Clara

Anqi Chi SID:460204008

22/10/2018

Data Clean

LDA

Table 1: Coefficients of Predictors in LDA

	Estimated Coefficients
BMR	0.001249
Energy Intake	-0.434204
Total mins spent sedentary	-0.000046
Sex	2.093857
Whether currently on a diet	0.363241
Carbohydrate diet	-0.005032
Fat diet	0.001558
Protein diet	0.002124

Comment:

The above output suggests the following interpretations for each of the variables. * People who have higher BMR are more likely to be obese. * Lower EIBMR1(Energy intake) increases obesity probability (since the coefficient -0.434 is negative). * Lower ADTOTSE(Total mins spent sitting or lying down) increases obesity probability. * Sex = 1 for males. So females are more likely to obese. * People on a diet are more likely to be obese compared with people not on a diet * People who have high fat or high protein diet type are more likely to be obese * People who have high carbon diet type decrease the probability of obesity

CV error

```
## [1] 0.1823225
```

Comment: The CV error for LDA is 18.232 percent
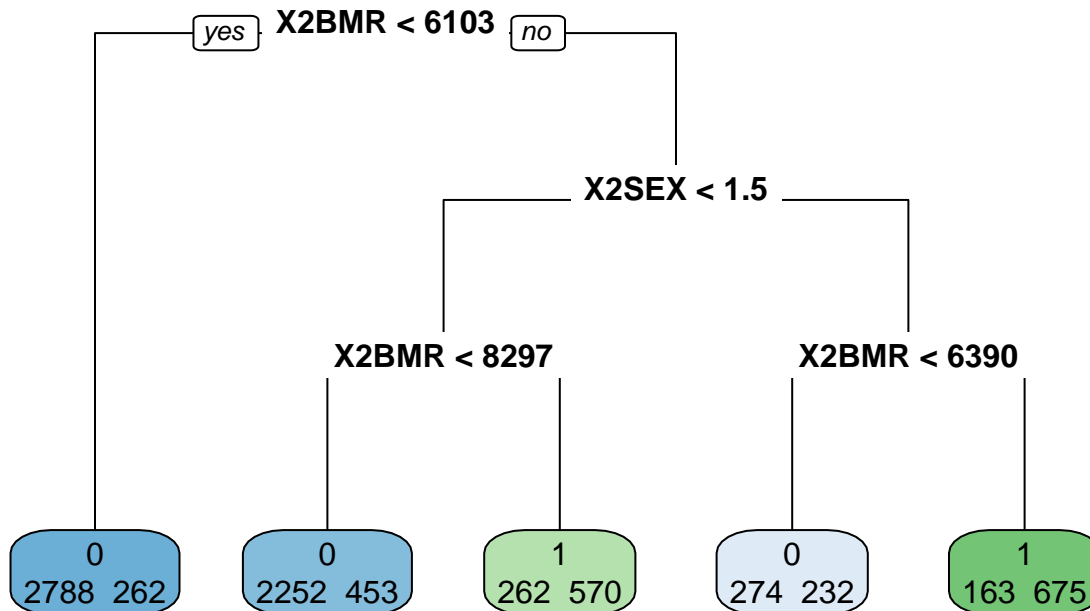
CART Fit for Obesity

X2BMR < 6103 *yes* *no*

X2SEX < 1.5

X2BMR < 8297   **X2BMR < 6390**

0	0	1	0	1
2788 262	2252 453	262 570	274 232	163 675

Table 2: Obesity Rate for Male with BMR greater than 6103

	$6103 < BMR < 8297$	$BMR > 8297$
estimated obesity rate	6.7%	68.5%

Table 3: Obesity Rate for Female with BMR greater than 6103

	$6103 < BMR < 6390$	$BMR > 6390$
estimated obesity rate	45.8%	80.1%

Comment:

- 91.4 percent of people(including male and female) whose BMR are less than 6103 are normal (not obese).
- 83.3 percent of male whose BMR are between 6103 and 8297 are normal.
- 68.5 percent of male whose BMR are greater than 8297 are obese.
- 54.2 percent of female whose BMR are between 6103 and 6390 are normal.
- 80.1 percent of female whose BMR are greater than 6390 are obese.

CV error

```
## [1] 0.1755138
```

Comment: The CV error for CART is 17.551 percent

Logistic regression

```
##
## Call:
## glm(formula = y ~ ., family = binomial, data = X1)
##
## Deviance Residuals:
##     Min      1Q    Median      3Q      Max
## -2.7799  -0.6546  -0.3851   0.4044   3.2858
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.777e+01  6.188e-01 -28.717  < 2e-16 ***
## BMR          1.879e-03  5.203e-05  36.109  < 2e-16 ***
## EIBMR1      -7.113e-01  7.220e-02  -9.852  < 2e-16 ***
## ADTOTSE     -5.952e-05  2.449e-05  -2.430   0.0151 *
## SEX          3.423e+00  1.174e-01  29.159  < 2e-16 ***
## BDYMSQ04     3.802e-01  8.342e-02   4.557 5.18e-06 ***
## CHOPER1     -1.014e-02  3.948e-03  -2.567   0.0103 *
## FATPER1      1.507e-03  4.357e-03   0.346   0.7295
## PROPER1     -9.982e-04  5.854e-03  -0.171   0.8646
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 9350.7  on 7930  degrees of freedom
## Residual deviance: 6587.0  on 7922  degrees of freedom
## AIC: 6605
##
## Number of Fisher Scoring iterations: 5
```

Comment:

The full model has the coefficeints for BMR, EIBMR1, ADTOTSE, SEX, BDYMSQ04, CHOPER1 as statisticlly signiffcantly different from zero at the 0.05 level. The fitted model is

$$\text{logit}(p) = -0.1093 + 0.001879 \cdot \text{BMR} - 3.423 \cdot \text{SEX} - 0.7113 \cdot \text{EIBMR1} - 0.00005952 \cdot \text{ADTOTSE} + 0.3802 \cdot \text{BDYMSQ04} - 0.01014 \cdot \text{CH}$$

where p is the probability of obesity.

CV error for glms.

```
## [1] 0.1777834
```

Comment: The CV error for glm is 17.78 percent

summary

```
##
## Attaching package: 'huxtable'

## The following object is masked from 'package:kableExtra':
##
##     add_footnote
```

Table 4: Summary of the CV Errors

Methods	Errors
LDA	18.232
CART	17.551
Logistic Regression	17.778