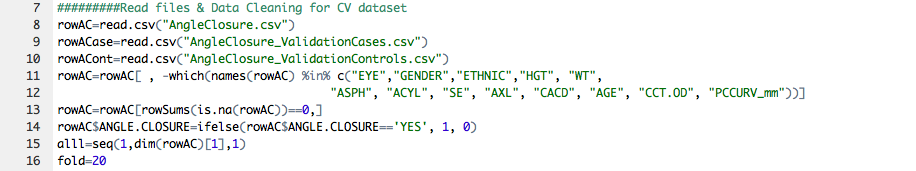
**ISYE 6740 TAKE HOME EXAM 3 (CODE & REPORT)**

**CODE**

1. Packages used for 5 models and AUC calculation in cross validation (neural network, SVM, random forest, KNN, logistic regression)



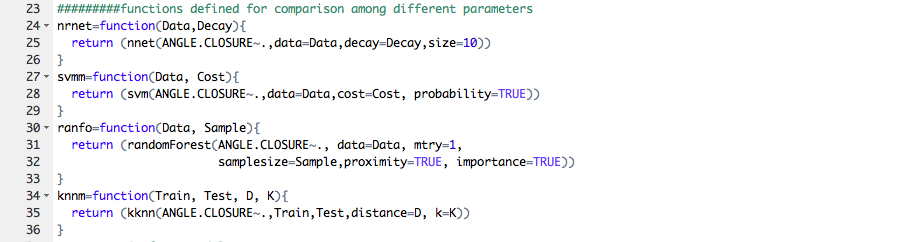
1. Read files & data manipulation for the cross-validation data set



1. Tuning parameters for each of the 5 models



1. Functions defined for each of the 5 models: take tuning parameters as inputs



1. Function for cross validation & output of cross validation



1. Find the best tuning parameter for each of the 5 models & find the solution for stacked models



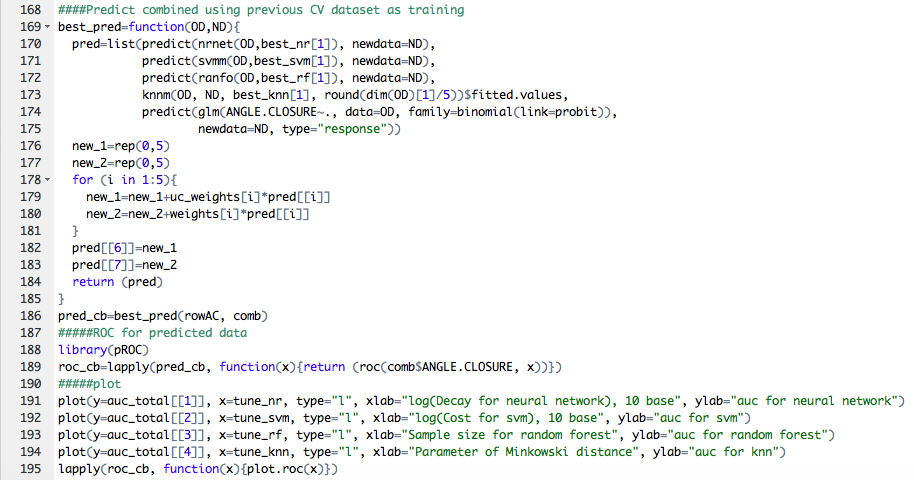
1. Functions for manipulating the two testing data sets: match column names and replace missing values



1. Data manipulation and combination of the two testing data sets



1. Predict and plot



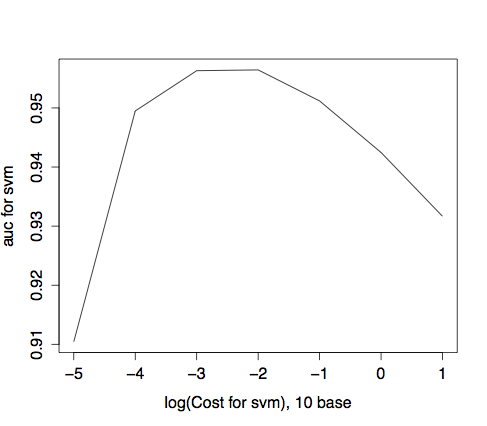
**REPORT**

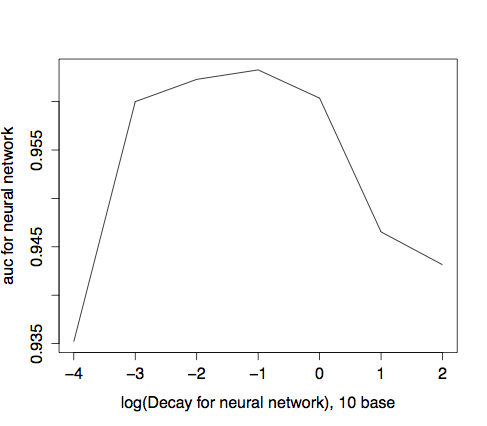
**AUC: Cross Validation Models**

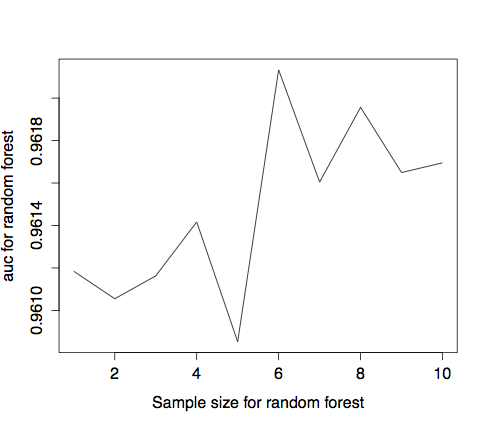
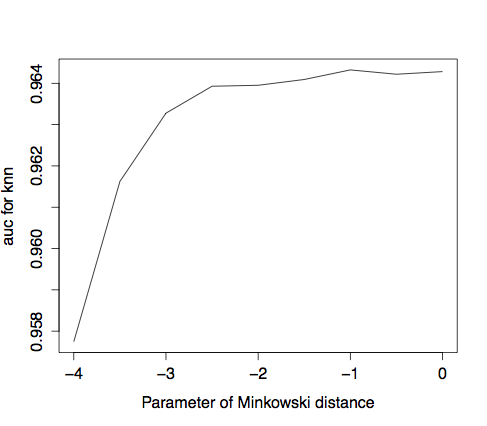
(Since logistic model has no tuning parameter, no AUC curve for logistic model)

Left 1: Neural Network

Right 1: SVM

Left 2: Random Forest

Right 2: KNN



|  | Neural Network | SVM | Random Forest | KNN | Logistic |
| --- | --- | --- | --- | --- | --- |
| Parameter Name | Decay | Cost | Sample Size | Minkowski Distance Parameter | NA |
| Parameter Range | 10^seq(-4, 2, 1) | 10^seq(-5, 1, 1) | seq(1, 10, 1) | 2^seq(-4, 0, 0.5) | NA |
| Optimal Parameter | 0.1 | 0.01 | 6 | 2^0.5 | NA |
| Max AUC | 0.9633 | 0.9564 | 0.9621 | 0.9643 | 0.9617 |

**ROC: 5 Models and 2 Stacked Models**

| Model | | AUC | |
| --- | --- | --- | --- |
| Neural Network | | 0.9687 | |
| SVM | | 0.9186 | |
| Random Forest | | 0.956 | |
| KNN | | 0.96 | |
| Logistic | | 0.9509 | |
| Unconstrained Stacked | | 0.958 | |
| Constrained Stacked | | 0.9615 | |
| Model | Unconstrained Weights | | Constrained Weights |
| Neural Network | 1.0330 | | 0.2778 |
| SVM | 4.4781 | | 0 |
| Random Forest | 1.0558 | | 0.2677 |
| KNN | 1.2125 | | 0.1662 |
| Logistic | 1.0136 | | 0.2883 |

**ROC: 5 Models and 2 Stacked Models**

Left 1: Neural Network

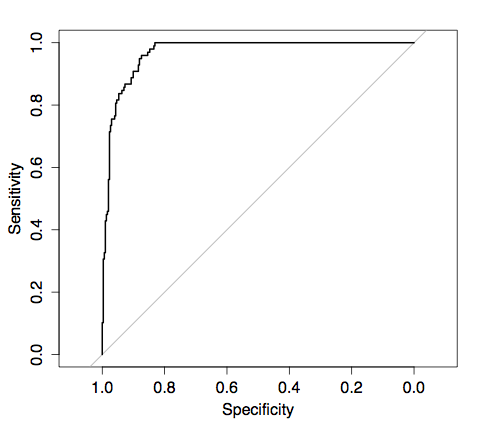
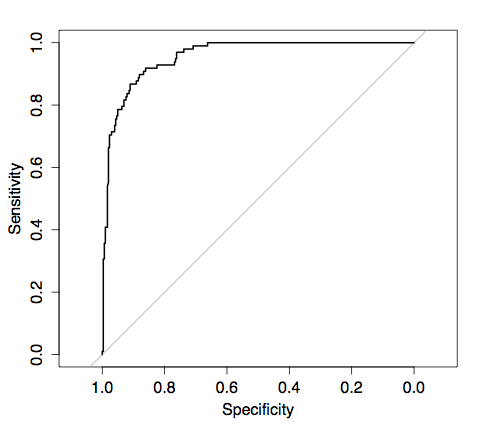
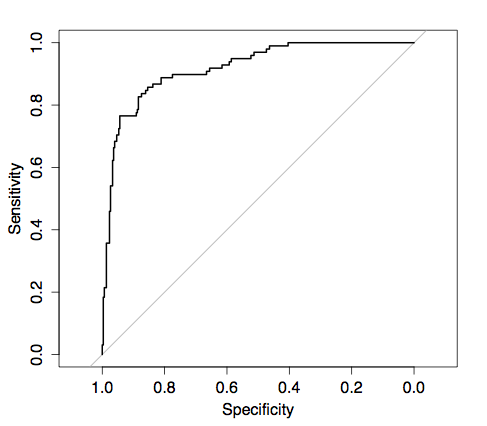
Middle 1: SVM

Right 1: Random Forest

Left 2: KNN

Middle 2: Logistic

Left 3: Unconstrained Stacked Model

Middle 3: Constrained Stacked Model

