

# STAT542, Homework 3

*Shuhui Guo*

## Problem 1

a)

After formulating the primal of the linear separable SVM optimization problem, we can solve it by solve.QP and get the estimations of  $\beta_0, \beta_1, \beta_2$ :

$$\hat{\beta}_0 = 2.7824$$

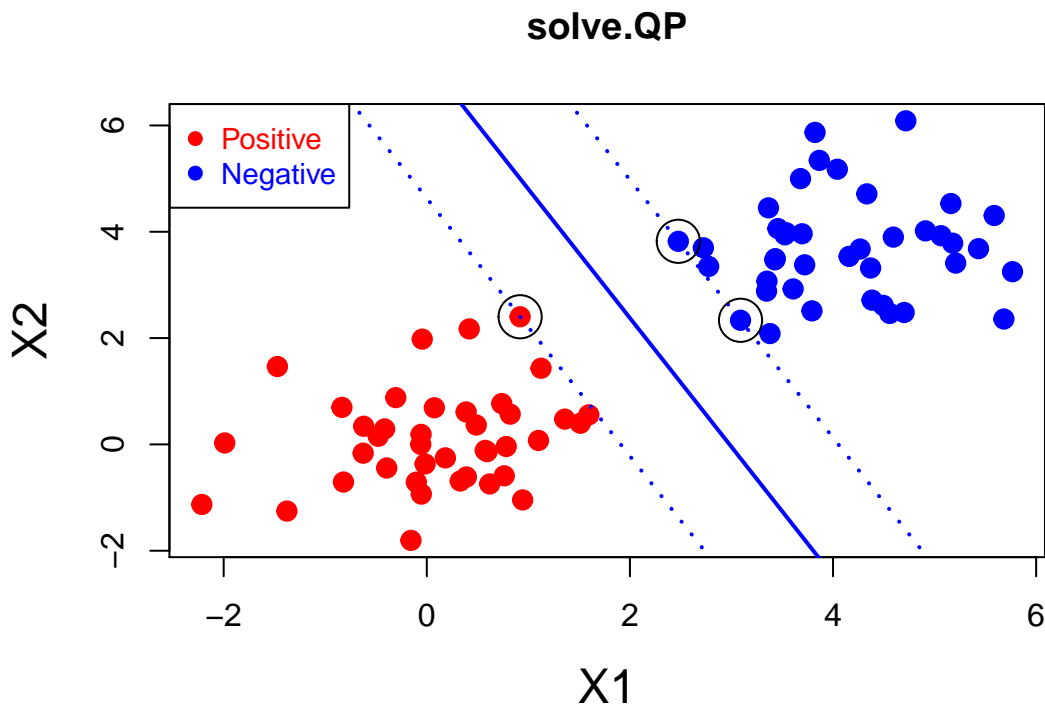
$$\hat{\beta}_1 = -0.9334$$

$$\hat{\beta}_2 = -0.385$$

Let  $y_i(x_i^T \beta + \beta_0) = 1$ , we can get the support vectors:

$$(0.919, 2.4016), (2.4764, 3.8204), (3.0891, 2.335)$$

Plot the results produced by solve.QP as follow:



In this plot, the blue solid line is the decision line using the results produced by solve.QP. The circled points are the support vectors.

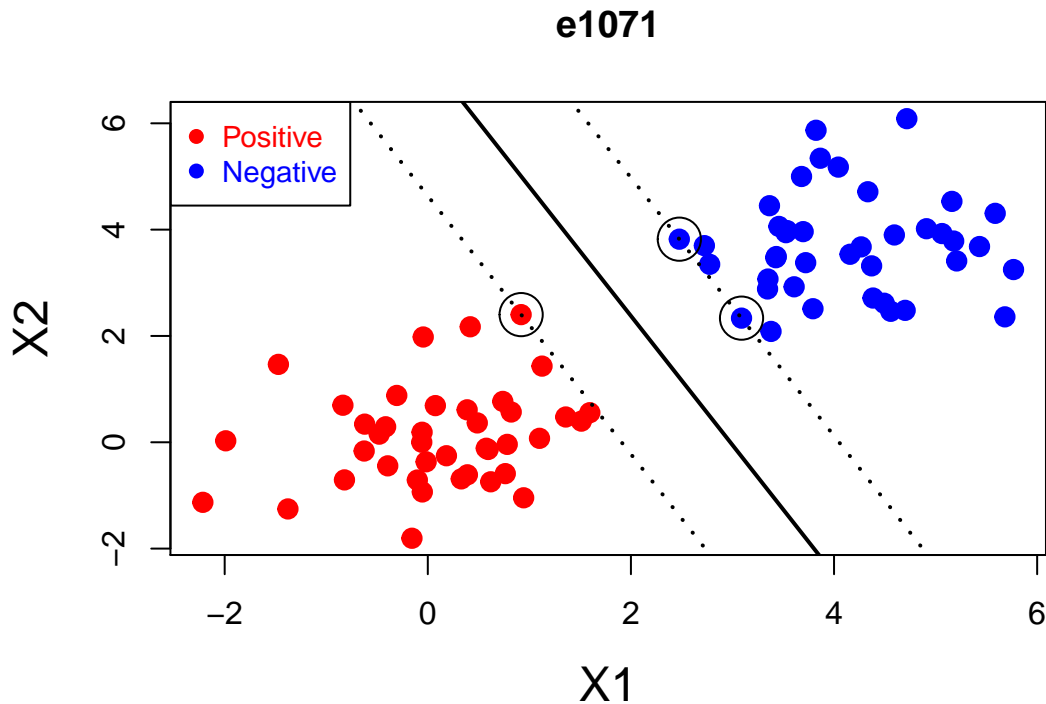
Then use e1071 package to solve the primal of the linear separable SVM optimization problem. We can get the estimations of  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$ :

$$\begin{aligned}\hat{\beta}_0 &= 2.782 \\ \hat{\beta}_1 &= -0.9339 \\ \hat{\beta}_2 &= -0.3845\end{aligned}$$

And the support vectors:

$$(0.919, 2.4016), (2.4764, 3.8204), (3.0891, 2.335)$$

Plot the results produced by e1071 package as follow:



In this plot, the black solid line is the decision line using the results produced by e1071 package. The circled points are the support vectors.

From the above report, we can see that my solutions and the results produced by e1071 package are almost the same.

b)

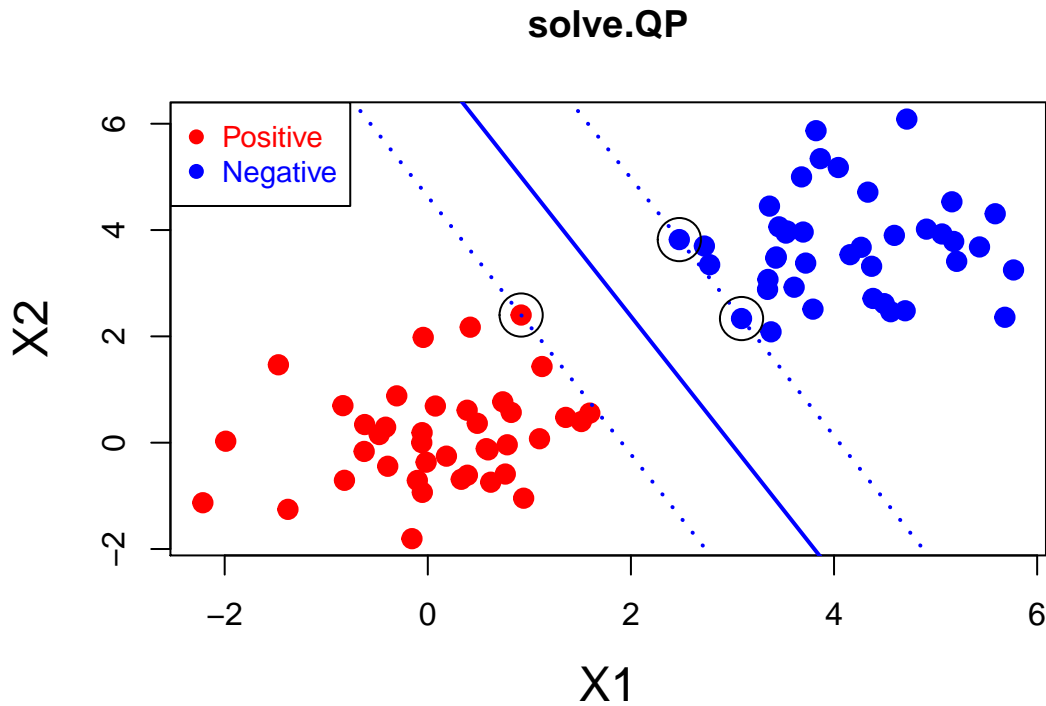
After formulating the dual form of the linear separable SVM, we can solve it by solve.QP and get the estimations of  $\beta_0, \beta_1, \beta_2$ :

$$\begin{aligned}\hat{\beta}_0 &= 2.7824 \\ \hat{\beta}_1 &= -0.9334 \\ \hat{\beta}_2 &= -0.385\end{aligned}$$

Let  $\alpha_i > 0$ , we can get the support vectors:

$$(0.919, 2.4016), (2.4764, 3.8204), (3.0891, 2.335)$$

Plot the results produced by solve.QP as follow:



In this plot, the blue solid line is the decision line using the results produced by solve.QP. The circled points are the support vectors.

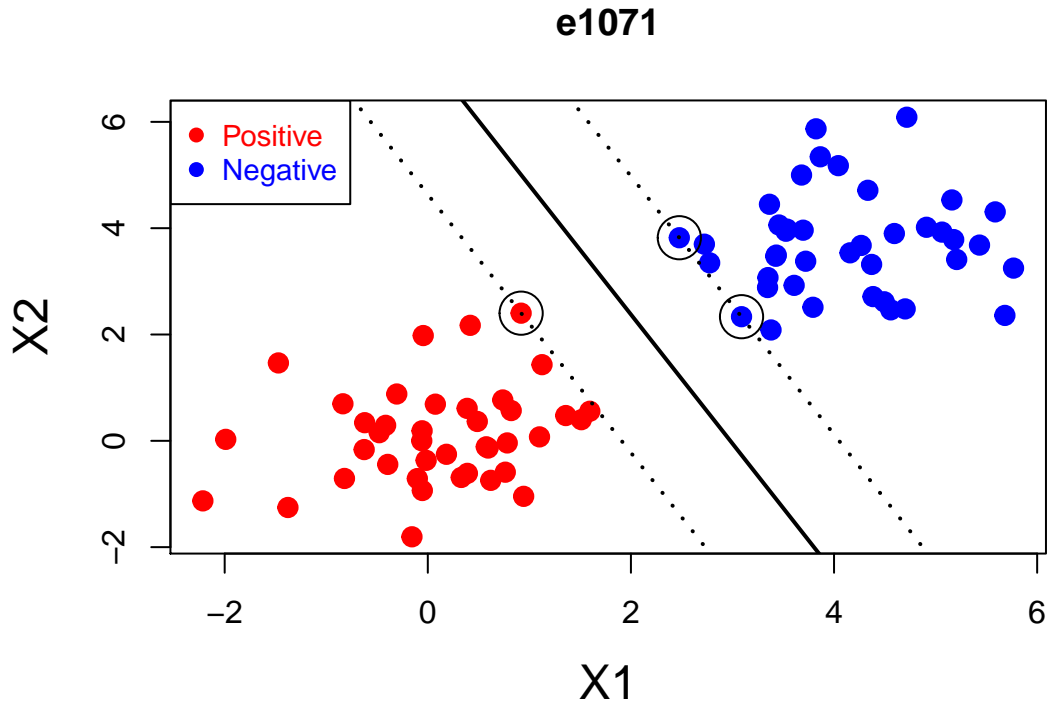
Then use e1071 package to solve this. We can get the estimations of  $\beta_0, \beta_1, \beta_2$ :

$$\begin{aligned}\hat{\beta}_0 &= 2.782 \\ \hat{\beta}_1 &= -0.9339 \\ \hat{\beta}_2 &= -0.3845\end{aligned}$$

And the support vectors:

$$(0.919, 2.4016), (2.4764, 3.8204), (3.0891, 2.335)$$

Plot the results produced by e1071 package as follow:



In this plot, the black solid line is the decision line using the results produced by e1071 package. The circled points are the support vectors.

From the above report, we can see that my solutions and the results produced by e1071 package are almost the same.

c)

Generate the nonseparable data:

```

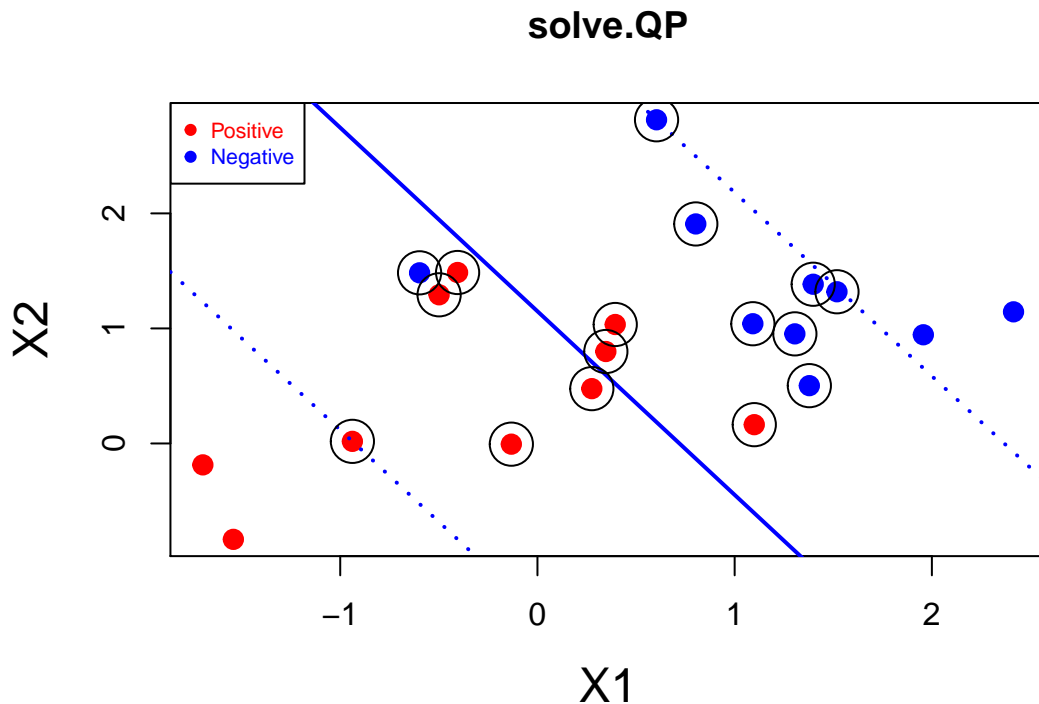
set.seed(70)
n <- 10 # number of data points for each class
p <- 2 # dimension
xpos <- matrix(rnorm(n*p,mean=0,sd=1),n,p)
xneg <- matrix(rnorm(n*p,mean=1.5,sd=1),n,p)
x <- rbind(xpos,xneg)
y <- matrix(c(rep(1,n),rep(-1,n)))

```

Set  $C = 0.1$ . After formulating the dual form of the linear nonseparable SVM, we can solve it by solve.QP and get the estimations of  $\beta_0, \beta_1, \beta_2$ :

$$\begin{aligned}\hat{\beta}_0 &= 0.4372 \\ \hat{\beta}_1 &= -0.6072 \\ \hat{\beta}_2 &= -0.3805\end{aligned}$$

Let  $0 < \alpha_i < C$ , we can get the support vectors. Plot the results produced by solve.QP as follow:



In this plot, the blue solid line is the separating line using the results produced by solve.QP. The circled points are the support vectors.

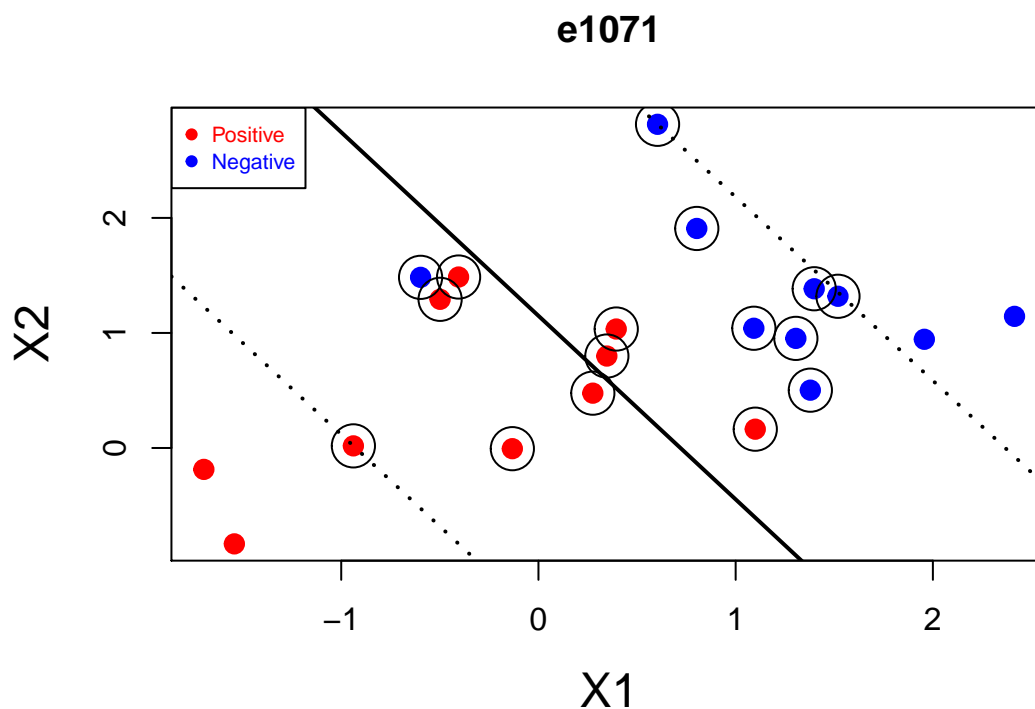
Then use e1071 package to solve this. We can get the estimations of  $\beta_0, \beta_1, \beta_2$ :

$$\hat{\beta}_0 = 0.4372$$

$$\hat{\beta}_1 = -0.6072$$

$$\hat{\beta}_2 = -0.3805$$

Plot the results produced by e1071 package as follow:



In this plot, the black solid line is the separating line using the results produced by e1071 package. The circled points are the support vectors.

From the above report, we can see that my solutions and the results produced by e1071 package are almost the same.

d)

In this problem, I firstly shuffled the original dataset, then used my SVM code to do 5-fold cross validation with different  $C$  and found the highest prediction accuracy is 0.721. The corresponding tuning parameter can be  $C = 5$ .

To select the tuning parameter, I tried different  $C$  in the interval  $[0.01, 100]$  and found the prediction accuracy fluctuated around 0.72. Also, when  $C \geq 5$ , the prediction accuracies are very close. So in the previous paragraph, I say that  $C = 5$  can be the tuning parameter. And I choose  $C = 0.05, 0.1, 0.5, 1, 5$  to report the results:

*When  $C = 0.05$ , the prediction accuracy is 0.708*  
*When  $C = 0.10$ , the prediction accuracy is 0.712*  
*When  $C = 0.50$ , the prediction accuracy is 0.718*  
*When  $C = 1.00$ , the prediction accuracy is 0.721*  
*When  $C = 5.00$ , the prediction accuracy is 0.721*

I then used e1071 package to do 5-fold cross validation and found the highest prediction accuracy is 0.727. The corresponding tuning parameter is  $C = 0.1$ .

The report of the results produced by e1071 package with  $C = 0.05, 0.1, 0.5, 1, 5$  is as follow:

*When  $C = 0.05$ , the prediction accuracy is 0.714*  
*When  $C = 0.10$ , the prediction accuracy is 0.727*  
*When  $C = 0.50$ , the prediction accuracy is 0.725*  
*When  $C = 1.00$ , the prediction accuracy is 0.716*  
*When  $C = 5.00$ , the prediction accuracy is 0.712*

From the above report, we can see that the results produced by my SVM code and the results produced by e1071 package are very close.

## Problem 2

When  $X \leq \xi_1$ ,

$$f(X) = \sum_{j=0}^3 \beta_j X^j$$

Since  $f(X)$  is linear for  $X \leq \xi_1$ ,  $f'(X)$  is a constant. So

$$f''(X) = 2\beta_2 + 6\beta_3 X = 0$$

To make this equation holds for all  $X \leq \xi_1$ , we can get  $\beta_2$  and  $\beta_3$  are both 0. So 1) is proved. Hence,  $f(X)$  can be written as:

$$f(X) = \beta_0 + \beta_1 X + \sum_{k=1}^K \theta_k (X - \xi_k)_+^3$$

When  $X \geq \xi_K$ ,

$$f(X) = \beta_0 + \beta_1 X + \sum_{k=1}^K \theta_k (X - \xi_k)^3$$

Since  $f(X)$  is linear for  $X \geq \xi_K$ ,  $f'(X)$  is a constant. So

$$f''(X) = 6 \sum_{k=1}^K \theta_k (X - \xi_k) = 6 \left( \sum_{k=1}^K \theta_k X - \sum_{k=1}^K \theta_k \xi_k \right) = 0$$

To make this equation holds for all  $X \geq \xi_K$ , we can get  $\sum_{k=1}^K \theta_k = 0$  and  $\sum_{k=1}^K \theta_k \xi_k = 0$ . So 2) and 3) are proved.

With the results established, we can get

$$\begin{aligned} \sum_{k=1}^{K-2} \alpha_k d_k(X) &= \sum_{k=1}^{K-2} \theta_k [(X - \xi_k)_+^3 - (X - \xi_K)_+^3] \\ &= \sum_{k=1}^{K-2} \theta_k (X - \xi_k)_+^3 - (X - \xi_K)_+^3 \sum_{k=1}^{K-2} \theta_k \\ &= \sum_{k=1}^{K-2} \theta_k (X - \xi_k)_+^3 - (X - \xi_K)_+^3 (0 - \theta_K - \theta_{K-1}) \\ &= \sum_{k=1}^{K-2} \theta_k (X - \xi_k)_+^3 + (\theta_K + \theta_{K-1})(X - \xi_K)_+^3 \end{aligned}$$

And

$$\begin{aligned} \sum_{k=1}^{K-2} \alpha_k d_{K-1}(X) &= \left( \sum_{k=1}^{K-2} \theta_k \xi_K - \sum_{k=1}^{K-2} \theta_k \xi_k \right) d_{K-1}(X) \\ &= [0 - (\theta_K + \theta_{K-1})] \xi_K d_{K-1}(X) - [0 - (\theta_K \xi_K + \theta_{K-1} \xi_{K-1})] d_{K-1}(X) \\ &= -(\xi_K - \xi_{K-1}) \theta_{K-1} d_{K-1}(X) \\ &= -\theta_{K-1} [(X - \xi_{K-1})_+^3 - (X - \xi_K)_+^3] \end{aligned}$$

Therefore,



$$\begin{aligned}
\sum_{k=1}^{K-2} \alpha_k (d_k(X) - d_{K-1}(X)) &= \sum_{k=1}^{K-2} \alpha_k d_k(X) - \sum_{k=1}^{K-2} \alpha_k d_{K-1}(X) \\
&= \sum_{k=1}^{K-2} \theta_k (X - \xi_k)_+^3 + (\theta_K + \theta_{K-1})(X - \xi_K)_+^3 \\
&\quad + \theta_{K-1}[(X - \xi_{K-1})_+^3 - (X - \xi_K)_+^3] \\
&= \sum_{k=1}^K \theta_k (X - \xi_k)_+^3
\end{aligned}$$

So we have

$$\beta_0 + \beta_1 X + \sum_{k=1}^K \theta_k (X - \xi_k)_+^3 = \beta_0 + \beta_1 X + \sum_{k=1}^{K-2} \alpha_k (d_k(X) - d_{K-1}(X))$$

So the power series representation can be rewrote as

$$f(X) = \beta_0 + \beta_1 X + \sum_{k=1}^{K-2} \alpha_k (d_k(X) - d_{K-1}(X))$$

## Problem 3

### I. Introduction

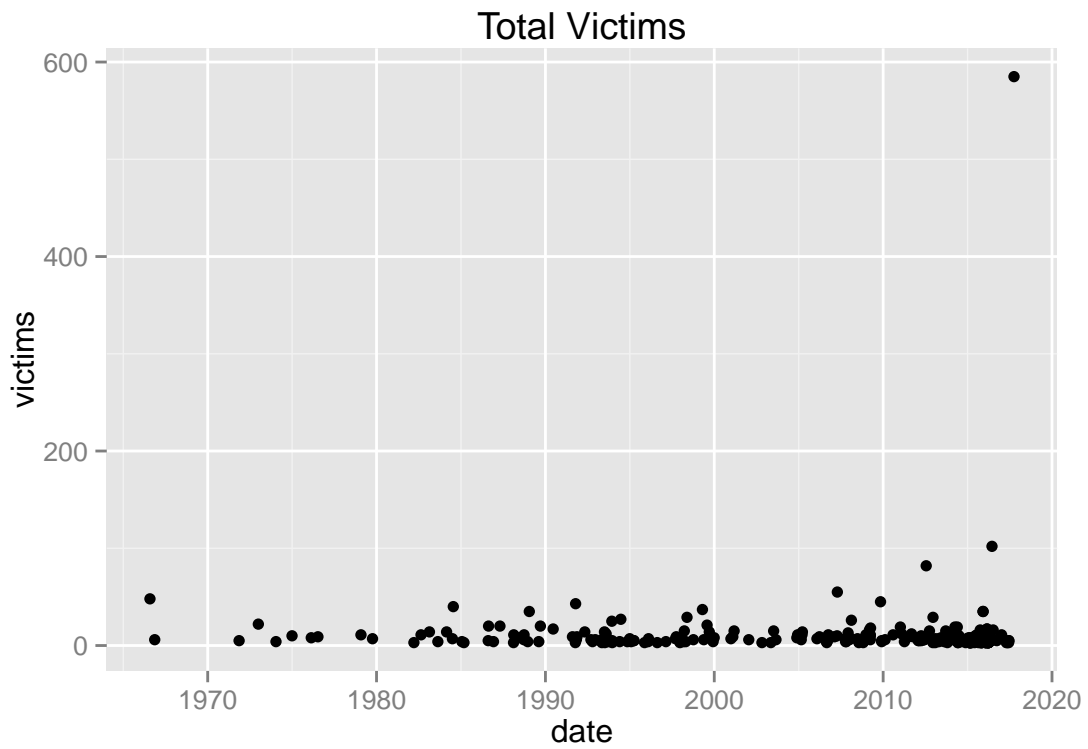
The dataset in this part collects detailed information of 398 mass shootings in the United States of America in 1966-2017. These mass shootings resulted in 1996 deaths and 2488 injured. The average number of mass shootings per year is 7 for the last 50 years.

### II. Question

In this part, I would like to calculate the number of victims per year and explore the trend of total victims in the latest years.

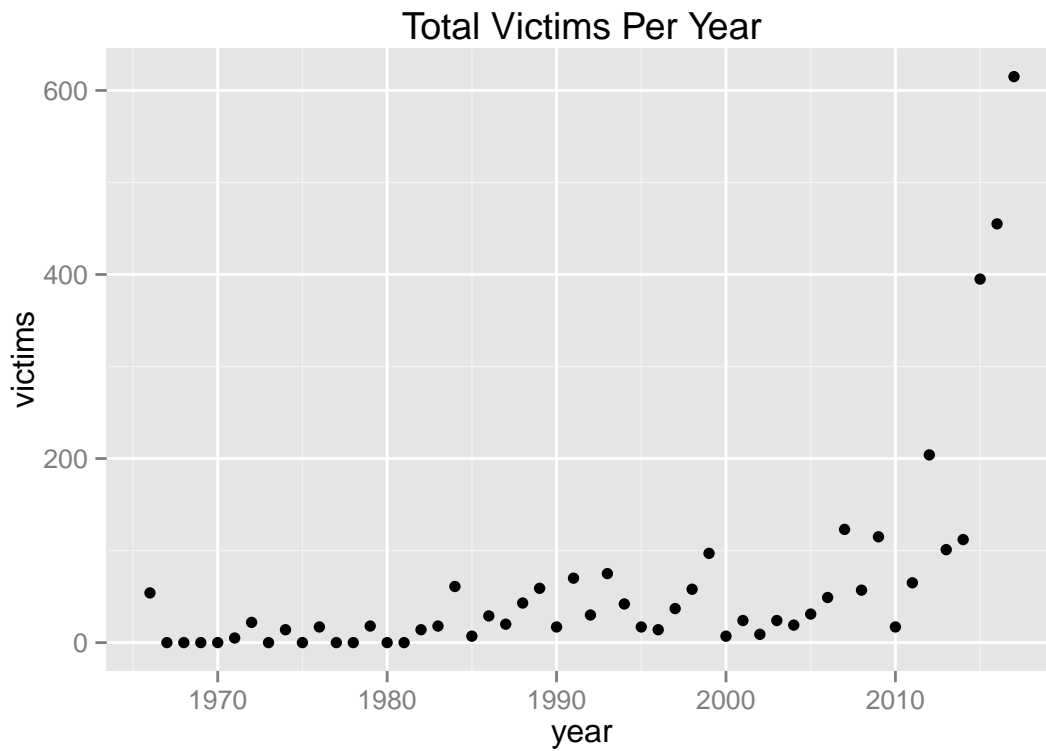
### III. Description Analysis

The scattered plot of the total victims in every incident during 1966-2017 is as follow:



From this plot, we can see that the number of total victims fluctuates around a relatively small value before 2010. Nevertheless, after 2010, the number of victims increased. The latest and also the worst mass shooting of October 2, 2017 caused 585 victims.

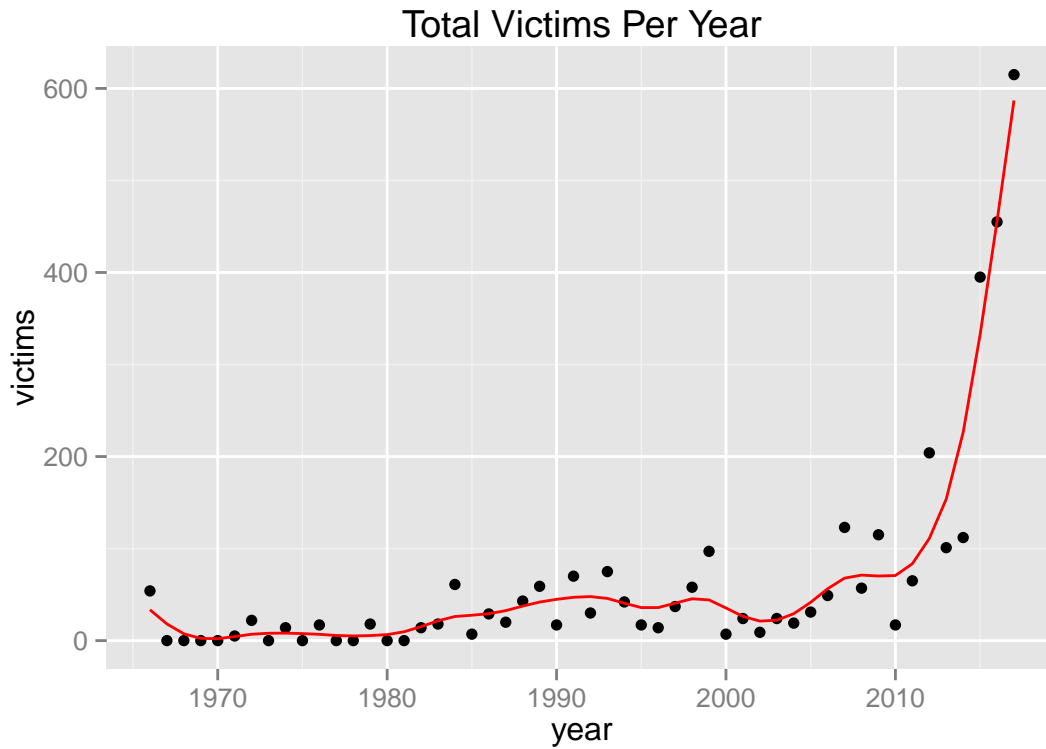
Then we can calculate the number of victims per year and plot the results:



From this plot, we can see that the number of victims per year is fluctuating with an increasing trend. After the year of 2010, the increasing trend is more obvious and the number of victims reaches the highest in 2017.

#### IV. Model Fitting

To explore the trend of total victims in the latest years, we choose the Smoothing Spline to fit the number of victims per year. While doing the model fitting, the Smoothing Spline method can select the number and location of knots automatically. Plot the fitting line as follow:



In this plot, the red line is the fitting line. We can see the increasing trend of the number of victims. After the year of 2010, there is an obvious increase which means the number of victims in the mass shooting increased rapidly.

## V. Conclusion

From the above discussions, we can obtain the number of victims per year and find the trend of the number of victims. Then we can draw the conclusion that the general trend is increasing. Especially after the year of 2010, the increasing trend is obvious, which means the number of victims in 2010-2017 increased rapidly.