MiniProj_4

Wenxiong Lu

1a

```
## Warning in nominalTrainWorkflow(x = x, y = y, wts = weights, info = trainInfo, :
## There were missing values in resampled performance measures.
```

```
## [1] 1.131119
```

b

```
## Subset selection object
## Call: regsubsets.formula(Quality ~ ., wine, method = c("exhaustive",
##     "backward", "forward", "seqrep"))
## 6 Variables  (and intercept)
##          Forced in Forced out
## Clarity      FALSE      FALSE
## Aroma        FALSE      FALSE
## Body         FALSE      FALSE
## Flavor       FALSE      FALSE
## Oakiness     FALSE      FALSE
## Region       FALSE      FALSE
## 1 subsets of each size up to 6
## Selection Algorithm: exhaustive
##          Clarity Aroma Body Flavor Oakiness Region
## 1  ( 1 ) " "     " "   " "  "*"    " "      " "
## 2  ( 1 ) " "     " "   " "  "*"    "*"      " "
## 3  ( 1 ) " "     "*"   " "  "*"    "*"      " "
## 4  ( 1 ) "*"     "*"   " "  "*"    "*"      " "
## 5  ( 1 ) "*"     "*"   "*"  "*"    "*"      " "
## 6  ( 1 ) "*"     "*"   "*"  "*"    "*"      "*"
```

```
## [1] 0.6137349 0.6417466 0.6776290 0.6801276 0.6769428 0.6666623
```

```
## Warning in nominalTrainWorkflow(x = x, y = y, wts = weights, info = trainInfo, :
## There were missing values in resampled performance measures.
```

```
## [1] 0.9922728
```

By checking z test, we can remove SkinThickness which is insiginificant and build the model using all other predictors.

c

```
## [1] 0.6137349 0.6417466 0.6776290 0.6801276 0.6769428 0.6666623
```

```
## Warning in nominalTrainWorkflow(x = x, y = y, wts = weights, info = trainInfo, :
## There were missing values in resampled performance measures.
```

```
## [1] 0.9922728
```

d

```
## [1] 0.6137349 0.6417466 0.6776290 0.6801276 0.6769428 0.6666623
```

```
## Warning in nominalTrainWorkflow(x = x, y = y, wts = weights, info = trainInfo, :
## There were missing values in resampled performance measures.
```

```
## [1] 0.9922728
```

e

```
## Warning: Option grouped=FALSE enforced in cv.glmnet, since < 3 observations per
## fold
```

```
## Warning in nominalTrainWorkflow(x = x, y = y, wts = weights, info = trainInfo, :
## There were missing values in resampled performance measures.
```

```
## [1] 0.9551514
```

f

```
## Warning: Option grouped=FALSE enforced in cv.glmnet, since < 3 observations per
## fold
```

```
## Warning in nominalTrainWorkflow(x = x, y = y, wts = weights, info = trainInfo, :
## There were missing values in resampled performance measures.
```

```
## [1] 0.9810641
```

g

```
##          a         b         c         d         e         f
## 1 1.131119 0.9922728 0.9922728 0.9922728 0.9551514 0.9810641
```

based on the table we should choose adjusted R2 in any king model selection method for smallest test error.

#####2 a

```
## [1] 0.2195
```

b

```
## Morgan-Tatar search since family is non-gaussian.
```

```
##
## Call:  glm(formula = y ~ ., family = family, data = Xi, weights = weights)
##
## Coefficients:
##                 (Intercept)              Pregnancies..
##                   -8.027315                    0.126371
##                   Glucose..             BloodPressure..
##                    0.033681                   -0.009581
##                   Insulin..                       BMI..
##                   -0.001212                    0.077874
## DiabetesPedigreeFunction..                       Age..
##                    0.889495                    0.012894
##
## Degrees of Freedom: 1999 Total (i.e. Null);  1992 Residual
## Null Deviance:       2569
## Residual Deviance: 1914  AIC: 1930
```

```
## [1] 0.2199697
```

c

```
## Morgan-Tatar search since family is non-gaussian.
```

```
##
## Call:  glm(formula = y ~ ., family = family, data = Xi, weights = weights)
##
## Coefficients:
##                 (Intercept)               Pregnancies..
##                   -8.027315                    0.126371
##                   Glucose..             BloodPressure..
##                    0.033681                   -0.009581
##                   Insulin..                       BMI..
##                   -0.001212                    0.077874
## DiabetesPedigreeFunction..                       Age..
##                    0.889495                    0.012894
##
## Degrees of Freedom: 1999 Total (i.e. Null);  1992 Residual
## Null Deviance:       2569
## Residual Deviance: 1914  AIC: 1930
```

```
## [1] 0.2214898
```

d

```
## Morgan-Tatar search since family is non-gaussian.
```

```
##
## Call:  glm(formula = y ~ ., family = family, data = Xi, weights = weights)
##
## Coefficients:
##                 (Intercept)               Pregnancies..
##                   -8.027315                    0.126371
##                   Glucose..             BloodPressure..
##                    0.033681                   -0.009581
##                   Insulin..                       BMI..
##                   -0.001212                    0.077874
## DiabetesPedigreeFunction..                       Age..
##                    0.889495                    0.012894
##
## Degrees of Freedom: 1999 Total (i.e. Null);  1992 Residual
## Null Deviance:       2569
## Residual Deviance: 1914  AIC: 1930
```

```
## [1] 0.2205147
```

e

```
## [1] 0.23
```

f

```
## [1] 0.221
```

g

```
t_all=as.data.frame(matrix(dat=c(t21,t22,t23,t24,t25,t26),nrow=1,byrow = TRUE))
names(t_all)<-c('a','b','c','d','e','f')
t_all
```

The smallest test error is of model in question 2a (Full-model). It the same as the last proj3