Spring 2017

Support Vector Machine and Its Application to Regression and Classification

Xiaotong Hu

SUPPORT VECTOR MACHINE AND ITS APPLICATION TO REGRESSION AND CLASSIFICATION

A Masters Thesis

Presented to

The Graduate College of

Missouri State University

In Partial Fulfillment

Of the Requirements for the Degree

Master of Science, Mathematics

By

Xiaotong Hu

April 2017

SUPPORT VECTOR MACHINE AND ITS APPLICATION TO REGRESSION AND CLASSIFICATION

Mathematics

Missouri State University, April 2017

Master of Science

Xiaotong Hu

ABSTRACT

Support Vector machine is currently a hot topic in the statistical learning area and is now widely used in data classification and regression modeling. In this thesis, we introduce the basic idea for support vector machine, its application in the classification area including both linear and nonlinear parts, and the idea of support vector regression contains the comparison of loss functions and the usage of kernel function. Two real life examples, which are taken from R package, are also provided for both classification and regression part respectively, talking about classification of glass type and prediction for Ozone pollution.

KEYWORDS: support vector machine, data Classification, regression model, hyperplane, statistical learning

This abstract is approved as to form and content

_____

Songfeng Zheng, PhD
Chairperson, Advisory Committee
Missouri State University

SUPPORT VECTOR MACHINE AND ITS APPLICATION

TO REGRESSION AND CLASSIFICATION

By

Xiaotong Hu

A Master Thesis
Submitted to the Graduate College
Of Missouri State University
In Partial Fulfillment of the Requirements
For the Degree of Master of Science, Mathematics

May, 2017

Approved:

_____
Songfeng Zheng, PhD

_____
Yingcai Su, PhD

_____
George Mathew, PhD

_____
Julie Masterson, PhD: Dean, Graduate College

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. INTRODUCTION

1.1 Basic Idea of Statistical Learning Theory

Statistical learning theory is a framework for machine learning area related to the fields of statistics and functional analysis. The goal for statistical learning is find a fittest function f, representing the systematic information of the response from its predictors, which will be used for prediction and inference.

To estimate this unknown function, the approach used in statistical learning is called training data, whose name come from the usage that train our method how to estimate f. Every point in the training data is a input-output pair, where the input maps an output. The statistical leaning problem consist of inferring the function that maps between the input and the output, such that the learned function can be used to predict output from future input.

Statistical learning theory is widely used in regression modeling and data classification. In the regression problem, the responses are quantitative values which take a continuous range of values, meanwhile, the responses in classification problem are qualitative values which are elements from a discrete set of labels. In the statistical learning area, the approach to find the function f are different in many aspects, for example, the choice of loss function, however, they also share some similarities. In the following chapter, we will go further in both of these two areas.

1.2 Introduction to Support Vector Machine

Support Vector machine are currently a hot topic in the statistical learning area. In late 1990s, the traditional neural network approaches suffered severe difficulties with generalization and producing models which is the main reason for the foundation of

support vector machines. It was developed in 1995 by Vladimir Vapnik, and soon gained popularity due to many attractive features.

In statistical learning, support vector machines are supervised learning method with assoxiated leaning algorithms that analyze dataset. It is first been introduced as an method for solving classification problems. However, due to many attractive features, it is recently extended to the area of regression analysis.

This thesis will mainly focus on the application of support vector machine used in classification and regression area and is structured as follows.

Chapter 2 introduces the basic idea of classification, including the concept of hyperplane, different type of classifier and their properties, and will be focused on the support vector classification in both linear and non-linear condition, and its application in real life examples operating by computational languages R.

Chapter 3 will focus on support vector regression which is separated into linear and non-linear regression. In both these two sections, it will introduce different type and choice for loss function and how will it influence the performance of regression modeling. It will also introduce an example at the end of these chapter to show how support vector regression deals with real life problem.

CHAPTER 2. SUPPORT VECTOR CLASSIFICATION

2.1 Hyperplane

In a p-dimensional space, a hyperplane is a flat affine subspace (a subspace need not pass through the origin) of dimension p-1. It is defined by equation

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p = 0 \qquad (2.1)$$

for parameter $\beta_1 \ldots \beta_p$. In this case, the point X=(x_1,x_2,...,x_p) lies on the hyperplane.

Now, suppose X does not satisfy 2.1, but

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p > 0 \qquad (2.2)$$

then we say X lies on the one side of the hyperplane. On the other hand, if

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p < 0 \qquad (2.3)$$

then X lies on the other side of the hyperplane.

Here we can consider that the hyperplane divides this p-dimensional space into two halves, which will be the basic idea of classification.

2.2 Maximal Margin Classifier

2.2.1 Classification using a Separating Hyperplane

Suppose we have a $n \times p$ data matrix that contains n observation in p-dimensional space

$$x_1 = \begin{pmatrix} x_{1\,1} \\ \vdots \\ x_{1\,p} \end{pmatrix}, \ldots, x_n = \begin{pmatrix} x_{n\,1} \\ \vdots \\ x_{n\,p} \end{pmatrix}$$

What we want is to separate these data into two class. The approach is to label $y_1,\ldots,y_n \in \{-1,1\}$, where -1 represent a class and 1 represent the other class. Our goal is to develop a classifier that can classify the observation using its feature measurement, and

it is based on the concept of separating hyperplane (James, Witten, Hastie & Tibshirani, 2009).

The separating hyperplane with parameter $\beta_1 \cdots \beta_p$ has the property that

$$\begin{cases} \beta_0 + \beta_1 x_{i\,1} + \beta_2 x_{i\,2} + \ldots + \beta_p x_{i\,p} > 0, & if\ y_i = 1 \\ \beta_0 + \beta_1 x_{i\,1} + \beta_2 x_{i\,2} + \ldots + \beta_p x_{i\,p} < 0, & if\ y_i = -1 \end{cases} \qquad (2.4)$$

which equivalently means

$$y_i(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p) > 0 \qquad (2.5)$$

Thus, if a hyperplane exist, the observation dataset can be assigned a class based on which side if the hyperplane it is located. In figure 2.1, we can clearly classify the observation x_i based on hyperplane $\beta_0 + \beta_1 x_1 + \beta_2 x_2 = 0$. That is, if $f(x_i) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ is positive, it is assigned to class 1; if negative, on the other hand, it is assigned to class -1.

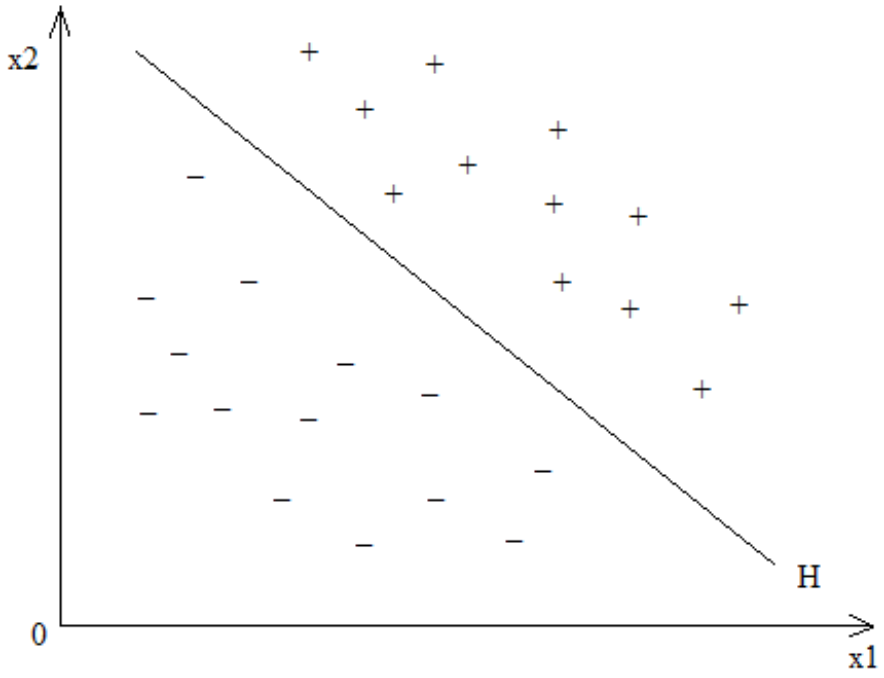

Figure 2.1 Separating Hyperplane $\beta_0 + \beta_1 x_1 + \beta_2 x_2 = 0$ in 2-dimensional space

2.2.2 The Maximal Margin Classifier

In general, if our data can be perfectly separated using a hyperplane, then there will exist infinitely many such hyperplane. In order to decide which separating hyperplane to use, we need to use the maximal margin classifier, which is also known as the optimal separating hyperplane. The maximal margin classifier is farthest from the training observation. That is, when we compute the distance of each observation to the separating hyperplane, the minimal one is what we called 'the margin'. The maximal margin classifier is the hyperplane for which the margin is the largest (James, Witten, Hastie & Tibshirani, 2009).

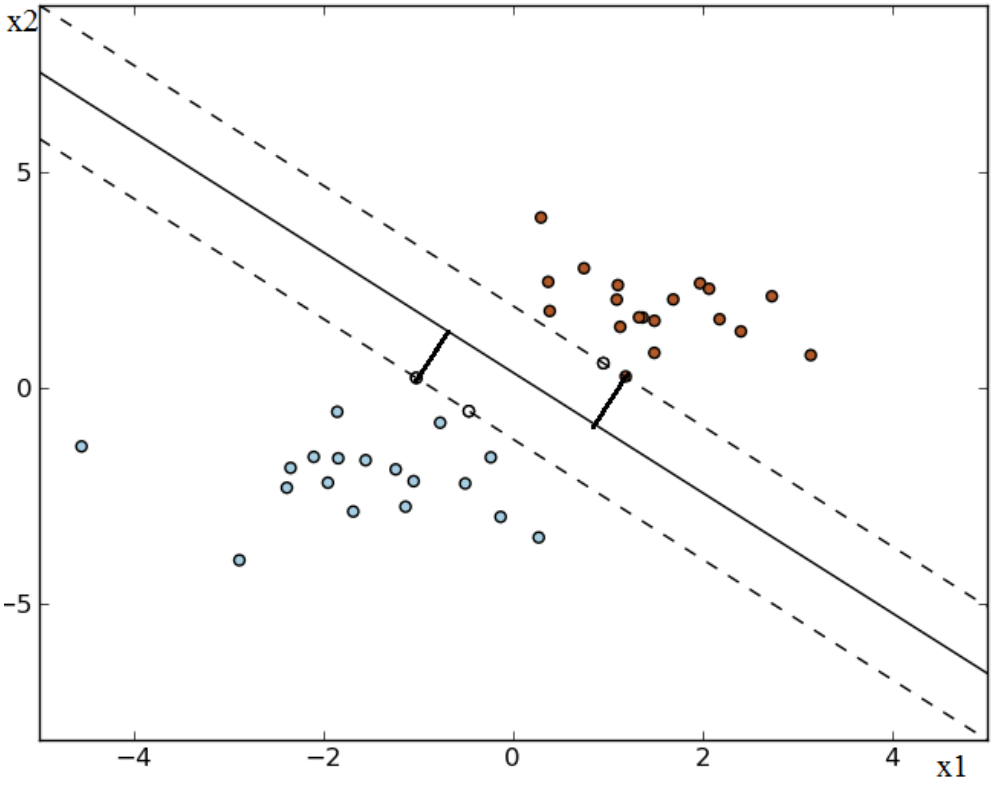

Figure 2.2 Maximal Margin Classifier and the margin in 2-dimisional space

In general, the maximal margin classifier is the solution to the optimization problem

$$\max_{\beta_0,\beta_1,\ldots,\beta_p} M \qquad (2.6)$$

$$subject\ to \sum_{j=1}^{p} \beta_j^2 = 1 \qquad (2.7)$$

$$y_i(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p) \geq M\ for\ all\ i = 1,\ldots,n \qquad (2.8)$$

Equation 2.5 guarantees M is positive, and with constraints from 2.6 to 2.8, the perpendicular distance for ith observation to the hyperplane is actually given by $y_i(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p)$. Thus, M is the margin of the hyperplane, and the optimization problem is exactly the definition of the maximal margin hyperplane.

2.3 Support Vector Classifier

2.3.1 Overview of Support vector Classifier

Sometimes in practical situation, the observations that belong to two classes are not necessarily separable by the hyperplane. In fact, the maximal margin hyperplane may not be desirable even if it does exist. A classifier based on the maximal margin hyperplane could be changed dramatically with addition or miss one single observation which may result in overfitting the training data. An example is shown in figure 2.3, where a single observation could dramatically change the hyperplane from the left-handle panel to the right handle panel.

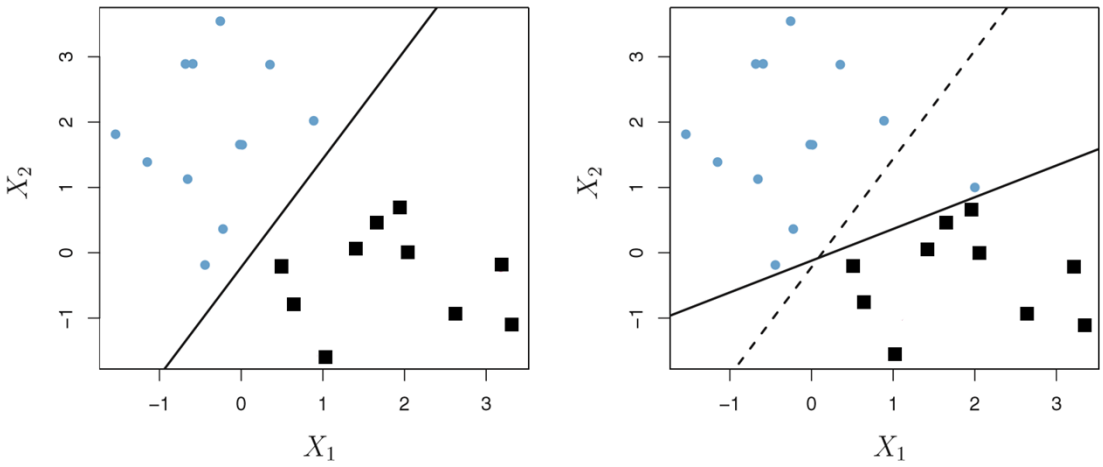

Figure 2.3 Hyperplane change leaded by a single observation

In this case, another approach to classify the observation, which is known as soft margin classifier or support vector classifier is needed. Unlike maximal margin classifier, the soft margin classifier allows some observation on the wrong side of the margin or even on the wrong side of the hyperplane.

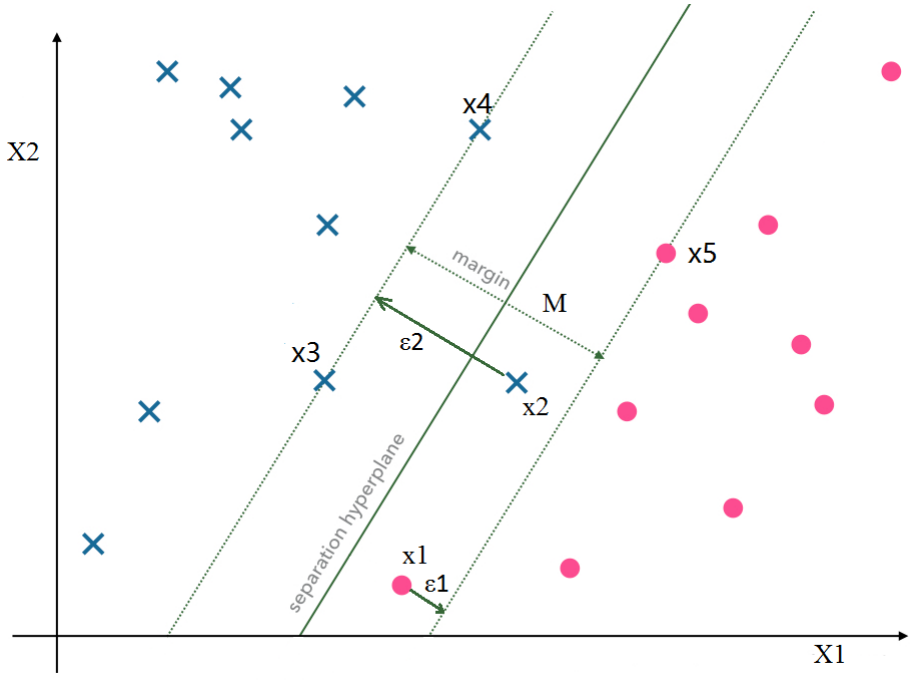

Figure 2.4 Soft Margin Classifier

Figure 2.4 shows an example of soft margin classifier. Here, observation x1 is on the wrong side of margin but right side of hyperplane, and observation x2 is on both wrong side of margin and hyperplane. Observation x3, 4 and 5, which lie on the edge of the margin, these 3 observations and observation x1 and x2 which lie inside the soft margin are what we called 'the support vectors', and we will detail this part in 2.3.2.

Similar with the maximal margin classifier, the soft margin classifier is the solution to the optimization problem

$$\max_{\beta_0,\beta_1,\ldots,\beta_p} M \qquad (2.9)$$

$$subject\ to \sum_{j=1}^{p} \beta_j^2 = 1 \qquad (2.10)$$

$$y_i(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p) \geq M(1 - \varepsilon_i) \quad for\ all\ i = 1,\ldots,n \qquad (2.11)$$

$$\varepsilon_i \geq 0, \sum_{i=1}^{n} \varepsilon_i \leq C \qquad (2.12)$$

where C is a nonnegative tuning parameter. As showing in Figure 2.3, M is the width of the margin, which we seek to make this value as large as possible. ε_i is what we called ' slack variables' that allows individual observation to be on the wrong side of margin or hyperplane. By the constraints shows from 2.9 to 2.11, we can easily tell that if $0 \leq \varepsilon_i \leq 1$, the observation is on the wrong side of margin but right side of hyperplane (x_1 in Figure 2.3 with slack variable ε_1); if $\varepsilon_i \geq 1$, the observation is on the both wrong side of margin and hyperplane (x_2 in Figure 2.3 with slack variable ε_2) .

2.3.2 Computing the Support Vector Classifier

We can rewrite 2.9-2.12 into

$$\max_{\beta,\beta_0,||\beta||=1} M \qquad (2.13)$$

$$y_i(x_i^T\beta + \beta_0) \geq M(1-\varepsilon_i) \quad for\ all\ i = 1,...,n \qquad (2.14)$$

$$\varepsilon_i \geq 0, \sum_{i=1}^{n} \varepsilon_i \leq C \qquad (2.15)$$

where $x_i \in R^p$ with unit vector $||\beta|| = 1$. If we then drop the norm constraint on β, where we define $M = 1/||\beta||$, 2.13-2.15 will be represented as

$$min||\beta|| \qquad (2.16)$$

subject to $y_i(x_i^T\beta + \beta_0) \geq 1 - \varepsilon_i \quad for\ all\ i = 1,...,n \qquad (2.17)$

$$\varepsilon_i \geq 0, \sum_{i=1}^{n} \varepsilon_i \leq C \qquad (2.18)$$

The problem 2.16-2.18 is a quadratic with linear inequality constraints, hence, we can describe the programming solution using Lagrange multipliers (Smola & Scholkopf, 2003). The problem is equivalent to

$$\min \frac{1}{2}||\beta||^2 + C\sum_{i=1}^{n} \varepsilon_i \qquad (2.19)$$

subject to $y_i(x_i^T\beta + \beta_0) \geq 1 - \varepsilon_i \quad for\ all\ i = 1,...,n \qquad (2.20)$

$$\varepsilon_i \geq 0, \qquad (2.21)$$

The Lagrange function is

$$L = \frac{1}{2}||\beta||^2 + C\sum_{i=1}^{n} \varepsilon_i - \sum_{i=1}^{N} \alpha_i\left[y_i(x_i^T\beta + \beta_0) - (1-\varepsilon_i)\right] - \sum_{i=1}^{N} \mu_i\varepsilon_i \qquad (2.22)$$

Taking partial derivative to zero, we get

$$\beta = \sum_{i=1}^{N} \alpha_i y_i x_i \qquad (2.23)$$

$$\sum_{i=1}^{N} \alpha_i y_i = 0 \qquad (2.24)$$

$$\alpha_i = C - \mu_i, for\ i = 1,...N \qquad (2.25)$$

where the constraints $\alpha_i, \mu_i, \varepsilon_i \geq 0$. By substituting 2.23-2.25 into 2.22

$$\max L = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j x_i^T x_j \qquad (2.26)$$

$$subject\ to\ 0 \leq \alpha_i \leq C \qquad (2.27)$$

$$\sum_{i=1}^{N} \alpha_i y_i = 0 \qquad (2.28)$$

In addition, the Karush-Kuhn-Trcker conditions that are satisfied by the solution are,

$$\alpha_i \left[y_i (x_i^T \beta + \beta_0) - (1 - \varepsilon_i) \right] = 0 \qquad (2.29)$$

$$\mu_i \varepsilon_i = 0 \qquad (2.30)$$

$$\left[y_i (x_i^T \beta + \beta_0) - (1 - \varepsilon_i) \right] \geq 0 \qquad (2.31)$$

For $i = 1,...N$. All the equations from 2.23-2.29 uniquely characterize the solution to the primal and dual problems (Smola & Scholkopf, 2003).

Meanwhile, from 2.23, we can get the solution for β which is in the form of $\beta = \sum_{i=1}^{N} \alpha_i y_i x_i$, and the observations with constraints in 2.31 are what we called support vectors, which refer to those observation lies in the soft-margin. Among those support

vectors, some meet the constrains in 2.29, which are observations lies on the edge of the margin, and we normally use them to solve β_0.

2.4 Classification with Nonlinear Decision Boundaries

In general, the soft margin classifier is a approach for linear boundary classification. However, we sometimes need to deal with nonlinear boundary in practice when linear classifier cannot perfectly perform. In this case, we need to enlarge the feature space using functions of predictors, like quadratic or cubic. For example, with p observation $x_1,...x_p$, we can fit a vector classifier along with their quadratic form, i.e. $x_1, x_1^2,...,x_p,x_p^2$. Thus, the optimization problem show from 2.9 to 2.12 will become

$$\max_{\beta_0,\beta_{11},...,\beta_{p1},\beta_{12},...\beta_{p2}} \quad M \qquad (2.32)$$

$$subject\ to \sum_{j=1}^{p}\sum_{k}^{2}\beta_{jk}^2 = 1 \qquad (2.33)$$

$$y_i(\beta_0 + \sum_{j=1}^{p}\beta_{j1}x_{ij} + \sum_{j=1}^{p}\beta_{j2}x_{ij}^2) \geq M(1-\varepsilon_i) \quad for\ all\ i = 1,...,n \qquad (2.34)$$

$$\varepsilon_i \geq 0, \sum_{i=1}^{n}\varepsilon_i \leq C \qquad (2.35)$$

The solution will lead to a nonlinear decision boundary, which is known as support vector machine.

The key idea of support vector machine is what we called the kernel (James, Witten, Hastie & Tibshirani, 2009). A kernel is a function which quantifies the similarity

of two observation, representing as $K(x_i, x_{i'})$. For example, we can represent linear support vector classifier by taking inner product as kernel, which means

$$K(x_i, x_{i'}) = \sum_{j=1}^{p} x_{ij} x_{i'j} \qquad (2.36)$$

and the support vector classifier can be represented as

$$f(x) = \beta_0 + \sum_{j=1}^{p} \alpha_i K(x_i, x_{i'}) \qquad (2.37)$$

which is also the general form of support vector machine if $K(x_i, x_{i'})$ is not specific. Thus, if we would like a nonlinear classifier, another form of kernel function will be needed. For instance,

$$K(x_i, x_{i'}) = (1 + \sum_{j=1}^{p} x_{ij} x_{i'j})^d \qquad (2.38)$$

which is known as a polynomial kernel of degree d. When d > 1, the support vector classifier will lead to a more flexible, nonlinear form.

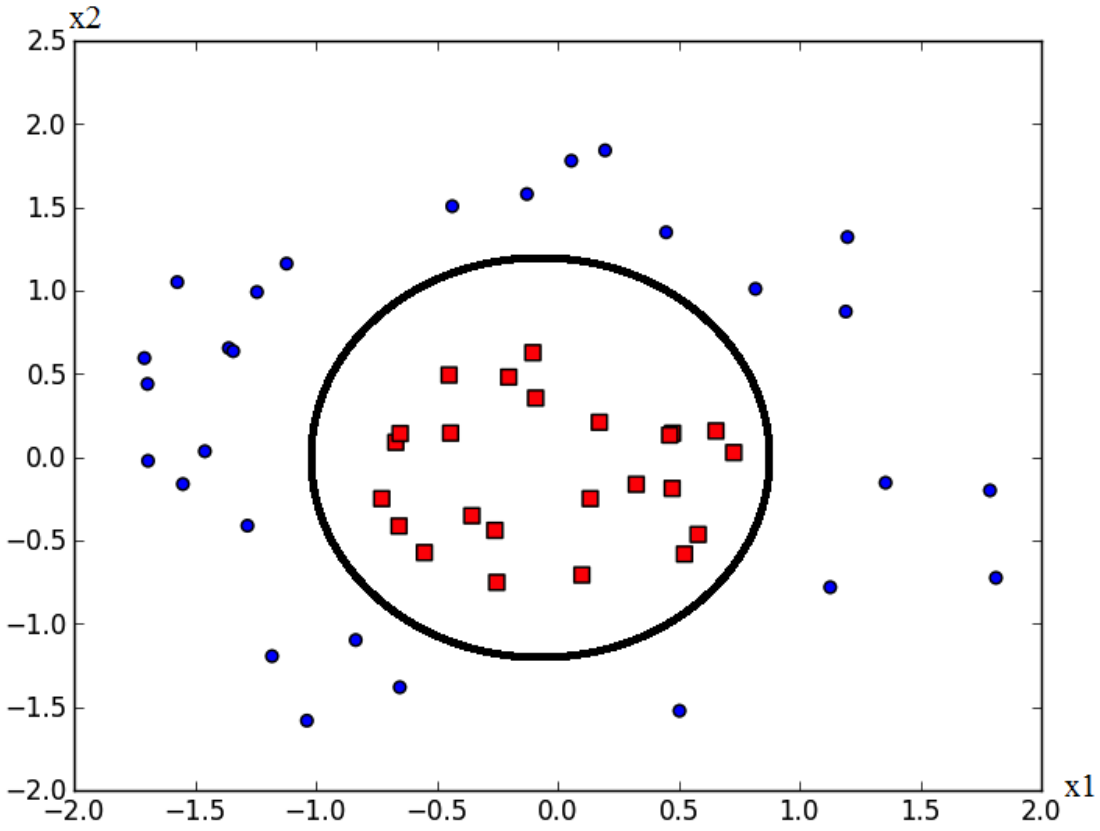

Figure 2.5 Support Vector Machine with polynomial kernel of degree 3

Another popular choice is the radial kernel, also is known as the Gaussian kernel, which will be discussed in Chapter 3, as it is more widely used in Support Vector Regression area.

2.5 Application of Support Vector Classification to Real World Data

The data 'Glass' used in this chapter, taken from UCI Repository Of Machine Learning Databases, containing examples of the chemical analysis of 7 different type of glass. It is a perfect data set for classification and our task is to forecast the type of glass on basis of its chemical analysis.

Table 1. Different Type of Glass and their Chemical Analysis Data (Partial)

Observation	RI	Na	Mg	Al	Si	K	Ca	Ba	Fe	Type
1	1.52101	13.64	4.49	1.1	71.78	0.06	8.75	0	0	1
2	1.51761	13.89	3.6	1.36	72.73	0.48	7.83	0	0	1
3	1.51618	13.53	3.55	1.54	72.99	0.39	7.78	0	0	1
4	1.51766	13.21	3.69	1.29	72.61	0.57	8.22	0	0	1
5	1.51742	13.27	3.62	1.24	73.08	0.55	8.07	0	0	1
6	1.51596	12.79	3.61	1.62	72.97	0.64	8.07	0	0.26	1
7	1.51743	13.3	3.6	1.14	73.09	0.58	8.17	0	0	1
8	1.51756	13.15	3.61	1.05	73.24	0.57	8.24	0	0	1
9	1.51918	14.04	3.58	1.37	72.08	0.56	8.3	0	0	1
10	1.51755	13	3.6	1.36	72.99	0.57	8.4	0	0.11	1
11	1.51571	12.72	3.46	1.56	73.2	0.67	8.09	0	0.24	1
12	1.51763	12.8	3.66	1.27	73.01	0.6	8.56	0	0	1
13	1.51589	12.88	3.43	1.4	73.28	0.69	8.05	0	0.24	1
14	1.51748	12.86	3.56	1.27	73.21	0.54	8.38	0	0.17	1
15	1.51763	12.61	3.59	1.31	73.29	0.58	8.5	0	0	1
16	1.51761	12.81	3.54	1.23	73.24	0.58	8.39	0	0	1
17	1.51784	12.68	3.67	1.16	73.11	0.61	8.7	0	0	1
18	1.52196	14.36	3.85	0.89	71.36	0.15	9.15	0	0	1
19	1.51911	13.9	3.73	1.18	72.12	0.06	8.89	0	0	1
20	1.51735	13.02	3.54	1.69	72.73	0.54	8.44	0	0.07	1
21	1.5175	12.82	3.55	1.49	72.75	0.54	8.52	0	0.19	1

This dataset containing 214 observation on 10 variables, and the following table provides a basic summary of this dataset.

Table 2. Summary of Data ‘Glass’

	Min	1st Qu	Median	Mean	3rd Qu	Max
Ri	1.511	1.517	1.518	1.518	1.519	1.534
Na	10.730	12.910	13.300	13.410	13.820	17.380
Mg	0.000	2.115	3.480	2.685	3.600	4.490
Al	0.290	1.190	1.360	1.445	1.630	3.500
Si	69.810	72.280	72.790	72.650	73.090	75.410
K	0.000	0.123	0.555	0.497	0.610	6.210
Ca	4.430	8.240	8.600	8.957	9.172	16.190
Ba	0.000	0.000	0.000	0.175	0.000	3.150
Fe	0.000	0.000	0.000	0.057	0.100	0.510

Table 3. Quantities for Each Type of Glass

Type	1	2	3	5	6	7

Quantity	70	76	17	13	9	29

We are going to take R package ‘e1071’, which is developed for support vector machine to help us to do the classification and prediction. First of all, we need to start by splitting the data into train and test data set, by randomly select 71 observations (one-third of dataset) as test data and the remains as train data.

In this case, we are going to build our support vector classification nonlinearly, and the kernel we are going to use is radial kernel. Starting with setting cost as 100 and gamma in radial kernel as 1, the support vector classifier will be represented as

$$f(x_i) = \beta_0 + \sum_{j=1}^{p} a_i \exp\left(-||x_i - x_j||^2\right) \qquad (2.39)$$

And by taking code ‘svm’ from package ‘e1071’ in R, we got the summary of our model showing below:

Table 4. Summary of Base Model for Data ‘Glass’

Parameters:
SVM-Type: C-classification
SVM-Kernel: radial
cost: 100
gamma: 1

Class	1	2	3	5	6	7	Total
Number of support vector	49	55	17	11	7	20	159

The number of support vector is a measurement of the performance of a certain dataset fits a classification model. The lower the number is, the better it fits in this model.

The next step is to predict the test value, and the way for us to visualize the performance of the accuracy our prediction is to build the confusion matrix, which is also known as ‘the error matrix’.

By taking code ‘confusionMatrix’ from package ‘caret’, we obtained the confusion matrix and the summary of this model is showing below:

Confusion Matrix

		true					
		1	2	3	5	6	7
	1	6	0	0	0	0	0
	2	1	6	0	2	1	2
predict	3	0	0	0	0	0	0
	5	0	0	0	0	0	0
	6	0	0	0	0	1	0
	7	0	0	0	0	0	2

Table 5. Statistics of Base Model for Data ‘Glass’

Overall Statistics:

Accuracy : 0.7143
95% CI : (0.4782, 0.8872)
No Information Rate : 0.3333
P-Value [Acc > NIR] : 0.0004045

Statistics by Class:

Class 1	Class 2	Class 3	Class 5	Class 6	Class 7

Sensitivity	0.8571	1.0000	N/A	0.0000	0.5000	0.5000
Specificity	1.0000	0.6000	1	1.0000	1.0000	1.0000
Balanced Accuracy	0.9286	0.8000	N/A	0.5000	0.7500	0.7500

The accuracy quantifies the performance of a specific support vector machine classification model Balanced accuracy is calculated as the average of the proportion corrects of each class individually. In our case, the overall accuracy is 0.7143. This means if we are going to predict a dataset, the probability of classifying a certain data point to its actual class is 0.7143.

The training accuracy is determined by several factors: cost, type of the kernel, and the parameters in a certain kernel. In our case, for instance, if we change our cost from 100 to 50, the outcomes will become

Table 6. Summary of Adjusted Model 1 for Data 'Glass'

Parameters:

SVM-Type: C-classification

SVM-Kernel: radial

cost: 50

gamma: 1

Class	1	2	3	5	6	7	Total
Number of support vector	47	53	16	11	8	21	156

It shows from the summary that the number of the support vector is lower than the situation when we set the cost equal to 100. Normally, the number of support vector will decreasing when the cost become lower, since the soft-margin is narrow down.

Also, we can get a another confusion matrix from this model.

Confusion matrix:

		true					
		1	2	3	5	6	7
	1	4	4	0	0	0	0
	2	1	4	1	1	0	0
predict	3	1	0	0	0	0	0
	5	0	0	0	1	0	0
	6	0	0	0	0	1	0
	7	0	0	0	0	0	3

Table 7. Statistics of Adjusted Model 1 for Data 'Glass'

Overall Statistics

Accuracy : 0.619

95% CI : (0.3844, 0.8189)

No Information Rate : 0.381

P-Value [Acc > NIR] : 0.2313

Statistics by Class:

Class 1	Class 2	Class 3	Class 5	Class 6	Class 7

Sensitivity	0.6667	0.5000	0.0000	0.5000	1.0000	1.0000
Specificity	0.7333	0.7692	0.9500	1.0000	1.0000	1.0000
Balanced Accuracy	0.7000	0.6346	0.4750	0.7500	1.0000	1.0000

Another factor that can affect the training accuracy is the parameter in the kernel. In our case, gamma is the only parameter which can affect the result. For example, if we change gamma from 1 to 0.5, the outcome will become

Table 8. Summary of Adjusted Model 2 for Data 'Glass'

Parameters:

SVM-Type: C-classification

SVM-Kernel: radial

cost: 100

gamma: 0.5

Class	1	2	3	5	6	7	Total
Number of support vector	43	49	15	11	8	20	146

Same as the situation when we change the cost from 100 to 50, the number of support vector also decreases when we lower the parameter gamma, since it is also a factor which can affect the range of soft-margin. The summary of confusion matrix shows below:

Confusion matrix:

		true					
		1	2	3	5	6	7
predict	1	5	3	0	0	0	0
	2	0	4	1	1	0	0
	3	1	1	0	0	0	0

5	0	0	0	1	0	0
6	0	0	0	0	1	0
7	0	0	0	0	0	3

Table 9. Statistics of Adjusted Model 2 for Data 'Glass'

Overall Statistics:

Accuracy : 0.6667

95% CI : (0.4303, 0.8541)

No Information Rate : 0.381

P-Value [Acc > NIR] : 0.00751

Statistics by Class:

	Class 1	Class 2	Class 3	Class 5	Class 6	Class 7
Sensitivity	0.8333	0.5000	0.0000	0.5000	1.0000	1.0000
Specificity	0.8000	0.8462	0.9000	1.0000	1.0000	1.0000
Balanced Accuracy	0.8167	0.6731	0.4500	0.7500	1.0000	1.0000

We can realized in these two cases the training accuracy is lower than the situation when we set the cost equal to 100 and gamma equal to 1. This is common, even though in both cases, the number of support vector do decrease. This situation is called 'overfitting', which occurs when the model is complex, i.e., has too many parameters related to the observation.

Our task is to fit the model to the dataset as accurate as possible, which means we need to find a certain combination of cost and gamma in the kernel. The approach which is commonly used is cross-validation, a model validation technique for assessing how the results of statistical analysis will generalize to an independent data set. It is mainly used in settings where the goal is prediction, and one wants to estimate

how accurately a predictive model will perform in practice. Here we are going to take the approach 10-fold cross-validation, which randomly separate the original dataset in to 10 subset, and on each time select one as a test data and the other nine as training data.

In R, we can write a 'for loop' to build this cross-validation. We could set sequence which allows the cost runs from 100 to 1 for 100 times, and set gamma from 0.01 from 1 for 100 times. In each combination, the for loops allows R to test our dataset 10 times where randomly selected 1 subset from the original dataset as the test dataset. This means the R system need to loop $100 \times 100 \times 10$testing subgroup to get the best combination of parameter gamma and constant cost. In this test, the base case is the situation we set gamma equal to 1 and cost equal to 10. The result after this 10-fold cross validation is showing below.

Table 10. Summary of Fittest Model for Data 'Glass'

Parameters:
SVM-Type: C-classification
SVM-Kernel: radial
cost: 10
gamma: 0.43

Class	1	2	3	5	6	7	Total
Number of support vector	51	53	16	11	7	17	155

The results shows that the highest accuracy occurs when cost equal to 10 and gamma is 0.43. We can realized the number of support vector does not decrease, the reason we mentioned before since the narrower soft margin could cause the overfitting situation.

Also, the summary of confusion matrix is

Confusion matrix:

		true					
		1	2	3	5	6	7
	1	6	0	0	0	0	0
	2	1	6	0	1	1	2
predict	3	0	0	0	0	0	0
	5	0	0	0	1	0	0
	6	0	0	0	0	1	0
	7	0	0	0	0	0	2

Table 11. Statistics of Fittest Model for Data 'Glass'

Overall Statistics:

Accuracy : 0.7619

95% CI : (0.5283, 0.9178)

No Information Rate : 0.3333

P-Value [Acc > NIR] : 7.251e-05

Statistics by Class:

	Class 1	Class 2	Class 3	Class 5	Class 6	Class 7
Sensitivity	0.8571	1.0000	N/A	0.5000	0.5000	0.5000
Specificity	1.0000	0.6667	1.0000	1.0000	1.0000	1.0000
Balanced Accuracy	0.9286	0.8333	N/A	0.7500	0.7500	0.7500

The highest accuracy calculated by 10-fold cross validation is 0.7619, which is slightly higher than our base case. In fact, in this example, the combination of the original gamma and cost is not the worst situation compare to other cases.

In general, in machine learning cases, we would expect our confusion matrix for prediction has at least 70%-80% accuracy. From this point of view, we cannot conclude that the result of our experiment is inaccurate, however, it is not ideal, technically speaking.

There is two possible reason which cause this situation. The first is the complexity of our observation. The hyperplane could become harder to set along with the increase of the number of parameter. In this example, we have 9 parameters and we can realize that in some situation, even though it shows the accuracy increases in the overall statistics, for a specific class, the balanced accuracy for prediction could possibly decrease.

The second reason is the choice of test dataset. In our experiment, we separated the original dataset into several subset and chose one as a test dataset and loop approach. However, we do not have a 'unknown' dataset for prediction, which means this part of dataset is only treated as test dataset, not the training one. The design of our experiment has higher possibility to cause overfitting, which can be an important factor which lower the prediction accuracy.

Chapter 3 SUPPORT VECTOR REGRESSION

This chapter will discuss support vector machine applied to both linear and non-linear regression problem. Our goal is to is to find a function f(x) that has at most ε deviation from actually obtained target yi, for all the training data, and at the same time is

as flat as possible. In other words, the data points lie in between the two borders of the margin which is maximized under suitable conditions will avoid outlier inclusion.

3.1 Linear regression

Suppose we are given a set of data

$$D = \{(x_1, y_1), \ldots, (x_l, y_l)\}, x \in R^d, y \in R \qquad (3.1)$$

with linear function in form of

$$f(x) = \ < w, x > \ + b \qquad (3.2)$$

The optimal regression function is given by the minimum of the functional,

$$\phi(\omega, \zeta) = \frac{1}{2} \| \omega \|^2 + C \sum_{i=1}^{l} (\xi_i^- + \xi_i^+) \qquad (3.3)$$

where $C > 0$ is a pre-specified constant value, determined the trade-off between the flatness of f and the amount up to which deviations larger than ε are tolerated, and ξ^-, ξ^+ are slack variables representing upper and lower constraints on the outputs of the system.

3.1.1 ε-insensitive Loss Function

The ε-insensitive loss function is the following

$$L(y, f(x)) = |y - f(x)| \qquad (3.4)$$

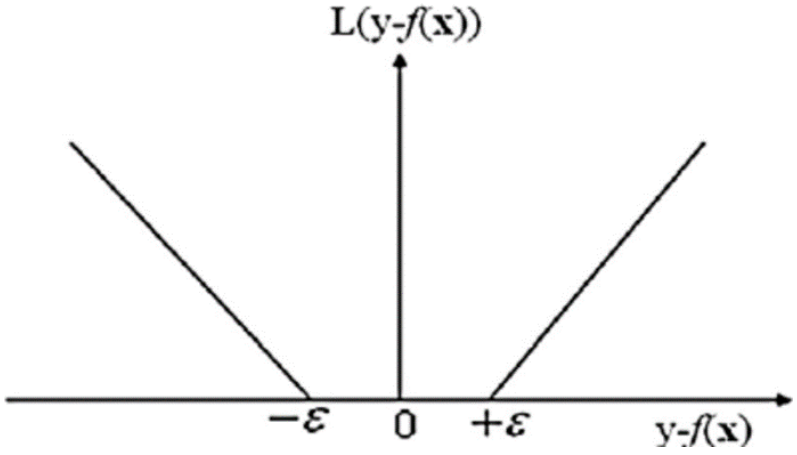

Figure 3.1 A plot of typical ε-insensitive loss function

This loss function is ideal when small amounts of error are acceptable, which is also refer to a 'soft-margin'. In ε-insensitive loss function, any points within some selected range ε are considered to have no error at all, which means the ε-insensitive loss function can be represented as

$$L_{\varepsilon}(y) = \begin{cases} 0 & for\ |f(x) - y| < \varepsilon \\ |f(x) - y| - \varepsilon & otherwise \end{cases} \quad (3.5)$$

This error-free margin makes the loss function an ideal candidate for support vector regression.

In linear function $f(x) = \ < w, x > \ + b,\ < \omega, x >$ denotes the dot product in R^d. Since we mentioned before, one of our goal is to ensure the flatness, which means seeking a small ω. One of our way is to minimize the norm, i.e. $\| \omega \|^2 = \ < \omega,\ \omega >$.

$$\text{minimize } \frac{1}{2} \| \omega \|^2 \quad (3.6)$$

$$\text{subject to } \begin{cases} y_i - < \omega, x_i > - b \leq \varepsilon \\ < \omega, x_i > + b - y_i \leq \varepsilon \end{cases} \quad (3.7)$$

Applying ε-insensitive loss function for minimizing the optimal regression function,

$$\text{minimize } \frac{1}{2} \| \omega \|^2 + C \sum_{i=1}^{l} (\xi_i^- + \xi_i^+) \quad (3.8)$$

$$subject\ to \quad \begin{cases} y_i - \ < \omega, x_i > - b \ \leq \ \varepsilon + \xi_i^- \\ < \omega, x_i > + b - y_i \ \leq \ \varepsilon + \xi_i^+ \\ \xi_i^-, \xi_i^+ \quad \geq 0 \end{cases} \quad (3.9)$$

Here, the key idea is to construct a Lagrange function (Smola & Scholkopf, 2003). We proceed as follows:

$$L\left(\eta_i,\eta^*_i,\alpha_i,\alpha^*_i\right)=\frac{1}{2}\parallel\omega\parallel^2+C\sum_{i=1}^{l}(\xi^-_i+\xi^+_i)-\sum_{i=1}^{l}\left(\eta_i\xi^-_i+\eta^*_i\xi^+_i\right)$$
$$-\sum_{i=1}^{l}\alpha_i(\varepsilon+\xi_i-y_i+<\omega,x_i>+b$$
$$-\sum_{i=1}^{l}\alpha^*_i(\varepsilon+\xi^*_i-y_i$$
$$+<\omega,x_i>+b \qquad (3.10)$$

Here L is the Lagrangian and $\eta_i,\eta^*_i,\alpha_i,\alpha^*_i$ are Lagrange multipliers. Taking partial derivative of 3.10 with primal variables $\left(\omega,\ b,\xi^-_i,\xi^+_i\right)$:

$$\partial_b L=\sum_{i=1}^{l}\left(\alpha^*_i-\alpha_i\right)=0 \qquad (3.11)$$

$$\partial_\omega L=\omega-\sum_{i=1}^{l}\left(\alpha_i-\alpha^*_i\right)x_i=0 \qquad (3.12)$$

$$\partial_{\xi^+_i}L=C-\alpha^*_i-\eta^*_i=0 \qquad (3.13)$$

Substituting the 3.11-3.13 into L, the solution is given by,

$$\max\ -\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\left(\alpha_i-\alpha^*_i\right)\left(\alpha_j-\alpha^*_j\right)<x_i,x_j>+\sum_{i=1}^{l}\alpha_i(y_i-\varepsilon)-\alpha^*_i\left(y_i+\varepsilon\right) \quad (3.14)$$

or alternatively,

$$min\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\left(\alpha_i-\alpha^*_i\right)\left(\alpha_j-\alpha^*_j\right)<x_i,x_j>-\sum_{i=1}^{l}\left(\alpha_i-\alpha^*_i\right)y_i+\sum_{i=1}^{l}\left(\alpha_i+\alpha^*_i\right)\varepsilon \quad (3.15)$$

with constraints,

$$0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, \ldots, l \qquad (3.16)$$

$$\sum_{i=1}^{l} (\alpha_i - \alpha_i^*) = 0 \qquad (3.17)$$

Equation 3.12 can be rewritten as

$$\omega = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) x_i \qquad (3.18)$$

Thus

$$f(x) = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) < x_i, x > + b \qquad (3.19)$$

This is what we called Support Vector expansion, where ω can be described as a linear combination of x_i.

The Karush-Kuhn-Trcker conditions that are satisfied by the solution are,

$$\alpha_i \alpha_i^* = 0, \quad i = 1, \ldots, l \qquad (3.20)$$

which is also called complementary slackness (Smola & Scholkopf, 2003). This condition shows that there can never be a set of dual variables α_i, α_i^* which are both simultaneously nonzero, which allows us to conclude that

$$\varepsilon - y_i + < \omega, x_i > + b \geq 0 \quad and \quad \xi_i = 0 \quad if \quad \alpha_i < C$$
$$\varepsilon - y_i + < \omega, x_i > + b \leq 0 \quad if \quad \alpha_i < 0 \qquad (3.21)$$

Therefore, the support vectors are points where exactly one of the Lagrange multipliers is greater than zero. When ε = 0, we get L loss function and the optimization problem simplified,

$$min \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \beta_i \beta_j < x_i, x_j > - \sum_{i=1}^{l} \beta_i y_i \qquad (3.22)$$

with constraints,

$$-C \leq \beta_i \leq C, i = 1,\ldots, l \qquad (3.23)$$

$$\sum_{i=1}^{l} \beta_i = 0$$

and the regression function is given by Equation 3.2, where

$$\omega = \sum_{i=1}^{l} \beta_i x_i \qquad (3.24)$$

$$b = -\frac{1}{2} < \omega, (x_r + x_s) > \qquad (3.25)$$

where x_r and x_s are support vectors lie on the edge of the margin.

In addition, if $\varepsilon = 0$, the result is just the median regression.

3.1.2 Quadratic Loss Function

The quadratic function is the following

$$L(f(x) - y) = (f(x) - y)^2 \qquad (3.26)$$

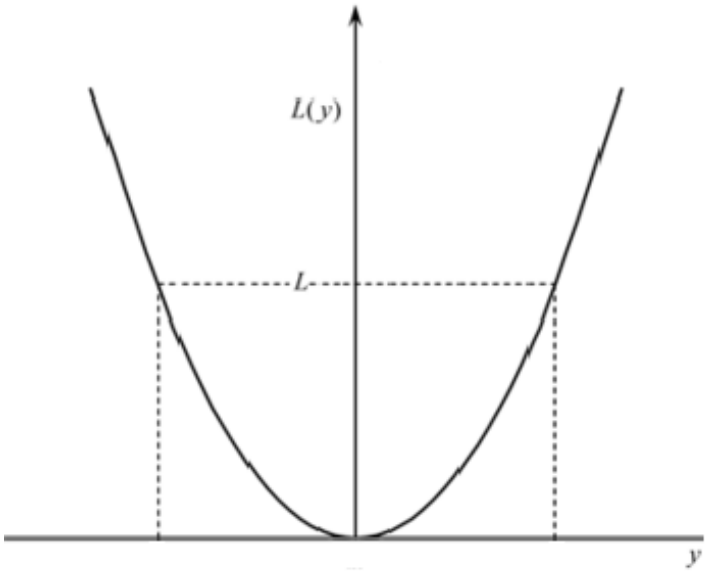

Figure 3.2 A plot of typical quadratic loss function

Quadratic loss is another function that is well-suited for the purpose of regression problems, although it makes the outliers in the data punished very heavily by the squaring of the error. As a result, datasets must be filtered for outliers first, or else the fit from this loss function may not be desirable.

Same as what we did in 3.1.1, by applying quadratic loss function, the solution is given by,

$$\max\ -\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\left(\alpha_i-\alpha_i^*\right)\left(\alpha_j-\alpha_j^*\right)<x_i,x_j>+\sum_{i=1}^{l}y_i\left(\alpha_i-\alpha_i^*\right)-\frac{1}{2C}\sum_{i=1}^{l}\left(\alpha_i{}^2-\left(\alpha_i^*\right)^2\right) \tag{3.27}$$

The corresponding optimisation can be simplified by exploiting the KKT conditions, which implies $\beta_i^* = |\beta_i|$. The resultant optimization problems is,

$$min\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\beta_i\beta_j<x_i,x_j>-\sum_{i=1}^{l}\beta_i y_i+\frac{1}{2C}\sum_{i=1}^{l}\beta_i^2 \tag{3.28}$$

with constraints

$$\sum_{i=1}^{l}\beta_i=0 \tag{3.29}$$

and the regression function is given by Equation 3.2, where

$$\omega=\sum_{i=1}^{l}\beta_i x_i \tag{3.30}$$

$$b=-\frac{1}{2}<\omega,\left(x_r+x_s\right)> \tag{3.31}$$

3.1.3 Huber Loss Function

The Huber loss function is the following,

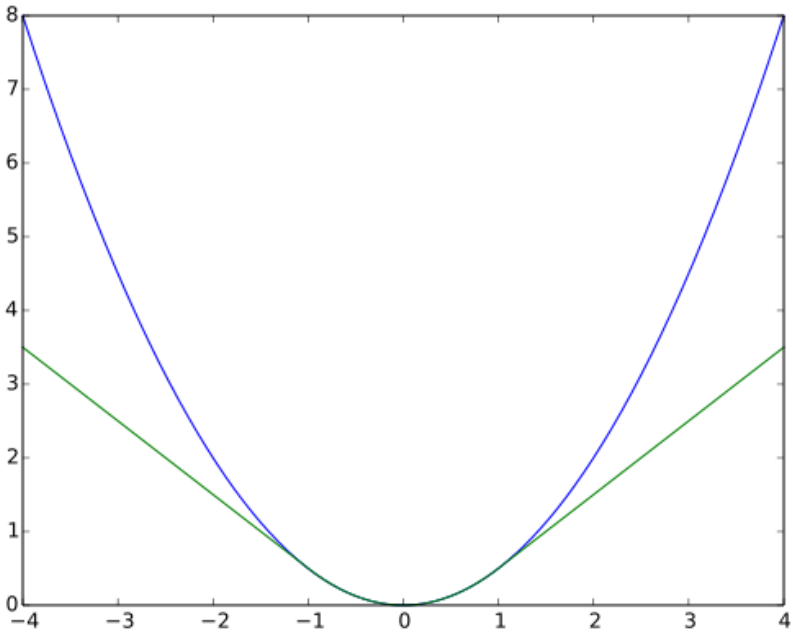

Figure 3.3 Huber loss and quadratic loss as a function of $y - f(x)$

$$L_{huber}(f(x)-y)\begin{cases}\frac{1}{2}(f(x)-y)^2 & for \quad |f(x)-y|<\mu \\ \mu|f(x)-y|-\frac{\mu^2}{2} & otherwise\end{cases} \tag{3.32}$$

Huber proposed the loss function as a robust loss function that has optimal properties when the distribution of the data is unknown (Yeh, Huang & Lee, 2011). Compared to quadratic loss function, Huber loss is less sensitive to outliers in data. As defined above, the Huber loss function is convex in a uniform neighborhood of its minimum $f(x) - y = 0$, at the boundary of this uniform neighborhood, the Huber loss function has a differentiable extension to an affine function at points $f(x) - y = -\mu$ and $f(x) - y = \mu$. These properties allow it to combine much of the sensitivity of the mean-unbiased.

By applying Huber loss function, the solution is given by,

$$\text{maximize} \quad -\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\left(\alpha_i - \alpha_i^*\right)\left(\alpha_j - \alpha_j^*\right) < x_i, x_j > + \sum_{i=1}^{l} y_i\left(\alpha_i - \alpha_i^*\right) - \frac{1}{2C}\sum_{i=1}^{l}\mu\left(\alpha_i^2 + \left(\alpha_i^*\right)^2\right) \quad (3.23)$$

The resultant optimization problem is,

$$min\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\beta_i\beta_j < x_i, x_j > - \sum_{i=1}^{l}\beta_i y_i + \frac{1}{2C}\sum_{i=1}^{l}\beta_i^2\mu \quad (3.24)$$

with constraints

$$-C \le \beta_i \le C, i = 1,\ldots, l \quad (3.25)$$

$$\sum_{i=1}^{l}\beta_i = 0 \quad (3.26)$$

and the regression function is given by Equation 3.2, where

$$\omega = \sum_{i=1}^{l}\beta_i x_i \quad (3.27)$$

$$b = -\frac{1}{2} < \omega, \left(x_r + x_s\right) > \quad (3.28)$$

3.2 Nonlinear Regression

In the situation of nonlinear support vector regression approach, it is always achieved by mapping x_i into a high dimensional feature space, i.e. there exist a map Φ: $\chi \rightarrow F$, which F is a space the standard SV linear regression performs. The most widely adopted method is by using kernel approach.

Same as linear SVR, we adopting ε-insensitive loss function, and the objective function and constraints are showing as

minimize $\frac{1}{2}\| \omega \|^2 + C\sum_{i=1}^{l}(\xi_i^- + \xi_i^+)$ (3.29)

$$subject\ to \quad \begin{cases} y_i - < \omega, \Phi(x_i) > - b \leq \varepsilon + \xi_i^- \\ < \omega, \Phi(x_i) > + b - y_i \leq \varepsilon + \xi_i^+ \\ \xi_i^-, \xi_i^+ \geq 0 \end{cases} \quad (3.30)$$

where the regression hyperplane is derived as

$$f(x) = < w, \Phi(x) > + b \quad (3.31)$$

To solve 3. , one can also introduce the Lagrange function as the linear SVR situation and taking partial derivatives with respect to the primal variables and set the resulting derivatives to zero. The solution is given by

$$\max -\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K < x_i, x_j > + \sum_{i=1}^{l} y_i(\alpha_i^* - \alpha_i) - \sum_{i=1}^{l}\varepsilon(\alpha_i + \alpha_i^*) \quad (3.33)$$

with constraints,

$$0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, \ldots, l \quad (3.34)$$

$$\sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0 \quad (3.35)$$

Here, $K < x_i, x_j >$ is the kernel function which represents the inner product $< \Phi(x_i), \Phi(x_j) >$.

The most widely used kernel function is the radial basis function, which is also called Gaussian kernel. The function is defined as

$$K < x_i, x_j \geq = < \Phi(x_i), \Phi(x_j) > = exp(- \gamma||x_i - x_j||^2) \quad (3.36)$$

where $||x_i - x_j||^2$ is recognized as the squared Euclidean distance between the two feature vectors, and γ is the width parameter of Gaussian kernel. Solving 3.33 with the constraints equation, the regression function is given by,

$$f(x) = \sum_{i=1}^{l} \left(\alpha_i - \alpha_i^*\right) K < x_i, x > + b \qquad (3.37)$$

where

$$< \omega, x > = \sum_{i=1}^{l} \left(\alpha_i - \alpha_i^*\right) K < x_i, x_j > \qquad (3.39)$$

$$b = -\frac{1}{2} \sum_{i=1}^{l} \left(\alpha_i - \alpha_i^*\right) (K < x_i, x_r > + K < x_i, x_s >) \qquad (3.40)$$

As with the SVC the equality constraint may be dropped if the Kernel contains a bias term, b being accommodated within the Kernel function, and the regression function is given by,

$$f(x) = \sum_{i=1}^{l} \left(\alpha_i - \alpha_i^*\right) K < x_i, x > \qquad (3.41)$$

2.5 Application of Support Vector Regression to Real World Data

The data 'Ozone' used in this chapter, taken from Department of Statistics, UC Berkeley, containing observation of Los Angeles daily ozone pollution in 1976, and other factors that related to ozone pullution. It is a perfect data set for support vector regression and our task is to predict the daily maximum one-hour-average ozone reading based on the dataset.

Table 12. Los Angeles daily ozone pollution in 1976 (Partial)

Observation	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13
1	1	1	4	3	5480	8	20	*NA*	*NA*	5000	-15	30.56	200

2	1	2	5	3	5660	6	*NA*	38	*NA*	*NA*	-14	*NA*	300
3	1	3	6	3	5710	4	28	40	*NA*	2693	-25	47.66	250
4	1	4	7	5	5700	3	37	45	*NA*	590	-24	55.04	100
5	1	5	1	5	5760	3	51	54	45.32	1450	25	57.02	60
6	1	6	2	6	5720	4	69	35	49.64	1568	15	53.78	60
7	1	7	3	4	5790	6	19	45	46.4	2631	-33	54.14	100
8	1	8	4	4	5790	3	25	55	52.7	554	-28	64.76	250
9	1	9	5	6	5700	3	73	41	48.02	2083	23	52.52	120
10	1	10	6	7	5700	3	59	44	*NA*	2654	-2	48.38	120
11	1	11	7	4	5770	8	27	54	*NA*	5000	-19	48.56	120
12	1	12	1	6	5720	3	44	51	54.32	111	9	63.14	150
13	1	13	2	5	5760	6	33	51	57.56	492	-44	64.58	40
14	1	14	3	4	5780	6	19	54	56.12	5000	-44	56.3	200
15	1	15	4	4	5830	3	19	58	62.24	1249	-53	75.74	250
16	1	16	5	7	5870	2	19	61	64.94	5000	-67	65.48	200
17	1	17	6	5	5840	5	19	64	*NA*	5000	-40	63.32	200
18	1	18	7	9	5780	4	59	67	*NA*	639	1	66.02	150
19	1	19	1	4	5680	5	73	52	56.48	393	-68	69.8	10
20	1	20	2	3	5720	4	19	54	*NA*	5000	-66	54.68	140
21	1	21	3	4	5760	3	19	54	53.6	5000	-58	51.98	250

Format:

V1. Month

V2. Day of month

V3. Day of week

V4. Daily maximum one-hour-average ozone reading

V5. 500 millibar pressure height (m) measured at Vandenberg AFB

V6. Wind speed (mph) at Los Angeles International Airport (LAX)

V7. Humidity (%) at LAX

V8. Temperature (degrees F) measured at Sandburg, CA

V9. Temperature (degrees F) measured at El Monte, CA

V10. Inversion base height (feet) at LAX

V11. Pressure gradient (mm Hg) from LAX to Daggett, CA

V12. Inversion base temperature (degrees F) at LAX

V13. Visibility (miles) measured at LAX

We are also going to take R package 'e1071' as what we did in support vector classification to help us construct regression models and predict. We are going to separate our data into training and testing data, however, since we our task is to build a regression model, we need to omit the missing data in the dataset, which is the observation with parameters show not available in the table.

After omitting the missing data, we separate the observation into 134 observations as training data and 69 observations as testing data, since we are going to construct a one-third test, where we set one-third of the dataset as testing data and another part as training data. We are going to construct a nonlinear regression model with ε-insensitive Loss Function. Also, since it is nonlinear regression, we need a kernel to help us construct the model and here we will go with Gaussian kernel. Starting with setting cost as 1000 and gamma in radial kernel as 0.0001 and using code 'svm' in package 'e1071', we got the summary of our model showing below:

Table 13. Summary of Base Model for Data 'Ozone'

Parameters:
SVM-Type: eps-regression
SVM-Kernel: radial
cost: 1000
gamma: 1e-04
Number of Support Vectors: 118

Same as support vector classification, the number of support vector is a measurement of the performance of a certain dataset fits this regression model. The lower the number is, the better it fits in this model.

The next step is to predict the test value, and the way for us to quantify the performance of fitness of our prediction is to calculate pseudo r square.

$$R^2 = 1 - \frac{ln\hat{L}(M_{full})}{ln\hat{L}(M_{intercept})} \qquad (3.43)$$

The equation showing above is called Mcfadden's pseudo r square, which we are going to use in our prediction, where $ln\hat{L}(M_{full})$ represent log likelihood for model with predictors, and $ln\hat{L}(M_{intercept})$ represent log likelihood for model withour predictors (Yeh, Huang & Lee, 2011). The log likelihood of the intercept model is treated as a total sum of squares, and the log likelihood of the full model is treated as the sum of squared errors. The ratio of the likelihoods suggests the level of improvement over the intercept model offered by the full model. A likelihood falls between 0 and 1, and a smaller ratio of log likelihoods indicates that the full model is a far better fit than the intercept model. Thus, if comparing two models on the same data, McFadden's would be higher for the model with the greater likelihood.

By computing likelihood for both full and intercept models from package 'e1071', we obtained the Mcfadden's pseudo r square equals to be 0.6793. Since the r square is describe to be as large as possible, the next step we are going to do is to change the parameters in our model, which is similar to what we did in Chapter 2. The only 2

parameters which decide r square is cost and gamma in the radial kernel. For example, if we change cost from 1000 to 500, the summary of our model will become

Table 14. Summary of Adjusted Model 1 for Data 'Ozone'

Parameters:
SVM-Type: eps-regression
SVM-Kernel: radial
cost: 500
gamma: 1e-04
Number of Support Vectors: 116

We can see the number of support vectors decrease, which means the new regression model fits our dataset slightly better. Also, by computing the r square, we got the value to be 0.7264, which also confirm our conclusion from summary.

Also, we can change gamma, for example, from 0.0001 to 0.001, and the summary is shown below as

Table 15. Summary of Adjusted Model 2 for Data 'Ozone'

Parameters:
SVM-Type: eps-regression
SVM-Kernel: radial
cost: 1000
gamma: 0.001
Number of Support Vectors: 108

And the r square equals to 0.6789, which is slightly lower than the original model. Since the number of support vector do decrease, we can consider it as an overfitting situation.

Our goal is find a model, which can fit our dataset as accurate as possible, so the approach we are going to take is still cross-validation as we did in Chapter 2. Here, we are going to take a 5-fold cross-validation, where we write a 'for loop' to achieve it in R. We could set a sequence which allows the cost runs from 1000 to 100 for 90 times, and set gamma from 0.0001 from 0.01 for 100 times. The base case is showing in our original model where cost is 1000 and gamma is 0.0001.

The result after the cross-validation shows below:

Table 16. Summary of Fittest Model for Data 'Ozone'

Parameters:
SVM-Type: eps-regression
SVM-Kernel: radial
cost: 100
gamma: 0.0011
Number of Support Vectors: 102

Here, the r square is equals to 0.7853. From the result, we can realize the number of support vector decreases significantly and meanwhile the r square increase. Since the cost in this model is the lowest model in our sequence and the number of support vector do decrease when we change cost from 1000 to 500, we may assume that the model fits better with the decrease of cost.

In fact, we can plot the r square for this cross-validation with parameter cost and gamma, as seen in Figure 3.4.

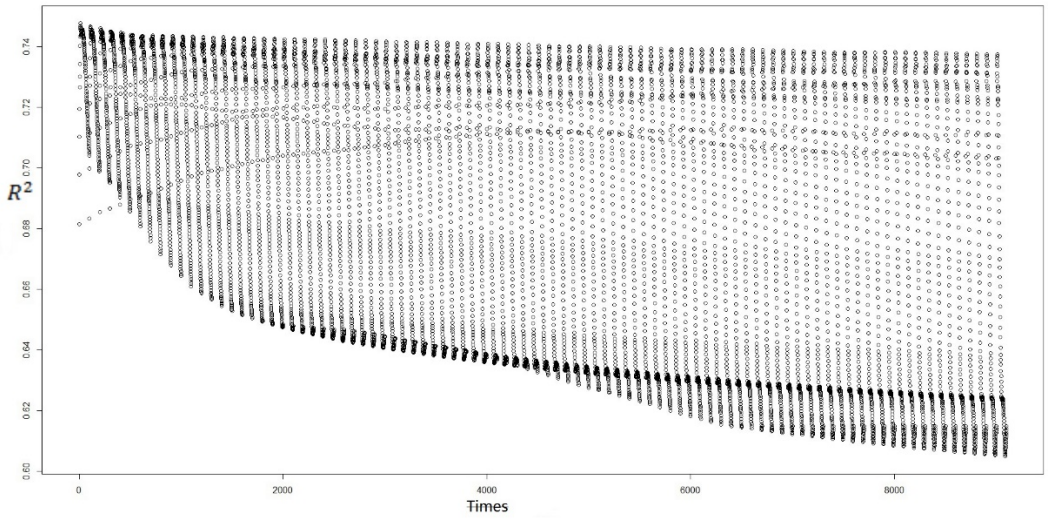

Figure 3.4 R square plot for 5-fold cross validation

We can see from this plot that the highest R square appears when cost and gamma are both relatively small. Since we mentioned before the model could be too sensitive for the dataset when we set the cost undersize and the overfitting situation may also appears, we can consider that this regression model with cost 100 and gamma 1 an relatively ideal result.

REFERENCE

James, G., Witten, D., Hastie, T. & Tibshirani, R. (2009). *An Introduction to Statistical Learning* (2nd ed.). New York: Springer.

Smola, A.J. & Scholkopf, B. (2003). A Tutorial on Support Vector Regression. *Statistics and Computing,* 14, 199–222.

Yeh, C. Y., Huang, C. W. & Lee, S. J. (2011). A Multiple-kernel support vector regression approach for stock market price forecasting. *Expert System with Application,* 38, 2177-2186.

APPENDICES

Appendix A. Glass Data

Observation	RI	Na	Mg	Al	Si	K	Ca	Ba	Fe	Type
1	1.52101	13.64	4.49	1.1	71.78	0.06	8.75	0	0	1
2	1.51761	13.89	3.6	1.36	72.73	0.48	7.83	0	0	1
3	1.51618	13.53	3.55	1.54	72.99	0.39	7.78	0	0	1
4	1.51766	13.21	3.69	1.29	72.61	0.57	8.22	0	0	1
5	1.51742	13.27	3.62	1.24	73.08	0.55	8.07	0	0	1

6	1.51596	12.79	3.61	1.62	72.97	0.64	8.07	0	0.26	1
7	1.51743	13.3	3.6	1.14	73.09	0.58	8.17	0	0	1
8	1.51756	13.15	3.61	1.05	73.24	0.57	8.24	0	0	1
9	1.51918	14.04	3.58	1.37	72.08	0.56	8.3	0	0	1
10	1.51755	13	3.6	1.36	72.99	0.57	8.4	0	0.11	1
11	1.51571	12.72	3.46	1.56	73.2	0.67	8.09	0	0.24	1
12	1.51763	12.8	3.66	1.27	73.01	0.6	8.56	0	0	1
13	1.51589	12.88	3.43	1.4	73.28	0.69	8.05	0	0.24	1
14	1.51748	12.86	3.56	1.27	73.21	0.54	8.38	0	0.17	1
15	1.51763	12.61	3.59	1.31	73.29	0.58	8.5	0	0	1
16	1.51761	12.81	3.54	1.23	73.24	0.58	8.39	0	0	1
17	1.51784	12.68	3.67	1.16	73.11	0.61	8.7	0	0	1
18	1.52196	14.36	3.85	0.89	71.36	0.15	9.15	0	0	1
19	1.51911	13.9	3.73	1.18	72.12	0.06	8.89	0	0	1
20	1.51735	13.02	3.54	1.69	72.73	0.54	8.44	0	0.07	1
21	1.5175	12.82	3.55	1.49	72.75	0.54	8.52	0	0.19	1
22	1.51966	14.77	3.75	0.29	72.02	0.03	9	0	0	1
23	1.51736	12.78	3.62	1.29	72.79	0.59	8.7	0	0	1
24	1.51751	12.81	3.57	1.35	73.02	0.62	8.59	0	0	1
25	1.5172	13.38	3.5	1.15	72.85	0.5	8.43	0	0	1
26	1.51764	12.98	3.54	1.21	73	0.65	8.53	0	0	1
27	1.51793	13.21	3.48	1.41	72.64	0.59	8.43	0	0	1
28	1.51721	12.87	3.48	1.33	73.04	0.56	8.43	0	0	1
29	1.51768	12.56	3.52	1.43	73.15	0.57	8.54	0	0	1
30	1.51784	13.08	3.49	1.28	72.86	0.6	8.49	0	0	1
31	1.51768	12.65	3.56	1.3	73.08	0.61	8.69	0	0.14	1
32	1.51747	12.84	3.5	1.14	73.27	0.56	8.55	0	0	1
33	1.51775	12.85	3.48	1.23	72.97	0.61	8.56	0.09	0.22	1
34	1.51753	12.57	3.47	1.38	73.39	0.6	8.55	0	0.06	1
35	1.51783	12.69	3.54	1.34	72.95	0.57	8.75	0	0	1
36	1.51567	13.29	3.45	1.21	72.74	0.56	8.57	0	0	1

Observation	RI	Na	Mg	Al	Si	K	Ca	Ba	Fe	Type
37	1.51909	13.89	3.53	1.32	71.81	0.51	8.78	0.11	0	1
38	1.51797	12.74	3.48	1.35	72.96	0.64	8.68	0	0	1
39	1.52213	14.21	3.82	0.47	71.77	0.11	9.57	0	0	1
40	1.52213	14.21	3.82	0.47	71.77	0.11	9.57	0	0	1
41	1.51793	12.79	3.5	1.12	73.03	0.64	8.77	0	0	1
42	1.51755	12.71	3.42	1.2	73.2	0.59	8.64	0	0	1
43	1.51779	13.21	3.39	1.33	72.76	0.59	8.59	0	0	1
44	1.5221	13.73	3.84	0.72	71.76	0.17	9.74	0	0	1
45	1.51786	12.73	3.43	1.19	72.95	0.62	8.76	0	0.3	1

46	1.519	13.49	3.48	1.35	71.95	0.55	9	0	0	1
47	1.51869	13.19	3.37	1.18	72.72	0.57	8.83	0	0.16	1
48	1.52667	13.99	3.7	0.71	71.57	0.02	9.82	0	0.1	1
49	1.52223	13.21	3.77	0.79	71.99	0.13	10.02	0	0	1
50	1.51898	13.58	3.35	1.23	72.08	0.59	8.91	0	0	1
51	1.5232	13.72	3.72	0.51	71.75	0.09	10.06	0	0.16	1
52	1.51926	13.2	3.33	1.28	72.36	0.6	9.14	0	0.11	1
53	1.51808	13.43	2.87	1.19	72.84	0.55	9.03	0	0	1
54	1.51837	13.14	2.84	1.28	72.85	0.55	9.07	0	0	1
55	1.51778	13.21	2.81	1.29	72.98	0.51	9.02	0	0.09	1
56	1.51769	12.45	2.71	1.29	73.7	0.56	9.06	0	0.24	1
57	1.51215	12.99	3.47	1.12	72.98	0.62	8.35	0	0.31	1
58	1.51824	12.87	3.48	1.29	72.95	0.6	8.43	0	0	1
59	1.51754	13.48	3.74	1.17	72.99	0.59	8.03	0	0	1
60	1.51754	13.39	3.66	1.19	72.79	0.57	8.27	0	0.11	1
61	1.51905	13.6	3.62	1.11	72.64	0.14	8.76	0	0	1
62	1.51977	13.81	3.58	1.32	71.72	0.12	8.67	0.69	0	1
63	1.52172	13.51	3.86	0.88	71.79	0.23	9.54	0	0.11	1
64	1.52227	14.17	3.81	0.78	71.35	0	9.69	0	0	1
65	1.52172	13.48	3.74	0.9	72.01	0.18	9.61	0	0.07	1
66	1.52099	13.69	3.59	1.12	71.96	0.09	9.4	0	0	1
67	1.52152	13.05	3.65	0.87	72.22	0.19	9.85	0	0.17	1
68	1.52152	13.05	3.65	0.87	72.32	0.19	9.85	0	0.17	1
69	1.52152	13.12	3.58	0.9	72.2	0.23	9.82	0	0.16	1
70	1.523	13.31	3.58	0.82	71.99	0.12	10.17	0	0.03	1
71	1.51574	14.86	3.67	1.74	71.87	0.16	7.36	0	0.12	2
72	1.51848	13.64	3.87	1.27	71.96	0.54	8.32	0	0.32	2
73	1.51593	13.09	3.59	1.52	73.1	0.67	7.83	0	0	2
74	1.51631	13.34	3.57	1.57	72.87	0.61	7.89	0	0	2
75	1.51596	13.02	3.56	1.54	73.11	0.72	7.9	0	0	2
76	1.5159	13.02	3.58	1.51	73.12	0.69	7.96	0	0	2
Observation	RI	Na	Mg	Al	Si	K	Ca	Ba	Fe	Type
77	1.51645	13.44	3.61	1.54	72.39	0.66	8.03	0	0	2
78	1.51627	13	3.58	1.54	72.83	0.61	8.04	0	0	2
79	1.51613	13.92	3.52	1.25	72.88	0.37	7.94	0	0.14	2
80	1.5159	12.82	3.52	1.9	72.86	0.69	7.97	0	0	2
81	1.51592	12.86	3.52	2.12	72.66	0.69	7.97	0	0	2
82	1.51593	13.25	3.45	1.43	73.17	0.61	7.86	0	0	2
83	1.51646	13.41	3.55	1.25	72.81	0.68	8.1	0	0	2
84	1.51594	13.09	3.52	1.55	72.87	0.68	8.05	0	0.09	2
85	1.51409	14.25	3.09	2.08	72.28	1.1	7.08	0	0	2

86	1.51625	13.36	3.58	1.49	72.72	0.45	8.21	0	0	2
87	1.51569	13.24	3.49	1.47	73.25	0.38	8.03	0	0	2
88	1.51645	13.4	3.49	1.52	72.65	0.67	8.08	0	0.1	2
89	1.51618	13.01	3.5	1.48	72.89	0.6	8.12	0	0	2
90	1.5164	12.55	3.48	1.87	73.23	0.63	8.08	0	0.09	2
91	1.51841	12.93	3.74	1.11	72.28	0.64	8.96	0	0.22	2
92	1.51605	12.9	3.44	1.45	73.06	0.44	8.27	0	0	2
93	1.51588	13.12	3.41	1.58	73.26	0.07	8.39	0	0.19	2
94	1.5159	13.24	3.34	1.47	73.1	0.39	8.22	0	0	2
95	1.51629	12.71	3.33	1.49	73.28	0.67	8.24	0	0	2
96	1.5186	13.36	3.43	1.43	72.26	0.51	8.6	0	0	2
97	1.51841	13.02	3.62	1.06	72.34	0.64	9.13	0	0.15	2
98	1.51743	12.2	3.25	1.16	73.55	0.62	8.9	0	0.24	2
99	1.51689	12.67	2.88	1.71	73.21	0.73	8.54	0	0	2
100	1.51811	12.96	2.96	1.43	72.92	0.6	8.79	0.14	0	2
101	1.51655	12.75	2.85	1.44	73.27	0.57	8.79	0.11	0.22	2
102	1.5173	12.35	2.72	1.63	72.87	0.7	9.23	0	0	2
103	1.5182	12.62	2.76	0.83	73.81	0.35	9.42	0	0.2	2
104	1.52725	13.8	3.15	0.66	70.57	0.08	11.64	0	0	2
105	1.5241	13.83	2.9	1.17	71.15	0.08	10.79	0	0	2
106	1.52475	11.45	0	1.88	72.19	0.81	13.24	0	0.34	2
107	1.53125	10.73	0	2.1	69.81	0.58	13.3	3.15	0.28	2
108	1.53393	12.3	0	1	70.16	0.12	16.19	0	0.24	2
109	1.52222	14.43	0	1	72.67	0.1	11.52	0	0.08	2
110	1.51818	13.72	0	0.56	74.45	0	10.99	0	0	2
111	1.52664	11.23	0	0.77	73.21	0	14.68	0	0	2
112	1.52739	11.02	0	0.75	73.08	0	14.96	0	0	2
113	1.52777	12.64	0	0.67	72.02	0.06	14.4	0	0	2
114	1.51892	13.46	3.83	1.26	72.55	0.57	8.21	0	0.14	2
115	1.51847	13.1	3.97	1.19	72.44	0.6	8.43	0	0	2
116	1.51846	13.41	3.89	1.33	72.38	0.51	8.28	0	0	2

Observation	RI	Na	Mg	Al	Si	K	Ca	Ba	Fe	Type
117	1.51829	13.24	3.9	1.41	72.33	0.55	8.31	0	0.1	2
118	1.51708	13.72	3.68	1.81	72.06	0.64	7.88	0	0	2
119	1.51673	13.3	3.64	1.53	72.53	0.65	8.03	0	0.29	2
120	1.51652	13.56	3.57	1.47	72.45	0.64	7.96	0	0	2
121	1.51844	13.25	3.76	1.32	72.4	0.58	8.42	0	0	2
122	1.51663	12.93	3.54	1.62	72.96	0.64	8.03	0	0.21	2
123	1.51687	13.23	3.54	1.48	72.84	0.56	8.1	0	0	2
124	1.51707	13.48	3.48	1.71	72.52	0.62	7.99	0	0	2
125	1.52177	13.2	3.68	1.15	72.75	0.54	8.52	0	0	2

126	1.51872	12.93	3.66	1.56	72.51	0.58	8.55	0	0.12	2
127	1.51667	12.94	3.61	1.26	72.75	0.56	8.6	0	0	2
128	1.52081	13.78	2.28	1.43	71.99	0.49	9.85	0	0.17	2
129	1.52068	13.55	2.09	1.67	72.18	0.53	9.57	0.27	0.17	2
130	1.5202	13.98	1.35	1.63	71.76	0.39	10.56	0	0.18	2
131	1.52177	13.75	1.01	1.36	72.19	0.33	11.14	0	0	2
132	1.52614	13.7	0	1.36	71.24	0.19	13.44	0	0.1	2
133	1.51813	13.43	3.98	1.18	72.49	0.58	8.15	0	0	2
134	1.518	13.71	3.93	1.54	71.81	0.54	8.21	0	0.15	2
135	1.51811	13.33	3.85	1.25	72.78	0.52	8.12	0	0	2
136	1.51789	13.19	3.9	1.3	72.33	0.55	8.44	0	0.28	2
137	1.51806	13	3.8	1.08	73.07	0.56	8.38	0	0.12	2
138	1.51711	12.89	3.62	1.57	72.96	0.61	8.11	0	0	2
139	1.51674	12.79	3.52	1.54	73.36	0.66	7.9	0	0	2
140	1.51674	12.87	3.56	1.64	73.14	0.65	7.99	0	0	2
141	1.5169	13.33	3.54	1.61	72.54	0.68	8.11	0	0	2
142	1.51851	13.2	3.63	1.07	72.83	0.57	8.41	0.09	0.17	2
143	1.51662	12.85	3.51	1.44	73.01	0.68	8.23	0.06	0.25	2
144	1.51709	13	3.47	1.79	72.72	0.66	8.18	0	0	2
145	1.5166	12.99	3.18	1.23	72.97	0.58	8.81	0	0.24	2
146	1.51839	12.85	3.67	1.24	72.57	0.62	8.68	0	0.35	2
147	1.51769	13.65	3.66	1.11	72.77	0.11	8.6	0	0	3
148	1.5161	13.33	3.53	1.34	72.67	0.56	8.33	0	0	3
149	1.5167	13.24	3.57	1.38	72.7	0.56	8.44	0	0.1	3
150	1.51643	12.16	3.52	1.35	72.89	0.57	8.53	0	0	3
151	1.51665	13.14	3.45	1.76	72.48	0.6	8.38	0	0.17	3
152	1.52127	14.32	3.9	0.83	71.5	0	9.49	0	0	3
153	1.51779	13.64	3.65	0.65	73	0.06	8.93	0	0	3
154	1.5161	13.42	3.4	1.22	72.69	0.59	8.32	0	0	3
155	1.51694	12.86	3.58	1.31	72.61	0.61	8.79	0	0	3
156	1.51646	13.04	3.4	1.26	73.01	0.52	8.58	0	0	3
Observation	RI	Na	Mg	Al	Si	K	Ca	Ba	Fe	Type
157	1.51655	13.41	3.39	1.28	72.64	0.52	8.65	0	0	3
158	1.52121	14.03	3.76	0.58	71.79	0.11	9.65	0	0	3
159	1.51776	13.53	3.41	1.52	72.04	0.58	8.79	0	0	3
160	1.51796	13.5	3.36	1.63	71.94	0.57	8.81	0	0.09	3
161	1.51832	13.33	3.34	1.54	72.14	0.56	8.99	0	0	3
162	1.51934	13.64	3.54	0.75	72.65	0.16	8.89	0.15	0.24	3
163	1.52211	14.19	3.78	0.91	71.36	0.23	9.14	0	0.37	3
164	1.51514	14.01	2.68	3.5	69.89	1.68	5.87	2.2	0	5
165	1.51915	12.73	1.85	1.86	72.69	0.6	10.09	0	0	5

166	1.52171	11.56	1.88	1.56	72.86	0.47	11.41	0	0	5
167	1.52151	11.03	1.71	1.56	73.44	0.58	11.62	0	0	5
168	1.51969	12.64	0	1.65	73.75	0.38	11.53	0	0	5
169	1.51666	12.86	0	1.83	73.88	0.97	10.17	0	0	5
170	1.51994	13.27	0	1.76	73.03	0.47	11.32	0	0	5
171	1.52369	13.44	0	1.58	72.22	0.32	12.24	0	0	5
172	1.51316	13.02	0	3.04	70.48	6.21	6.96	0	0	5
173	1.51321	13	0	3.02	70.7	6.21	6.93	0	0	5
174	1.52043	13.38	0	1.4	72.25	0.33	12.5	0	0	5
175	1.52058	12.85	1.61	2.17	72.18	0.76	9.7	0.24	0.51	5
176	1.52119	12.97	0.33	1.51	73.39	0.13	11.27	0	0.28	5
177	1.51905	14	2.39	1.56	72.37	0	9.57	0	0	6
178	1.51937	13.79	2.41	1.19	72.76	0	9.77	0	0	6
179	1.51829	14.46	2.24	1.62	72.38	0	9.26	0	0	6
180	1.51852	14.09	2.19	1.66	72.67	0	9.32	0	0	6
181	1.51299	14.4	1.74	1.54	74.55	0	7.59	0	0	6
182	1.51888	14.99	0.78	1.74	72.5	0	9.95	0	0	6
183	1.51916	14.15	0	2.09	72.74	0	10.88	0	0	6
184	1.51969	14.56	0	0.56	73.48	0	11.22	0	0	6
185	1.51115	17.38	0	0.34	75.41	0	6.65	0	0	6
186	1.51131	13.69	3.2	1.81	72.81	1.76	5.43	1.19	0	7
187	1.51838	14.32	3.26	2.22	71.25	1.46	5.79	1.63	0	7
188	1.52315	13.44	3.34	1.23	72.38	0.6	8.83	0	0	7
189	1.52247	14.86	2.2	2.06	70.26	0.76	9.76	0	0	7
190	1.52365	15.79	1.83	1.31	70.43	0.31	8.61	1.68	0	7
191	1.51613	13.88	1.78	1.79	73.1	0	8.67	0.76	0	7
192	1.51602	14.85	0	2.38	73.28	0	8.76	0.64	0.09	7
193	1.51623	14.2	0	2.79	73.46	0.04	9.04	0.4	0.09	7
194	1.51719	14.75	0	2	73.02	0	8.53	1.59	0.08	7
195	1.51683	14.56	0	1.98	73.29	0	8.52	1.57	0.07	7
196	1.51545	14.14	0	2.68	73.39	0.08	9.07	0.61	0.05	7
Observation	RI	Na	Mg	Al	Si	K	Ca	Ba	Fe	Type
197	1.51556	13.87	0	2.54	73.23	0.14	9.41	0.81	0.01	7
198	1.51727	14.7	0	2.34	73.28	0	8.95	0.66	0	7
199	1.51531	14.38	0	2.66	73.1	0.04	9.08	0.64	0	7
200	1.51609	15.01	0	2.51	73.05	0.05	8.83	0.53	0	7
201	1.51508	15.15	0	2.25	73.5	0	8.34	0.63	0	7
202	1.51653	11.95	0	1.19	75.18	2.7	8.93	0	0	7
203	1.51514	14.85	0	2.42	73.72	0	8.39	0.56	0	7
204	1.51658	14.8	0	1.99	73.11	0	8.28	1.71	0	7
205	1.51617	14.95	0	2.27	73.3	0	8.71	0.67	0	7

206	1.51732	14.95	0	1.8	72.99	0	8.61	1.55	0	7
207	1.51645	14.94	0	1.87	73.11	0	8.67	1.38	0	7
208	1.51831	14.39	0	1.82	72.86	1.41	6.47	2.88	0	7
209	1.5164	14.37	0	2.74	72.85	0	9.45	0.54	0	7
210	1.51623	14.14	0	2.88	72.61	0.08	9.18	1.06	0	7
211	1.51685	14.92	0	1.99	73.06	0	8.4	1.59	0	7
212	1.52065	14.36	0	2.02	73.42	0	8.44	1.64	0	7
213	1.51651	14.38	0	1.94	73.61	0	8.48	1.57	0	7
214	1.51711	14.23	0	2.08	73.36	0	8.62	1.67	0	7

Appendix B. Ozone Data

Observation	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13
16	1	16	5	7	5870	2	19	61	64.94	5000	-67	65.48	200
17	1	17	6	5	5840	5	19	64	*NA*	5000	-40	63.32	200
18	1	18	7	9	5780	4	59	67	*NA*	639	1	66.02	150
19	1	19	1	4	5680	5	73	52	56.48	393	-68	69.8	10
20	1	20	2	3	5720	4	19	54	*NA*	5000	-66	54.68	140

21	1	21	3	4	5760	3	19	54	53.6	5000	-58	51.98	250
22	1	22	4	4	5730	4	26	58	52.7	5000	-26	51.98	200
23	1	23	5	5	5700	5	59	69	51.08	3044	18	52.88	150
24	1	24	6	6	5650	5	70	51	*NA*	3641	23	47.66	140
25	1	25	7	9	5680	3	64	53	*NA*	111	-10	59.54	50
26	1	26	1	5	5780	3	*NA*	56	53.6	692	-25	67.1	0
27	1	27	2	6	5820	5	19	59	59.36	597	-52	70.52	70
28	1	28	3	6	5830	4	*NA*	59	60.08	*NA*	-44	*NA*	150
29	1	29	4	6	5810	5	19	64	56.66	1791	-15	64.76	150
30	1	30	5	11	5790	3	28	63	57.38	793	-15	65.84	120
31	1	31	6	10	5800	2	32	63	*NA*	531	-38	75.92	40
32	2	1	7	7	5820	5	19	62	*NA*	419	-29	75.74	120
33	2	2	1	12	5770	8	76	63	57.2	816	-7	66.2	6
34	2	3	2	9	5670	3	69	54	45.5	3651	62	49.1	30
35	2	4	3	2	5590	3	76	36	37.4	5000	70	37.94	100
36	2	5	4	3	5410	6	64	31	32.18	5000	28	32.36	200
37	2	6	5	3	5350	7	62	30	32.54	1341	18	45.86	60
38	2	7	6	2	5480	9	72	36	*NA*	5000	0	38.66	350
39	2	8	7	3	5600	7	76	42	*NA*	3799	-18	45.86	250
40	2	9	1	3	5490	11	72	37	38.48	5000	32	38.12	350
41	2	10	2	4	5560	10	72	41	40.46	5000	-1	37.58	300
42	2	11	3	6	5700	3	32	46	*NA*	5000	-30	45.86	300
43	2	12	4	8	5680	5	50	51	47.12	5000	-8	45.5	300
44	2	13	5	6	5700	4	86	55	49.28	2398	21	53.78	200
45	2	14	6	4	5650	5	61	41	*NA*	5000	51	36.32	100
46	2	15	7	3	5610	5	62	41	*NA*	4281	42	41.36	250
47	2	16	1	7	5730	5	66	49	*NA*	1161	27	52.88	200
48	2	17	2	11	5770	5	68	45	52.88	2778	2	55.76	200
Observation	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13
49	2	18	3	13	5770	3	82	55	55.4	442	26	58.28	40
50	2	19	4	4	5700	5	*NA*	45	38.12	*NA*	82	*NA*	2
51	2	20	5	6	5690	8	21	41	43.88	5000	-30	42.26	300
52	2	21	6	5	5700	3	19	45	*NA*	5000	-53	43.88	300
53	2	22	7	4	5730	11	19	51	*NA*	5000	-43	49.1	300
54	2	23	1	4	5690	7	19	53	50.18	5000	7	49.1	300

55	2	24	2	6	5640	5	68	50	37.4	5000	24	42.08	300
56	2	25	3	10	5720	6	63	60	53.06	1341	19	59.18	150
57	2	26	4	15	5740	3	54	54	56.48	1318	2	64.58	150
58	2	27	5	23	5740	3	47	53	58.82	885	-4	67.1	80
59	2	28	6	17	5740	3	56	53	*NA*	360	3	67.1	40
60	2	29	7	7	5670	7	61	44	*NA*	3497	73	49.46	40
61	3	1	1	2	5550	10	74	40	38.84	5000	73	40.1	80
62	3	2	2	3	5470	7	46	30	29.66	5000	44	29.3	300
63	3	3	3	3	5320	11	45	25	27.68	5000	39	27.5	200
64	3	4	4	5	*NA*	8	33	39	30.2	5000	15	30.02	500
65	3	5	5	4	5530	3	43	40	36.14	5000	-12	33.62	140
66	3	6	6	6	5600	3	21	45	*NA*	5000	-2	39.02	140
67	3	7	7	7	5660	7	57	51	*NA*	5000	30	42.08	140
68	3	8	1	7	5580	5	42	48	40.64	3608	24	39.38	100
69	3	9	2	6	5510	5	50	45	36.86	5000	38	32.9	140
70	3	10	3	3	5530	5	61	47	33.8	5000	56	35.6	200
71	3	11	4	2	5620	9	61	43	37.04	5000	66	34.34	120
72	3	12	5	8	5690	0	60	49	46.04	613	-27	59.72	300
73	3	13	6	12	5760	4	31	56	*NA*	334	-9	64.4	300
74	3	14	7	12	5740	3	66	53	*NA*	567	13	61.88	150
75	3	15	1	16	5780	5	53	61	57.92	488	-20	64.94	2
76	3	16	2	9	5790	2	42	63	57.02	531	-15	71.06	50
77	3	17	3	24	5760	3	60	70	58.64	508	7	66.56	70
78	3	18	4	13	5700	4	82	57	50.36	1571	68	56.3	17
79	3	19	5	8	5680	4	57	35	40.1	721	28	55.4	140
80	3	20	6	10	5720	5	21	52	*NA*	505	-49	67.28	140
81	3	21	7	8	5720	5	19	59	*NA*	377	-27	73.22	300
82	3	22	1	9	5730	4	32	67	59.54	442	-9	75.74	200
Observation	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13
83	3	23	2	10	5710	5	77	57	57.38	902	54	60.44	250
84	3	24	3	13	5750	6	70	*NA*	56.3	3188	53	58.64	80
85	3	25	4	14	5720	4	71	42	44.96	1381	4	56.3	60
86	3	26	5	9	5710	3	19	55	51.8	5000	-16	50	100
87	3	27	6	11	5600	6	45	40	*NA*	5000	38	46.94	150
88	3	28	7	7	5630	4	44	39	*NA*	1302	40	52.7	150

89	3	29	1	9	5690	7	70	57	46.58	1292	-5	53.6	200
90	3	30	2	12	5730	6	45	58	52.52	5000	-14	52.7	100
91	3	31	3	12	5710	3	46	62	52.52	472	34	62.96	300
92	4	1	4	8	5610	6	50	51	50	1404	42	54.5	120
93	4	2	5	9	5680	5	69	61	51.44	944	35	55.76	100
94	4	3	6	5	5620	6	67	34	*NA*	5000	75	35.24	200
95	4	4	7	4	5420	7	69	35	*NA*	5000	41	30.92	200
96	4	5	1	4	5540	5	54	35	33.26	5000	62	33.44	200
97	4	6	2	9	5590	6	51	48	38.12	5000	44	42.08	300
98	4	7	3	13	5690	6	63	59	52.88	2014	31	53.42	300
99	4	8	4	5	5550	7	63	41	37.58	5000	56	37.22	250
100	4	9	5	10	5620	7	57	58	46.76	5000	27	47.66	120
101	4	10	6	10	5630	6	61	51	*NA*	524	57	54.68	140
102	4	11	7	7	5580	7	78	46	*NA*	5000	55	38.48	200
103	4	12	1	5	5560	4	65	40	34.7	5000	59	35.24	140
104	4	13	2	4	5440	5	44	35	33.08	5000	24	032.54	80
105	4	14	3	7	5480	7	51	46	37.4	2490	29	47.48	300
106	4	15	4	3	5620	5	73	39	39.56	5000	107	31.28	100
107	4	16	5	4	5450	11	35	32	*NA*	5000	36	33.44	300
108	4	17	6	7	5660	6	35	47	*NA*	5000	28	39.38	200
109	4	18	7	11	5680	6	61	50	*NA*	1144	30	53.6	120
110	4	19	1	15	5760	4	50	65	56.3	547	1	66.92	100
111	4	20	2	22	5790	4	57	66	63.68	413	10	69.62	120
112	4	21	3	17	5720	5	68	69	60.8	610	46	63.68	60
113	4	22	4	7	5660	6	58	59	42.8	3638	81	51.26	120
114	4	23	5	10	5710	5	65	64	56.3	3848	45	56.84	100
115	4	24	6	19	5780	7	78	68	*NA*	1479	40	68	100
116	4	25	7	18	5750	7	73	49	*NA*	1108	55	65.48	27
Observation	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13
117	4	26	1	12	5700	5	41	52	49.64	869	0	58.1	40
118	4	27	2	6	5620	9	47	56	39.92	5000	43	38.3	140
119	4	28	3	9	5650	6	46	55	40.82	5000	49	37.94	150
120	4	29	4	19	5730	5	61	66	59.72	1148	31	60.8	100
121	4	30	5	21	5810	4	55	74	67.28	856	4	75.38	100
122	5	1	6	29	5790	4	60	76	*NA*	807	16	73.04	120

123	5	2	7	16	5740	8	78	70	*NA*	2040	46	63.5	150
124	5	3	1	5	*NA*	8	62	61	44.42	5000	63	42.62	100
125	5	4	2	11	5690	4	71	67	52.52	314	60	59	120
126	5	5	3	2	*NA*	4	67	45	41.72	5000	77	42.62	80
127	5	6	4	2	5680	6	77	41	42.98	5000	75	40.82	120
128	5	7	5	12	5650	8	66	61	51.8	1410	20	55.22	140
129	5	8	6	16	5730	6	74	68	*NA*	360	23	62.42	120
130	5	9	7	22	5730	3	78	69	*NA*	1568	32	67.64	70
131	5	10	1	20	5760	7	78	74	63.14	1184	40	68.72	80
132	5	11	2	27	5830	6	75	74	67.28	898	24	73.4	70
133	5	12	3	33	5880	3	80	80	73.04	436	0	86.36	40
134	5	13	4	25	5890	6	88	84	73.22	774	6	86	20
135	5	14	5	31	5850	4	76	78	71.24	1181	50	79.88	17
136	5	15	6	18	5820	6	63	80	*NA*	1991	47	69.62	40
137	5	16	7	16	*NA*	7	68	73	*NA*	2057	71	67.28	50
138	5	17	1	24	5800	7	78	76	*NA*	1597	56	68	50
139	5	18	2	16	5740	3	74	74	*NA*	1184	52	69.44	70
140	5	19	3	12	5710	7	63	66	*NA*	3005	58	59.18	80
141	5	20	4	9	5720	8	62	66	*NA*	2880	53	57.38	120
142	5	21	5	12	5710	7	69	63	*NA*	*NA*	66	56.3	120
143	5	22	6	16	5740	5	53	69	*NA*	2125	64	59	100
144	5	23	7	*NA*	5720	3	64	66	*NA*	1751	67	59.9	120
145	5	24	1	8	5690	9	62	62	*NA*	3720	74	50.9	120
146	5	25	2	9	5730	5	71	67	49.82	4337	66	59.36	200
147	5	26	3	29	5780	3	68	80	*NA*	2053	31	72.86	120
148	5	27	4	20	5790	7	79	76	*NA*	1958	70	70.52	40
149	5	28	5	5	5750	3	76	65	51.08	3644	86	59.36	70
150	5	29	6	5	5680	6	71	65	*NA*	1368	75	58.46	100
Observation	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13
151	5	30	7	11	5720	3	66	63	*NA*	3539	73	53.6	120
152	5	31	1	12	5770	4	81	62	*NA*	2785	49	63.32	100
153	6	1	2	19	5800	4	72	68	*NA*	984	26	69.26	120
154	6	2	3	17	5780	8	92	68	64.76	1804	56	68	70
155	6	3	4	19	5740	6	71	69	64.4	3234	77	62.78	80
156	6	4	5	16	5730	6	64	66	60.44	3441	67	60.98	100

157	6	5	6	14	5760	6	68	70	*NA*	1578	61	60.8	100
158	6	6	7	10	5770	7	59	70	*NA*	1850	76	60.8	120
159	6	7	1	9	5690	8	67	64	54.5	2962	80	59.36	120
160	6	8	2	7	5650	6	66	61	53.78	2670	54	55.4	120
161	6	9	3	5	5610	3	61	52	42.08	5000	76	42.08	150
162	6	10	4	2	5570	9	81	48	41.72	5000	57	40.82	140
163	6	11	5	12	5690	5	63	59	51.8	5000	46	51.26	140
164	6	12	6	22	5760	3	58	67	*NA*	987	28	63.86	140
165	6	13	7	17	5810	5	68	66	*NA*	1148	43	66.92	140
166	6	14	1	26	5830	4	71	74	71.78	898	-24	77.9	60
167	6	15	2	27	5880	6	67	83	72.5	777	-1	82.58	30
168	6	16	3	14	5860	3	64	78	68.72	1279	75	71.6	17
169	6	17	4	11	5830	6	64	75	66.2	1046	69	68.72	80
170	6	18	5	23	5870	4	69	84	74.12	1167	50	74.3	60
171	6	19	6	26	5860	3	77	81	*NA*	987	45	75.74	100
172	6	20	7	21	5800	3	61	79	*NA*	1144	57	71.24	120
173	6	21	1	15	5800	4	69	79	66.2	977	60	70.7	150
174	6	22	2	20	5770	5	64	65	65.12	770	26	75.56	120
175	6	23	3	15	5860	4	33	81	72.68	629	-11	86.36	140
176	6	24	4	18	5870	7	38	84	76.1	337	-14	89.78	140
177	6	25	5	26	5870	4	54	83	*NA*	590	26	85.1	120
178	6	26	6	19	5860	6	39	90	*NA*	400	19	83.3	120
179	6	27	7	13	5880	5	43	90	*NA*	580	9	87.26	80
180	6	28	1	30	5870	7	55	93	*NA*	646	25	89.24	140
181	6	29	2	26	5860	4	77	88	*NA*	826	41	84.38	140
182	6	30	3	15	5830	5	63	72	*NA*	823	52	74.48	150
183	7	1	4	16	5820	5	65	72	*NA*	2116	47	70.34	120
184	7	2	5	16	5820	8	64	70	*NA*	2972	52	64.4	120
Observation	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13
185	7	3	6	19	5860	6	68	78	*NA*	2752	41	69.98	140
186	7	4	7	23	5870	3	76	87	*NA*	1377	37	78.44	100
187	7	5	1	28	5890	6	71	91	*NA*	1486	33	79.88	50
188	7	6	2	34	5900	6	86	87	81.68	990	22	85.1	40
189	7	7	3	33	5890	5	65	91	*NA*	508	29	85.28	100
190	7	8	4	*NA*	5890	5	69	90	*NA*	688	37	83.3	80

191	7	9	5	24	5910	4	73	88	79.88	1204	56	79.88	100
192	7	10	6	17	5900	5	69	83	*NA*	2414	63	76.46	60
193	7	11	7	10	5860	3	64	78	*NA*	2385	67	70.34	50
194	7	12	1	14	5830	3	63	79	70.88	2326	64	71.24	70
195	7	13	2	13	5850	9	72	77	69.44	3389	56	68.72	80
196	7	14	3	17	5830	6	82	81	73.22	2818	58	71.78	80
197	7	15	4	15	*NA*	6	83	76	72.32	3083	75	72.32	80
198	7	16	5	22	5810	8	69	76	67.64	2394	54	69.62	90
199	7	17	6	19	5830	4	74	78	*NA*	2746	61	69.44	120
200	7	18	7	20	5830	5	69	75	*NA*	2493	55	72.5	120
201	7	19	1	25	5840	7	72	82	68	1528	42	73.94	100
202	7	20	2	28	5870	6	73	84	74.12	111	40	78.08	60
203	7	21	3	29	5870	4	90	86	74.48	1899	45	76.46	40
204	7	22	4	*NA*	5850	4	79	70	65.12	2020	37	73.22	50
205	7	23	5	23	5860	3	80	80	67.28	1289	32	75.2	40
206	7	24	6	26	5900	3	73	80	*NA*	984	35	78.8	70
207	7	25	7	14	5890	4	71	84	*NA*	836	28	81.5	80
208	7	26	1	13	5880	4	78	84	70.7	826	27	79.34	80
209	7	27	2	26	5890	6	80	81	69.8	1105	39	74.12	80
210	7	28	3	22	5870	8	74	85	71.42	1023	46	77.18	80
211	7	29	4	11	*NA*	6	70	79	67.82	1453	68	70.16	80
212	7	30	5	15	*NA*	6	71	72	66.02	2375	52	66.2	100
213	7	31	6	14	5820	6	63	73	*NA*	2956	46	67.28	120
214	8	1	7	13	5780	6	57	72	*NA*	2988	56	65.66	150
215	8	2	1	9	5770	3	55	68	62.6	4291	60	62.24	200
216	8	3	2	12	5790	4	65	65	56.48	3330	59	58.64	150
217	8	4	3	15	5820	6	*NA*	64	62.06	*NA*	31	*NA*	150
218	8	5	4	12	5840	6	*NA*	75	65.12	*NA*	35	*NA*	150
Observation	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13
219	8	6	5	15	5800	7	*NA*	69	63.32	*NA*	49	*NA*	150
220	8	7	6	25	5830	4	*NA*	69	*NA*	*NA*	30	*NA*	100
221	8	8	7	18	5800	3	*NA*	72	*NA*	*NA*	43	*NA*	100
222	8	9	1	14	5840	7	65	79	67.1	1233	30	70.52	100
223	8	10	2	22	*NA*	3	74	78	66.02	1450	36	69.8	30
224	8	11	3	24	5910	5	72	81	70.88	1069	28	74.3	80

225	8	12	4	19	5890	5	79	80	69.8	984	57	73.4	70
226	8	13	5	16	5870	6	62	76	*NA*	1653	71	68.72	60
227	8	14	6	7	5780	7	65	59	*NA*	3930	68	59.18	150
228	8	15	7	2	5730	5	77	55	*NA*	5000	73	51.62	200
229	8	16	1	4	5780	7	70	66	51.44	5000	45	51.26	200
230	8	17	2	6	5750	7	58	64	56.48	4212	46	56.84	200
231	8	18	3	12	5760	5	58	62	52.16	5000	52	49.82	250
232	8	19	4	9	5730	7	72	67	*NA*	5000	31	57.38	300
233	8	20	5	15	5730	5	77	74	64.04	1545	43	65.66	70
234	8	21	6	17	5790	4	57	74	*NA*	994	44	69.62	300
235	8	22	7	13	5750	3	67	70	*NA*	1125	55	68	150
236	8	23	1	20	5880	3	73	77	*NA*	636	16	73.94	300
237	8	24	2	22	5890	7	70	83	70.88	748	32	77	30
238	8	25	3	24	5880	4	73	81	73.76	692	44	77.72	100
239	8	26	4	26	5870	7	73	73	75.2	807	39	78.8	100
240	8	27	5	32	5900	6	71	87	76.46	869	19	78.98	17
241	8	28	6	33	5920	4	77	89	*NA*	800	24	85.64	20
242	8	29	7	27	5930	3	68	92	*NA*	393	6	91.76	4
243	8	30	1	38	5950	5	62	92	82.4	557	0	90.68	70
244	8	31	2	23	5950	8	61	93	81.68	620	27	85.64	30
245	9	1	3	19	5900	5	71	93	82.58	1404	33	84.74	70
246	9	2	4	19	5890	8	77	86	71.42	898	21	80.6	60
247	9	3	5	15	5860	7	71	76	69.44	377	-2	83.3	40
248	9	4	6	28	5840	5	67	81	*NA*	528	17	78.8	50
249	9	5	7	10	5800	6	74	78	*NA*	2818	26	72.68	70
250	9	6	1	14	5760	7	65	73	*NA*	3247	10	67.28	140
251	9	7	2	26	5810	6	82	80	71.78	895	0	78.08	100
252	9	8	3	17	5850	4	67	81	70.88	721	0	80.24	120
Observation	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13
253	9	9	4	3	*NA*	5	73	69	66.92	774	-27	75.56	100
254	9	10	5	2	*NA*	6	74	59	61.88	134	0	77.18	70
255	9	11	6	3	5760	7	87	52	*NA*	5000	39	51.8	150
256	9	12	7	14	5860	4	71	63	*NA*	1965	13	60.98	50
257	9	13	1	29	5830	5	77	72	68.72	1853	10	70.88	70
258	9	14	2	18	5840	5	78	75	69.26	2342	7	71.42	40

259	9	15	3	3	5800	7	72	55	54.32	5000	56	51.62	70
260	9	16	4	7	*NA*	10	67	59	51.44	5000	37	47.48	120
261	9	17	5	9	5790	3	71	61	50.54	4028	35	55.04	140
262	9	18	6	19	5830	5	71	71	*NA*	2716	26	63.68	140
263	9	19	7	8	5810	5	76	71	*NA*	3671	31	65.84	100
264	9	20	1	*NA*	5770	5	76	66	62.96	3431	26	62.24	50
265	9	21	2	23	5780	6	76	72	66.38	3795	31	66.92	70
266	9	22	3	13	5800	6	73	75	67.1	3120	35	66.92	40
267	9	23	4	*NA*	5770	3	66	71	63.68	4133	28	63.5	40
268	9	24	5	7	5800	5	80	65	60.08	2667	17	63.5	100
269	9	25	6	3	5780	9	73	61	*NA*	5000	39	52.7	120
270	9	26	7	5	5790	8	80	60	*NA*	5000	36	48.92	120
271	9	27	1	11	5770	5	75	64	55.94	308	25	68.72	140
272	9	28	2	12	5750	4	68	61	59.54	2982	18	59.9	120
273	9	29	3	5	5640	5	93	63	54.32	5000	30	52.7	70
274	9	30	4	4	5640	7	57	62	54.32	5000	25	51.26	150
275	10	1	5	5	5650	3	70	59	50.9	5000	38	47.66	200
276	10	2	6	4	5710	6	65	56	*NA*	5000	35	47.84	200
277	10	3	7	10	5760	6	66	59	*NA*	3070	13	60.08	200
278	10	4	1	17	5840	4	73	72	63.14	830	0	72.14	70
279	10	5	2	26	5880	3	77	71	67.64	711	-9	75.56	40
280	10	6	3	30	5890	5	80	75	71.06	1049	-10	78.98	50
281	10	7	4	18	5890	4	73	71	70.88	511	-39	83.84	17
282	10	8	5	12	5890	5	19	71	70.52	5000	-40	67.64	80
283	10	9	6	7	5890	6	19	73	*NA*	5000	-34	69.44	250
284	10	10	7	15	5850	3	73	78	*NA*	377	-3	78.8	200
285	10	11	1	12	5830	5	76	73	*NA*	862	27	73.58	2
286	10	12	2	7	5830	8	77	71	67.1	337	-17	81.14	20

Observation	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13
287	10	13	3	28	5860	5	86	73	69.8	492	-2	82.22	7
288	10	14	4	22	5830	5	76	71	69.44	1394	13	75.02	30
289	10	15	5	18	5800	7	66	66	62.96	3146	27	64.04	50
290	10	16	6	14	5830	4	74	69	*NA*	2234	11	66.74	70
291	10	17	7	24	5790	5	71	69	*NA*	2109	21	69.62	17
292	10	18	1	10	5730	4	84	64	56.3	5000	23	54.5	80

293	10	19	2	14	5780	5	74	65	63.14	2270	-7	68.9	50
294	10	20	3	9	5740	7	48	54	62.96	2191	-13	68.72	60
295	10	21	4	12	5710	8	75	62	58.64	3448	12	58.64	60
296	10	22	5	7	5690	6	74	56	52.34	5000	13	48.92	80
297	10	23	6	7	5670	4	67	55	*NA*	5000	11	49.46	50
298	10	24	7	6	5760	4	75	58	*NA*	2719	25	56.84	50
299	10	25	1	13	5820	5	71	48	*NA*	1899	21	62.06	40
300	10	26	2	5	5790	3	35	54	*NA*	5000	-41	52.52	40
301	10	27	3	3	5760	5	23	57	53.42	5000	-21	50.9	300
302	10	28	4	7	5800	6	19	60	57.02	5000	-19	54.32	200
303	10	29	5	8	5810	7	59	61	55.76	2385	10	60.44	150
304	10	30	6	10	5750	4	60	63	*NA*	1938	0	62.6	100
305	10	31	7	12	5840	0	38	65	*NA*	590	-11	69.98	100
306	11	1	1	7	5860	3	*NA*	66	65.12	*NA*	-32	*NA*	60
307	11	2	2	5	5870	6	*NA*	68	68.9	*NA*	-42	*NA*	150
308	11	3	3	6	5920	3	22	71	69.08	328	-40	80.6	150
309	11	4	4	4	5900	0	*NA*	70	68.72	*NA*	-43	*NA*	200
310	11	5	5	5	5860	7	19	70	62.78	5000	-29	61.7	300
311	11	6	6	11	5840	3	*NA*	70	*NA*	*NA*	-9	*NA*	120
312	11	7	7	20	5840	0	45	68	*NA*	597	-22	73.58	30
313	11	8	1	4	5850	5	*NA*	64	64.04	*NA*	-25	*NA*	100
314	11	9	2	14	5810	2	47	69	60.98	469	-4	71.78	50
315	11	10	3	16	5770	2	73	59	57.2	1541	18	63.14	20
316	11	11	4	5	5710	4	67	49	44.24	5000	24	41.9	200
317	11	12	5	3	5500	9	56	39	41.36	5000	15	41.72	120
318	11	13	6	5	5660	3	54	50	*NA*	5000	27	44.6	300
319	11	14	7	1	5700	3	71	46	*NA*	5000	54	42.8	200
320	11	15	1	5	5810	5	59	54	54.5	5000	-28	53.6	70
Observation	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13
321	11	16	2	4	5860	0	25	60	61.52	5000	-38	63.5	140
322	11	17	3	11	5900	0	24	62	62.6	5000	-36	60.08	150
323	11	18	4	6	5850	5	41	65	59.54	2014	-20	69.98	200
324	11	19	5	8	5780	3	50	66	59.72	436	1	70.34	4
325	11	20	6	14	5790	0	76	66	*NA*	830	3	66.02	40
326	11	21	7	18	5780	2	82	63	*NA*	1112	-8	66.38	30

327	11	22	1	12	5770	2	81	62	60.62	1210	-17	67.82	30
328	11	23	2	9	5750	2	85	60	59.72	501	-22	70.88	2
329	11	24	3	7	5780	5	76	63	60.44	875	-15	68.9	0
330	11	25	4	14	5790	5	66	60	*NA*	1601	7	62.06	30
331	11	26	5	4	5750	6	58	58	42.62	5000	59	41.9	60
332	11	27	6	3	5670	8	19	34	*NA*	5000	-63	37.04	150
333	11	28	7	3	5760	0	19	36	*NA*	5000	-52	41	100
334	11	29	1	3	5770	4	19	44	51.26	2280	-54	55.76	250
335	11	30	2	3	5810	2	19	53	55.94	2047	-43	63.5	150
336	12	1	3	3	5810	2	19	52	57.74	5000	-69	56.48	200
337	12	2	4	3	5870	3	19	53	60.8	3720	-50	61.34	200
338	12	3	5	3	5830	2	27	58	59	311	-24	69.98	200
339	12	4	6	6	5760	0	64	55	*NA*	2536	28	56.48	80
340	12	5	7	6	5680	0	52	50	*NA*	1154	-22	61.52	60
341	12	6	1	5	5780	4	19	48	54.14	2933	-40	59.9	300
342	12	7	2	3	5810	3	19	51	58.28	3064	-33	62.78	200
343	12	8	3	4	5760	0	32	62	56.12	826	-16	64.76	300
344	12	9	4	7	5680	0	58	40	46.94	5000	2	42.98	50
345	12	10	5	5	5750	0	26	44	52.88	111	-52	68.18	40
346	12	11	6	5	5790	5	19	49	*NA*	5000	-48	54.68	70
347	12	12	7	4	5770	3	19	53	*NA*	5000	-37	55.58	150
348	12	13	1	3	5750	0	19	53	51.98	5000	-26	51.08	150
349	12	14	2	2	5720	0	19	53	52.7	5000	-31	51.44	70
350	12	15	3	5	5760	3	19	55	58.1	948	-48	70.7	200
351	12	16	4	3	5780	0	19	51	54.32	5000	-50	50.9	120
352	12	17	5	4	5660	4	19	54	49.64	5000	-22	48.56	150
353	12	18	6	4	5610	2	58	48	*NA*	3687	-10	46.94	150
354	12	19	7	6	5640	0	51	53	*NA*	5000	0	44.24	60
Observation	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13
355	12	20	1	6	5680	3	52	49	48.38	5000	-19	45.68	70
356	12	21	2	3	5650	5	19	48	47.12	5000	-28	45.32	150
357	12	22	3	4	5710	4	19	51	51.08	5000	-25	48.38	300
358	12	23	4	3	5680	4	57	47	45.32	508	-10	58.64	100
359	12	24	5	8	5630	4	50	50	*NA*	2851	-5	50	70
360	12	25	6	5	5770	0	*NA*	49	*NA*	*NA*	-35	*NA*	40

361	12	26	7	3	5800	7	19	51	*NA*	3143	*NA*	60.26	140
362	12	27	1	2	5730	3	53	51	49.28	111	-14	72.5	200
363	12	28	2	3	5690	3	23	51	49.28	5000	-36	51.26	70
364	12	29	3	5	5650	3	61	50	46.58	3704	18	46.94	40
365	12	30	4	1	5550	4	85	39	41	5000	8	39.92	100
366	12	31	5	2	*NA*	4	68	37	*NA*	5000	-3	37.22	70