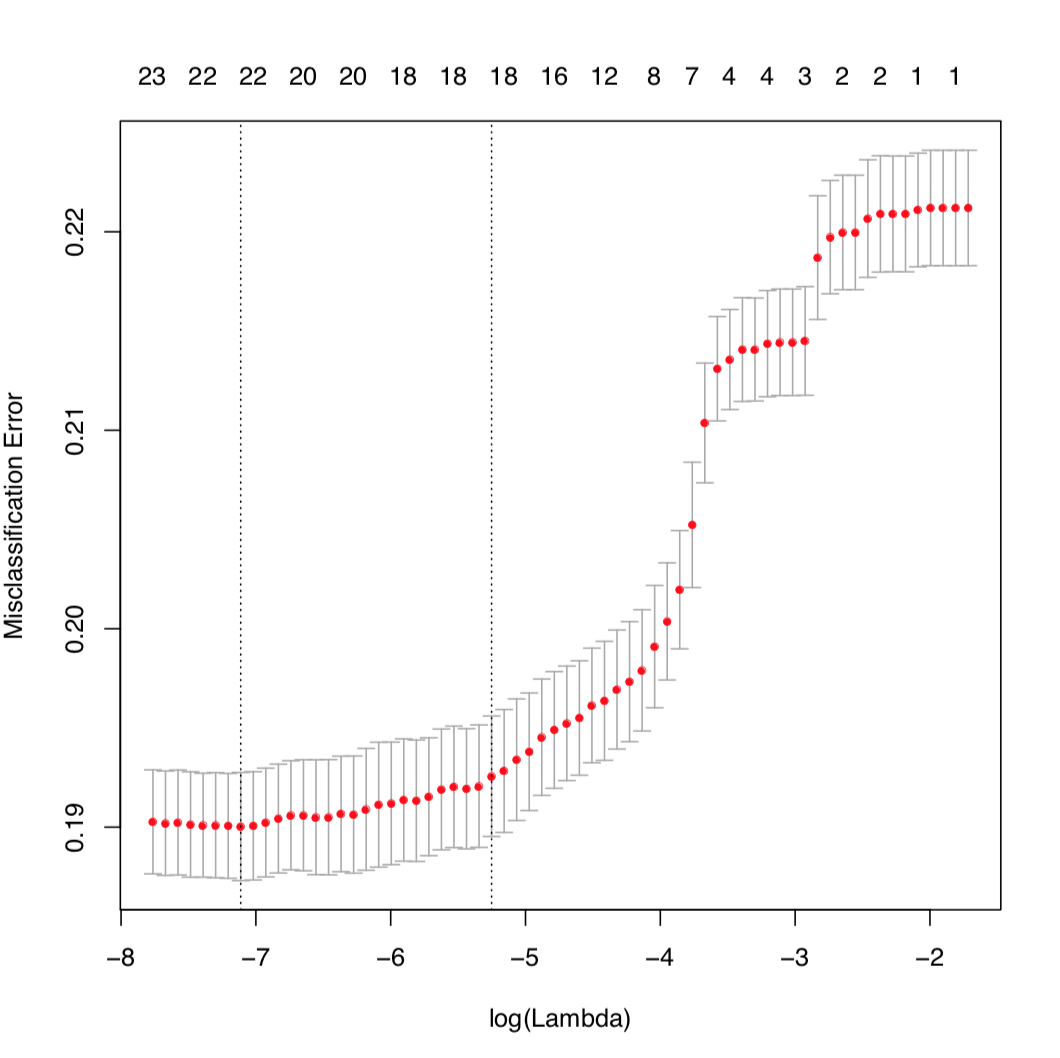
**Elastic Net Regression (alpha = 0.5)**

We got an accuracy of **80.96667%** in the elastic net model, with alpha=0.5. And the Misclassification Error vs Log(Lambda) plot is showed in the following figure:



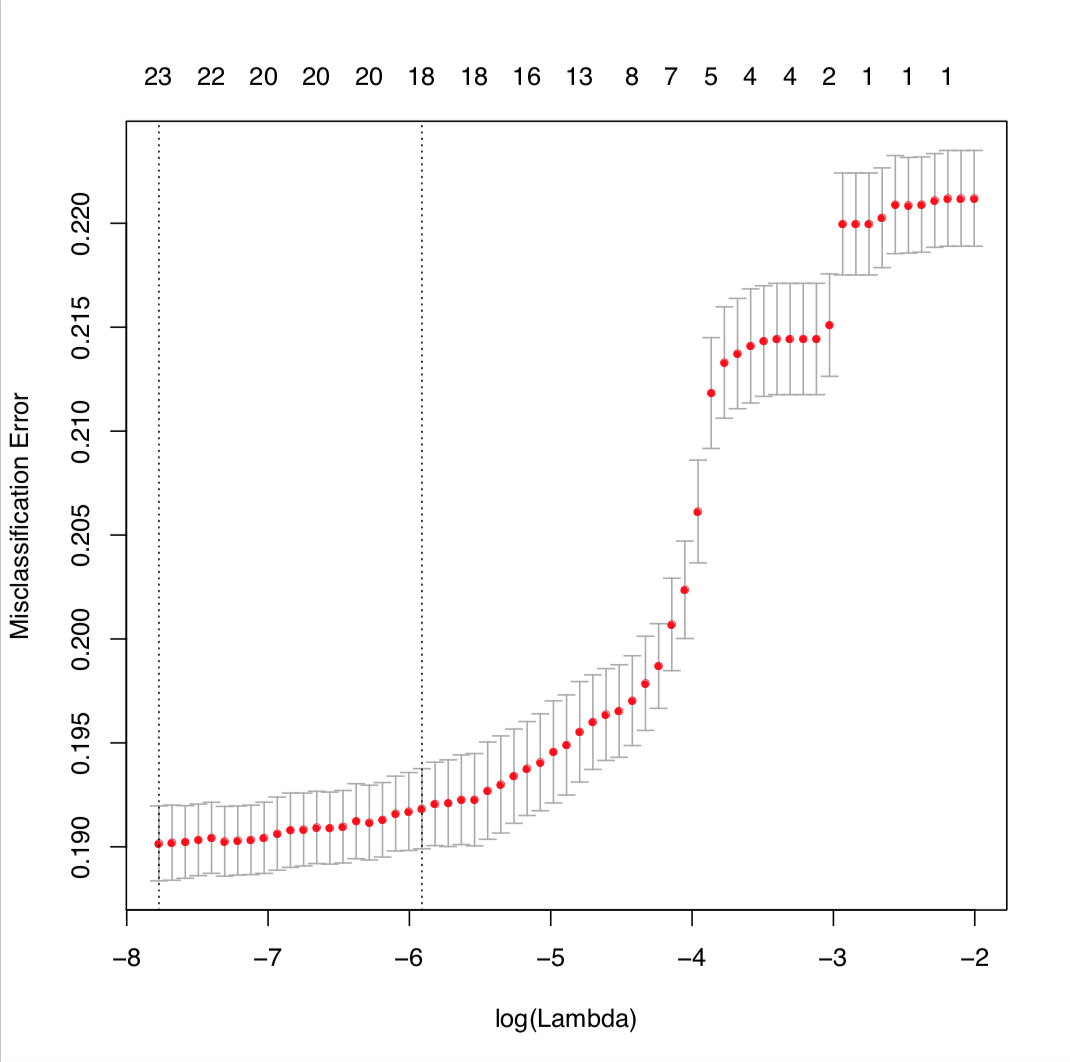
**Elastic Net Regression (alpha = 0.75)**

We got an accuracy of **80.99667%** in the elastic net model, with alpha=0.5. And the Misclassification Error vs Log(Lambda) plot is showed in the following figure:



**Lasso Regression** **(alpha = 1)**

We got an accuracy of **80.98333%** in the elastic net model, with alpha=0.5. And the Misclassification Error vs Log(Lambda) plot is showed in the following figure:



Overall, among all regression models with regularization schemas, elastic net regression and lasso regression seemed to have better accuracy than ridge regression.

However, we found out that logistic regression without any regularization seemed to have the best accuracy in this problem. And the differences of accuracy between the unregularized model and regularized models are significant compared to the differences of accuracy between all different regularized models.