Machine Learning

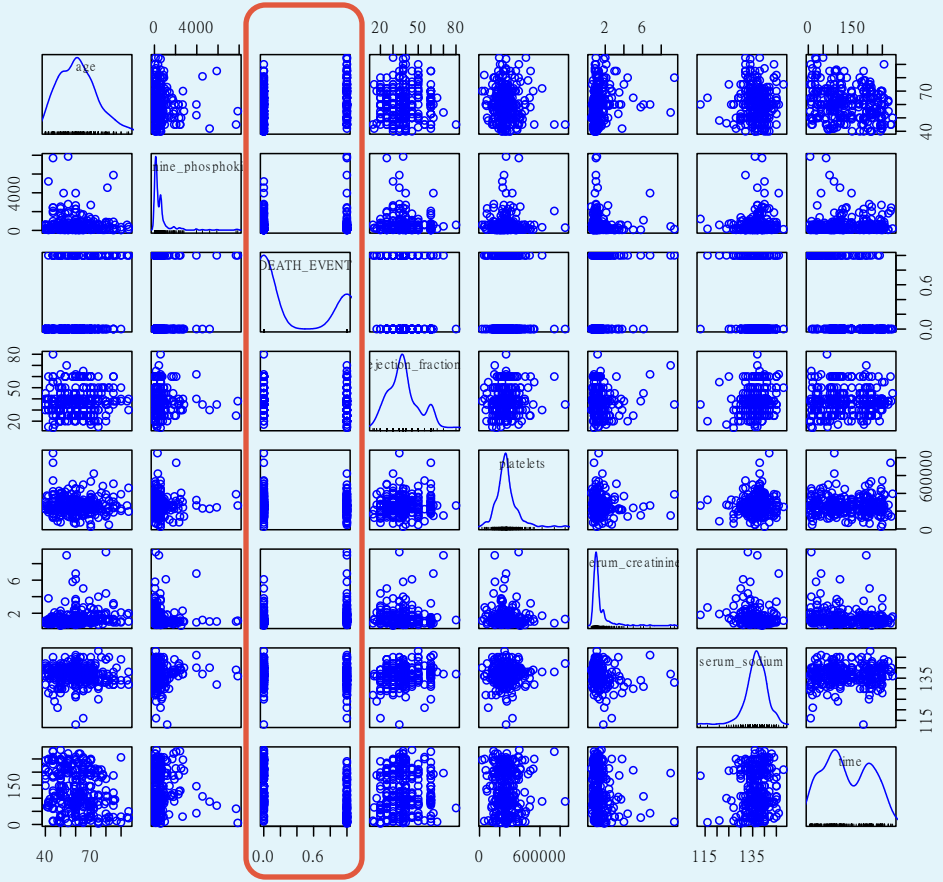
--HeartFailure Analysis

Group 6 :
ChihHao Luca Yuan
Ching Yu Hsu

Table of contents

1(a)-Overall view of relationship of

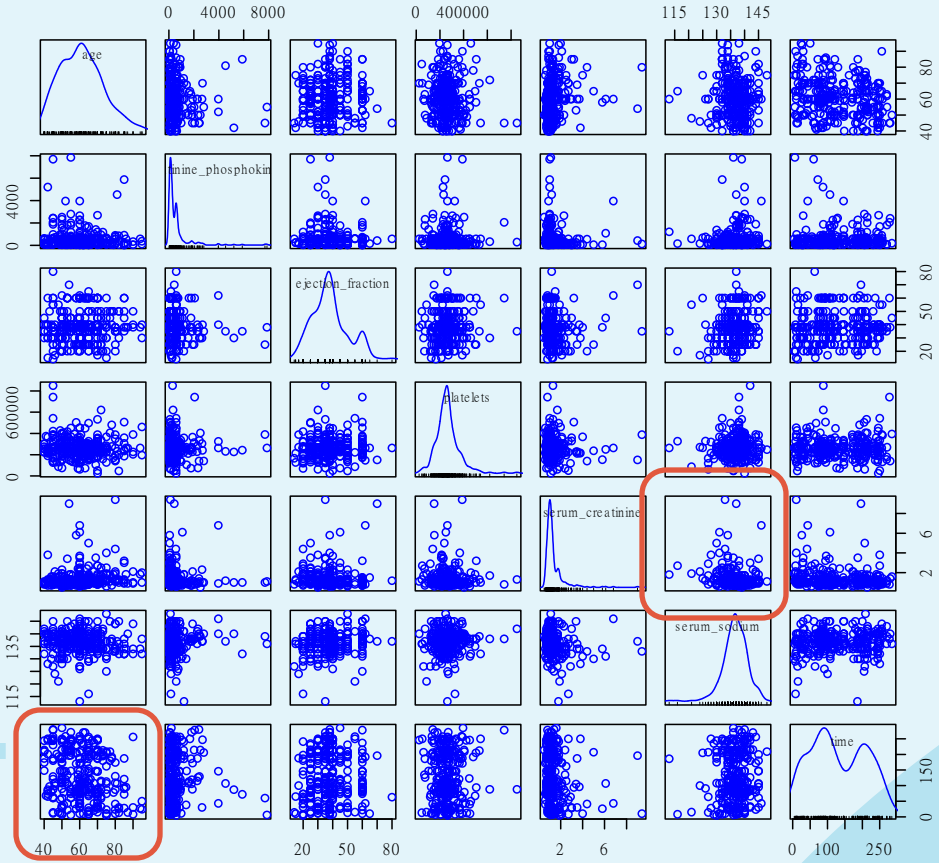
Death Event with all continuous variables



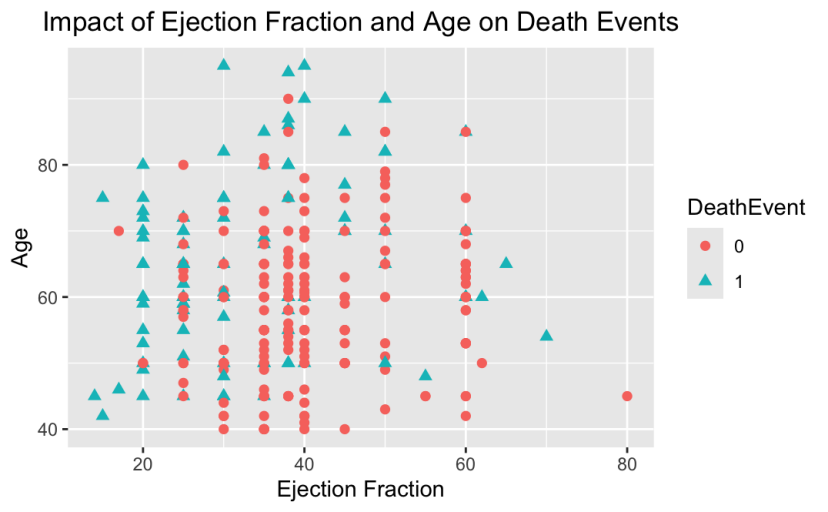
Correlation

	age	creatinine_phosphokinase	ejection_fraction	platelets	serum_creatinine	serum_sodium	time
age	1						
creatinine_phosphokinase	-0.0816	1					
ejection_fraction	0.0601	-0.044079554	1				
platelets	-0.0524	0.024463389	0.072177466	1			
serum_creatinine	0.1592	-0.01640848	-0.011302475	-0.0412	1		
serum_sodium	-0.046	0.059550156	0.175902282	0.0621	-0.18909521	1	
time	-0.2241	-0.009345653	0.041729235	0.0105	-0.149315418	0.08764	1

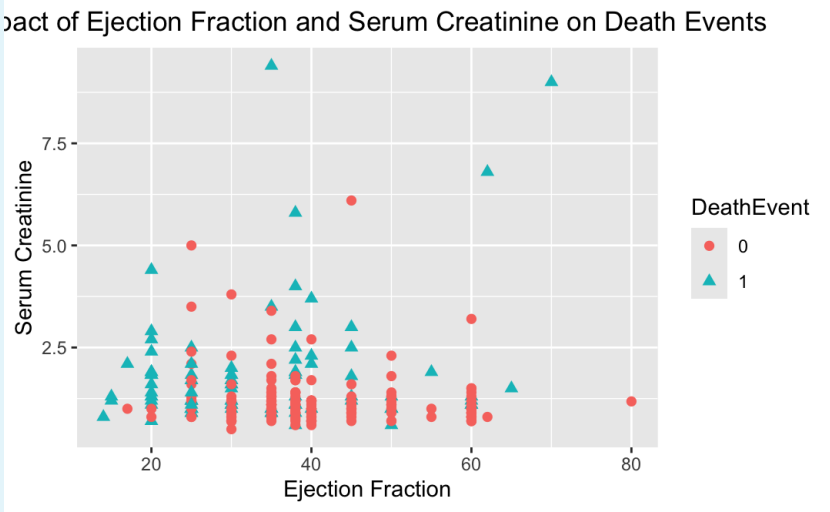
1(b)-Highlight at least two graphs where there are strong relationships between pairs of continuous variables



1(c)–The graph with pairs continuous variables and plot Death Event



Age & Ejection Fraction



Serum Creatinine & Ejection Fraction

2(a)-Perceptron with 2 variables: Model 1

```r
X <- Heart_train[, c("ejection_fraction", "age")] # Input Matrix
```

```r
Y <- Heart_train$DeathEvent
```

```r
set.seed(123)
perceptron(X, Y, 25)
```

```
Epoch #:  21
Number Incorrect:  73
Final Weight:  -182.2484 114.7661
Epoch #:  22
Number Incorrect:  68
Final Weight:  -161.2484 97.7661
Epoch #:  23
Number Incorrect:  69
Final Weight:  -160.2484 99.7661
Epoch #:  24
Number Incorrect:  70
Final Weight:  -159.2484 107.7661
Epoch #:  25
Number Incorrect:  69
Final Weight:  -161.2484 99.7661
```

	FALSE	TRUE
FALSE	57	33
TRUE	33	57

Accuracy : 0.633

2(a)-Perceptron with 2 variables: Model 2

```r
X2 <- Heart_train[, c("ejection_fraction", "serum_creatinine")]
```

```r
set.seed(123)

perceptron(X2, Y, 25)
```

```
Epoch #:  10
Number Incorrect:  76
Final Weight:  -14.24845 278.2261
Epoch #:  11
Number Incorrect:  76
Final Weight:  -18.24845 290.2861
Epoch #:  12
Number Incorrect:  75
Final Weight:  -19.24845 299.6161
Epoch #:  13
Number Incorrect:  76
Final Weight:  -0.2484496 310.2461
Epoch #:  14
Number Incorrect:  70
Final Weight:  -8.24845 315.7061
Epoch #:  15
Number Incorrect:  67
Final Weight:  -8.24845 316.1261
```

	FALSE	TRUE
FALSE	58	32
TRUE	32	58

Accuracy : 0.644

2(b)-Perceptron with 3 variables or more : Model 3

```{r}
X3 <- Heart_train[, c("ejection_fraction", "serum_creatinine", "age")] #
Input Matrix
```

```{r}
set.seed(123)
perceptron(X3, Y, 25)
```

```
Epoch #:  20
Number Incorrect:  69
Final Weight:  -176.2484 411.1361 45.17954
Epoch #:  21
Number Incorrect:  71
Final Weight:  -175.2484 433.0261 90.17954
Epoch #:  22
Number Incorrect:  71
Final Weight:  -198.2484 449.6461 56.17954
Epoch #:  23
Number Incorrect:  73
Final Weight:  -212.2484 470.7361 64.17954
Epoch #:  24
Number Incorrect:  69
Final Weight:  -184.2484 489.0261 105.1795
Epoch #:  25
Number Incorrect:  66
Final Weight:  -166.2484 503.5461 83.17954
```

	FALSE	TRUE
FALSE	60	30
TRUE	30	60

Accuracy : 0.6667

2(b)-Perceptron with 3 variables or more : Model 4

```{r}
X4 <- Heart_train[, c("age",
                      "ejection_fraction",
                      "serum_creatinine",
                      "serum_sodium")]
```

```{r}
set.seed(123)
perceptron(X4, Y, 25)
```

```
Epoch #:  15
Number Incorrect:  80
Final Weight:  634.7516 -791.2339 375.3895 -218.3397
Epoch #:  16
Number Incorrect:  79
Final Weight:  631.7516 -797.2339 403.1795 -94.33965
Epoch #:  17
Number Incorrect:  62
Final Weight:  523.7516 -730.2339 425.0195 -72.33965
Epoch #:  18
Number Incorrect:  64
Final Weight:  447.7516 -668.2339 444.4795 -50.33965
Epoch #:  19
Number Incorrect:  64
Final Weight:  371.7516 -611.2339 464.1395 -30.33965
Epoch #:  20
Number Incorrect:  64
Final Weight:  305.7516 -544.2339 484.1795 -9.339652
```

	FALSE	TRUE
FALSE	64	26
TRUE	26	64

Accuracy : 0.711

2(c)-Summarize the accuracy of all the perceptron models

	Variables	Accuracy
Model 1	• Ejection fraction • Age	0.633
Model 2	• Ejection fraction • Serum creatinine	0.644
Model 3	• Ejection fraction • Serum creatinine • Age	0.6667
Model 4	• Ejection fraction • Serum creatinine • Serum sodium • Age	0.711

3(a)- Develop the SVM with 2 variables : Model 1

```r
set.seed(123)
heart_svm1_model <-svm(DEATH_EVENT ~ age + smoking, data = heart_svm1_train)
```

y- variable : Death Event
x- variable : age, smoking

```
Call:
svm(formula = DEATH_EVENT ~ age + smoking, data = heart_svm1_train)


Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  radial
       cost:  1

Number of Support Vectors:  145

 ( 79 66 )


Number of Classes:  2

Levels:
 0 1
```

	FALSE	TRUE
FALSE	60	1
TRUE	27	2

Accuracy : 0.6889

3(a)- Develop the SVM with 2 variables : Model 2

```r
heart_svm2_model <-svm(DEATH_EVENT ~ diabetes + creatinine_phosphokinase,
data = heart_svm2_train)
```

y- variable : Death Event
x- variable : diabetes, creatinine phosphokinase

```
Call:
svm(formula = DEATH_EVENT ~ diabetes + creatinine_phosphokinase,
    data = heart_svm2_train)


Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  radial
       cost:  1

Number of Support Vectors:  151

 ( 84 67 )


Number of Classes:  2

Levels:
 0 1
```

	FALSE
FALSE	61
TRUE	29

Accuracy : 0.6778

3(b)- Develop the SVM with 3 variables or more : Model 3

```r
heart_svm3_model <-svm(DEATH_EVENT ~., data = heart_svm3_train)
```

y- variable : Death Event
x- variable : age, smoking, diabetes, creatinine phosphokinase

```
Call:
svm(formula = DEATH_EVENT ~ ., data = heart_svm3_train)


Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  radial
       cost:  1

Number of Support Vectors:  147

 ( 80 67 )


Number of Classes:  2

Levels:
 0 1
```

	FALSE	TRUE
FALSE	60	1
TRUE	26	3

Accuracy : 0.70

3(b)- Develop the SVM with 3 variables or more : Model 4

```r
heart_svm4_model <-svm(DEATH_EVENT ~., data = heart_svm4_train)
```

y- variable : Death Event
x- variable : age, smoking, diabetes, sex, high blood pressure, ejection fraction

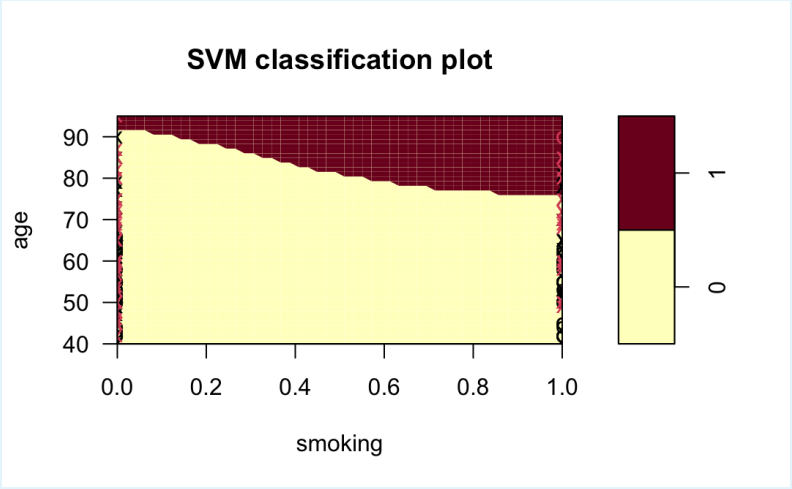
```
Call:
svm(formula = DEATH_EVENT ~ ., data = heart_svm4_train)


Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  radial
       cost:  1

Number of Support Vectors:  137

 ( 75 62 )


Number of Classes:  2

Levels:
 0 1
```

	FALSE	TRUE
FALSE	50	11
TRUE	17	12

Accuracy : 0.6889

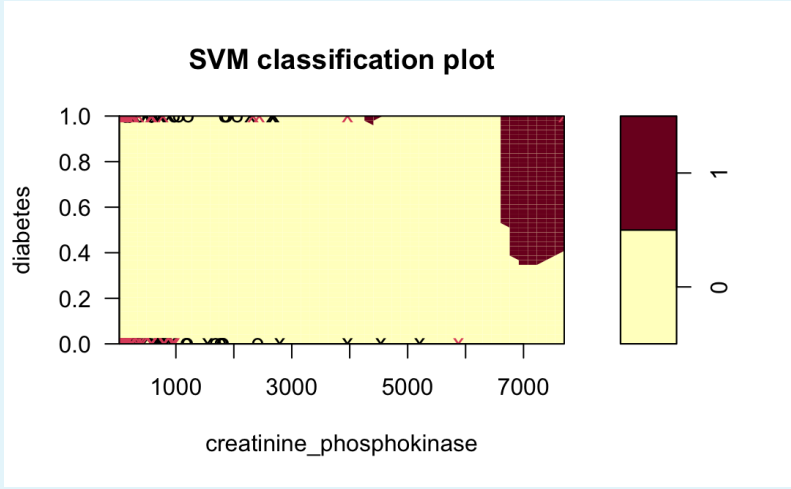
3(c)-SVM Results

```r
plot(heart_svm1_model, heart_svm1_train, age ~ smoking)
```



age & smoking

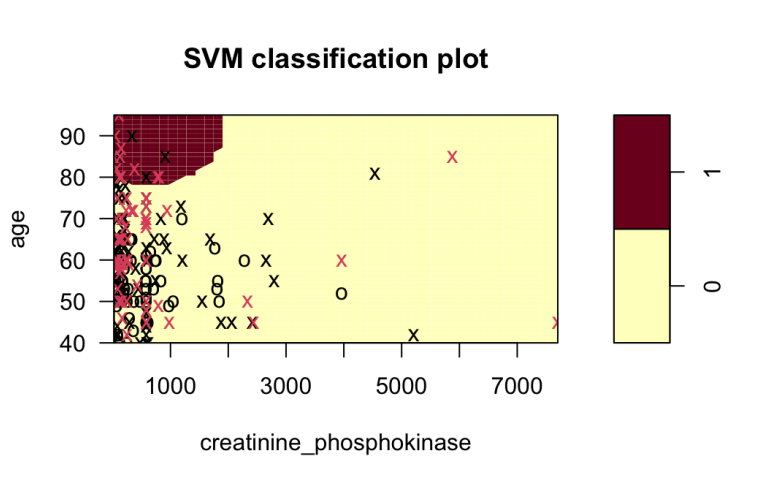
```r
plot(heart_svm2_model, heart_svm2_train, diabetes ~ creatinine_phosphokinase)
```



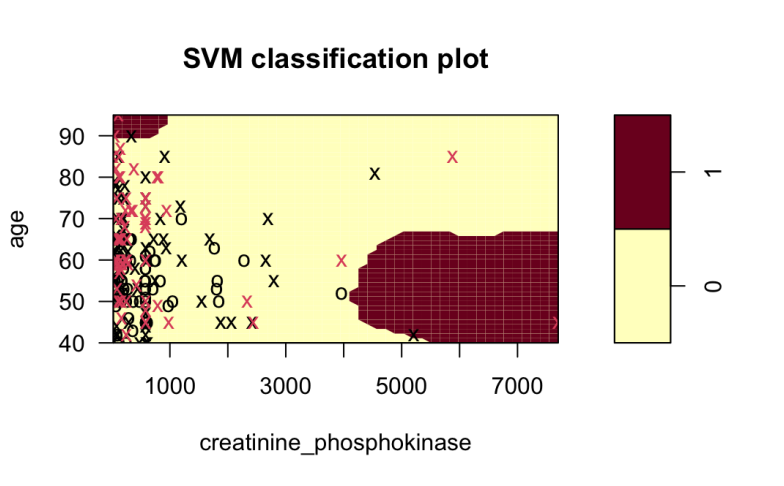
Diabetes & creatinine phosphokinase

3(c)-SVM Results

```
plot(heart_svm3_model, heart_svm3_train, age ~ creatinine_phosphokinase,
slice = list(smoking = 1, diabetes = 1))
```




```
plot(heart_svm3_model, heart_svm3_train, age ~ creatinine_phosphokinase,
slice = list(smoking = 0, diabetes = 1))
```

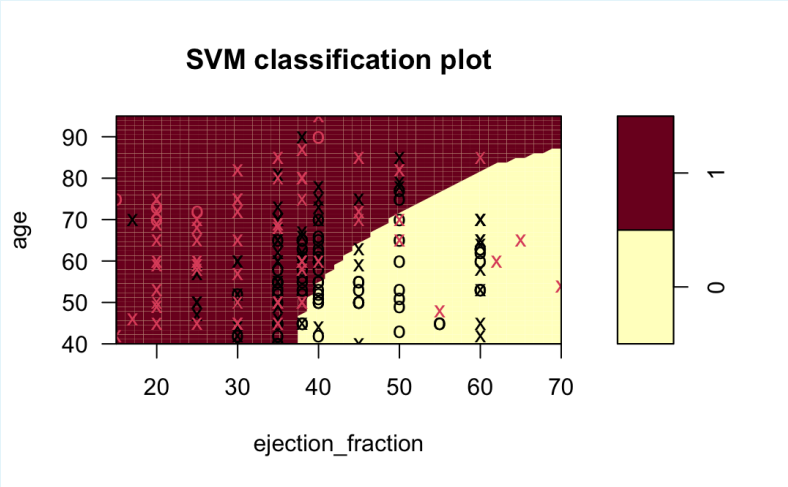


Compare "people who don't smoke (smoking = 0) and have diabetes (diabetes = 1)" with "people who smoke and have diabetes."

3(c)-SVM Results

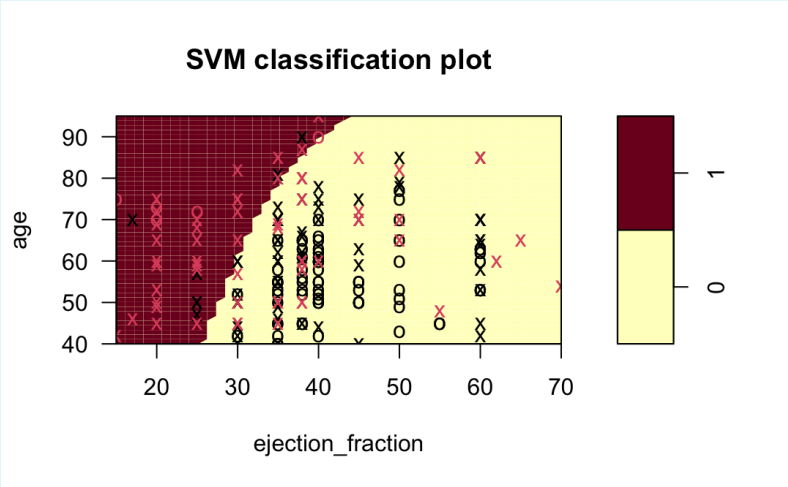
Female smoker with diabetes and high blood pressure

```r
plot(heart_svm4_model, heart_svm4_train, age ~ ejection_fraction,
     slice = list(smoking = 1, diabetes = 1, sex = 0, high_blood_pressure = 1))
```



Female smoker with no diabetes and no high blood pressure

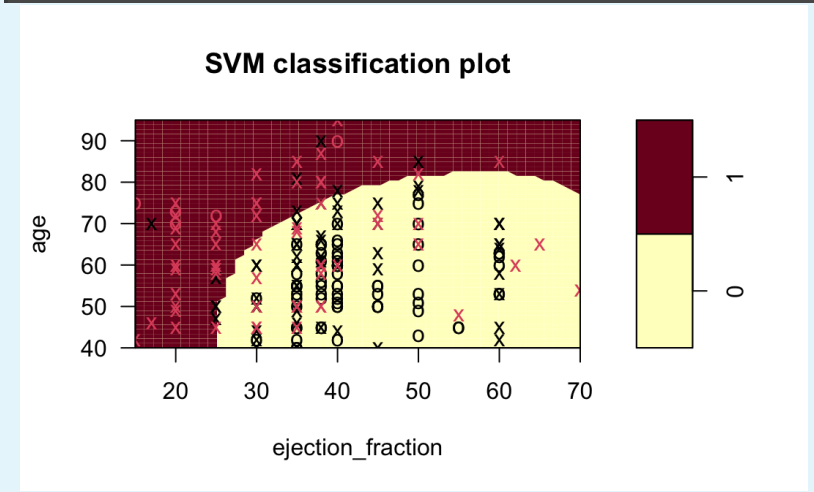
```r
plot(heart_svm4_model, heart_svm4_train, age ~ ejection_fraction,
     slice = list(smoking = 1, diabetes = 0, sex = 0, high_blood_pressure = 0))
```



3(c)-SVM Results

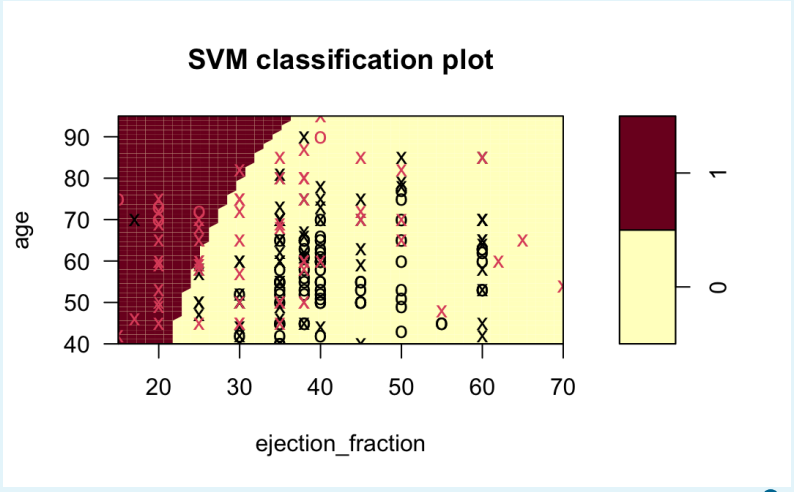
Female, not a smoker, with diabetes and high blood pressure

```r
plot(heart_svm4_model, heart_svm4_train, age ~ ejection_fraction,
     slice = list(smoking = 0, diabetes = 1, sex = 0, high_blood_pressure = 1))
```



Female , not a smoker with no diabetes and no high blood pressure

```r
plot(heart_svm4_model, heart_svm4_train, age ~ ejection_fraction,
     slice = list(smoking = 0, diabetes = 0, sex = 0, high_blood_pressure = 0))
```



3(d)-Summarize the accuracy of all the SVM models

	Variables	Accuracy
Model 1	• age • smoking	0.6889
Model 2	• diabetes • creatinine phosphokinase	0.6778
Model 3	• age • smoking • diabetes • creatinine phosphokinase	0.70
Model 4	• age • smoking • diabetes • sex • high blood pressure • ejection fraction	0.6889

4-A List of the lessons learned

4(a)-When do visualization help in this example?

We can use the plot to determine which modeling technique to use and see the relationships between variables.

When do visualization not help in this example?

When the Death Events (0,1) are too overlapped, it would be difficult to distinguish patterns.
If variables have no clear correlation with Death Events, the scatter plots become uninformative.

4(b)-Which choice model technique (perceptron and SVM) work in this example?

SVM

Which did not?

Perceptron

Why?

The perceptron can only solve problems that are linearly separable

4(c)-What other technique could you use in this example?

Decision Tree

Thanks!

Machine Learning

-- HeartFailure Analysis

Group 6