CS 5821
February 16, 2017
Mariia Kravtsova
Assignment 4

3. a. K-fold cross validation is implemented by randomly dividing the set of observations into k groups of similar length. Each group is a validation set, and the remaining set is the training set. We estimate the test error by averaging the MSE estimates.
b. i. The validation set approach is faster than LOOCV computationally, however the disadvantages are

- Highly variable validation estimate of the test error rate
- Validation set error rate overestimates the test error rate because only a subset is used to fit the model

Ii. The main advantage that Leave One Out Cross Validation has is having less bias than the validation set method, but it might also have higher variance which is a disadvantage. Also, compared to the k-fold cross validation it is computationally expensive method.

6. a. The estimated standard errors for the coefficients associated with income is ~4.985, and balance is ~2.274, intercept is ~4.348.

b. Function, requires library(boot):

```
boot.fn=function(data,index)
  return (coef(glm(default ~ income + balance, data=data, subset=index, family="binomial")))
```

c. Standard errors of the logistic regression coefficients are the following:
for income ~ 4.5825, for balance ~ 2.2679, and intercept ~ 4.2393.

d. The glm summary estimates and bootstrap estimates are fairly close to each other, which means we have two good methods for resampling for this data set.