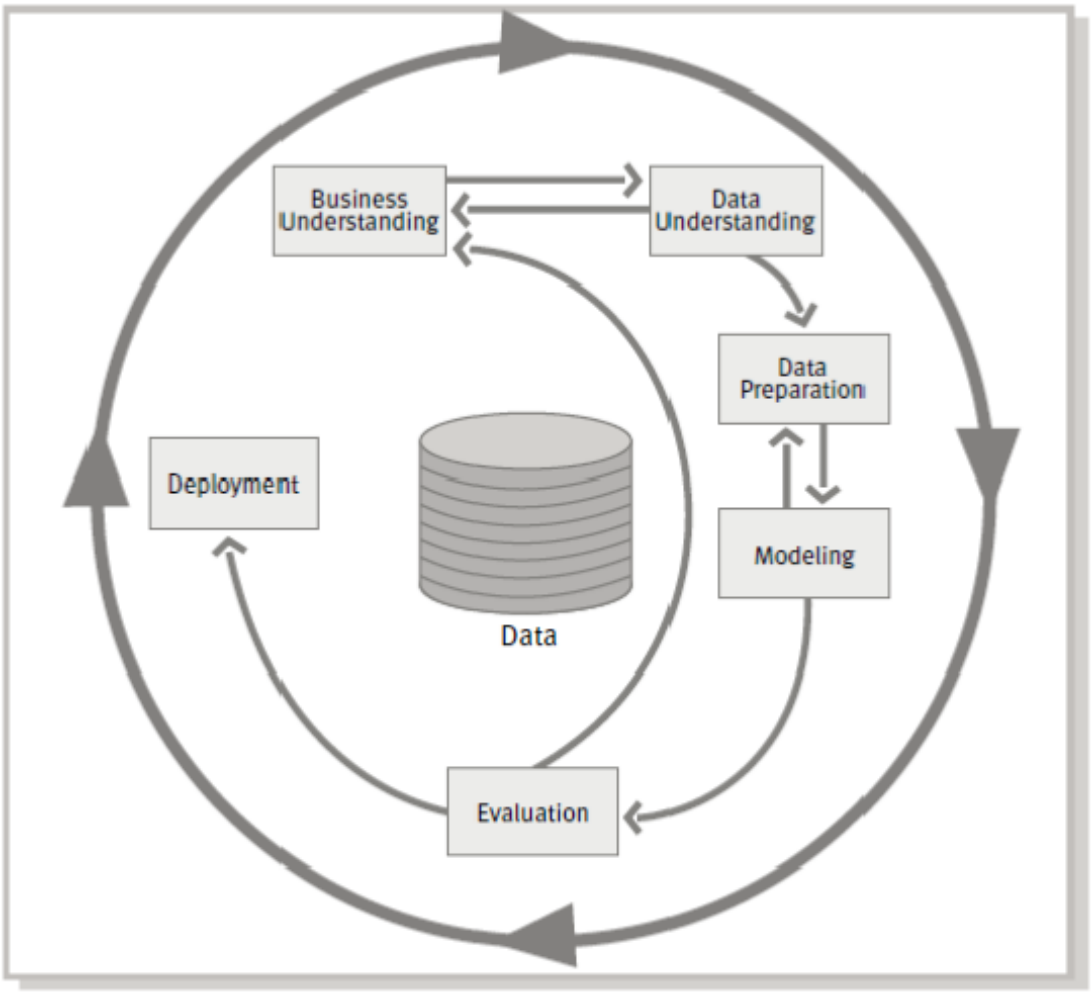
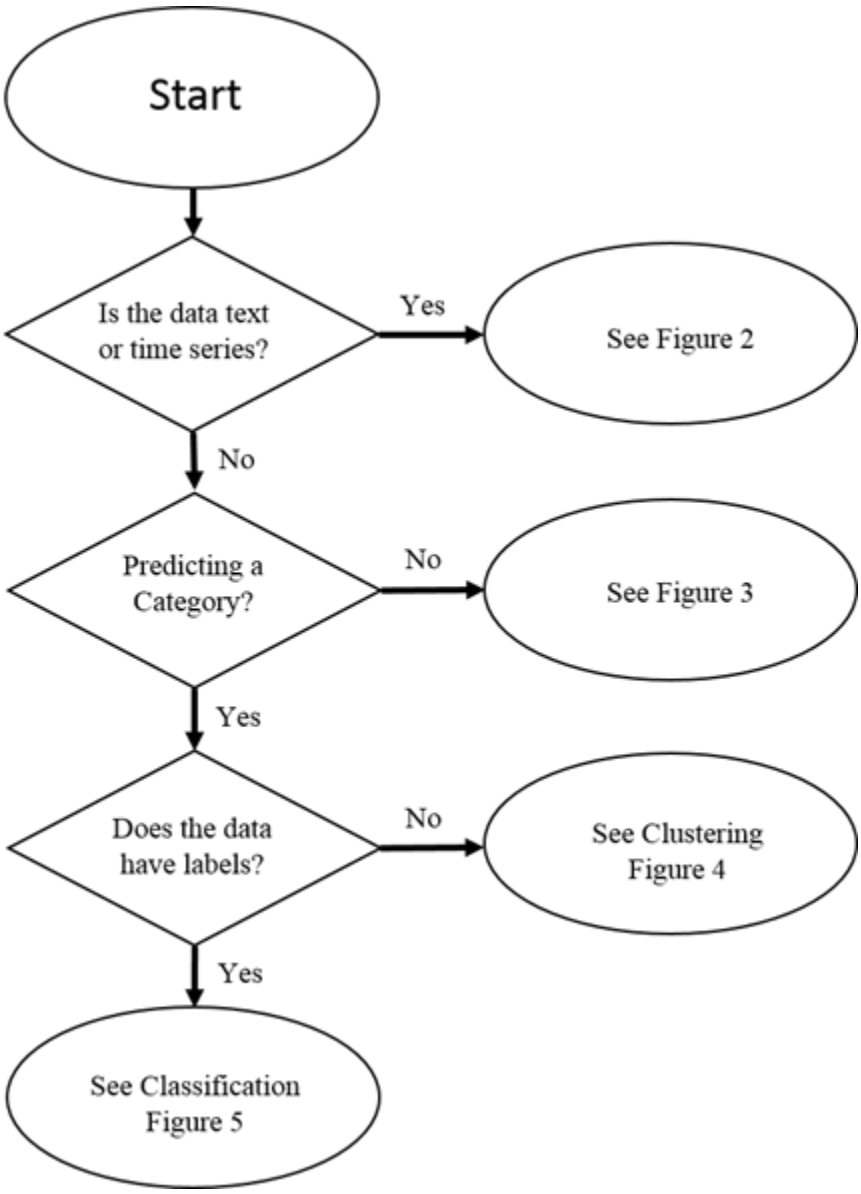
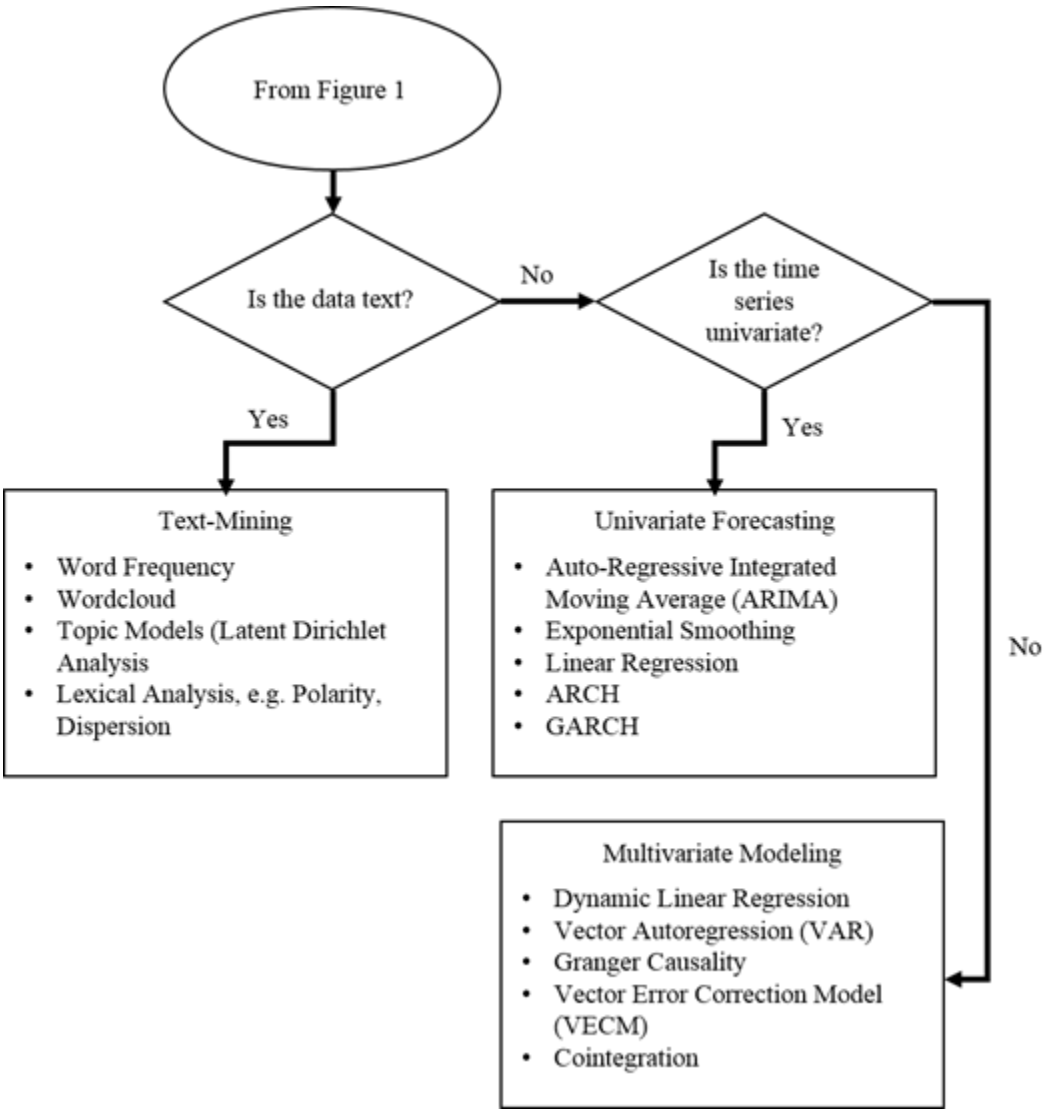
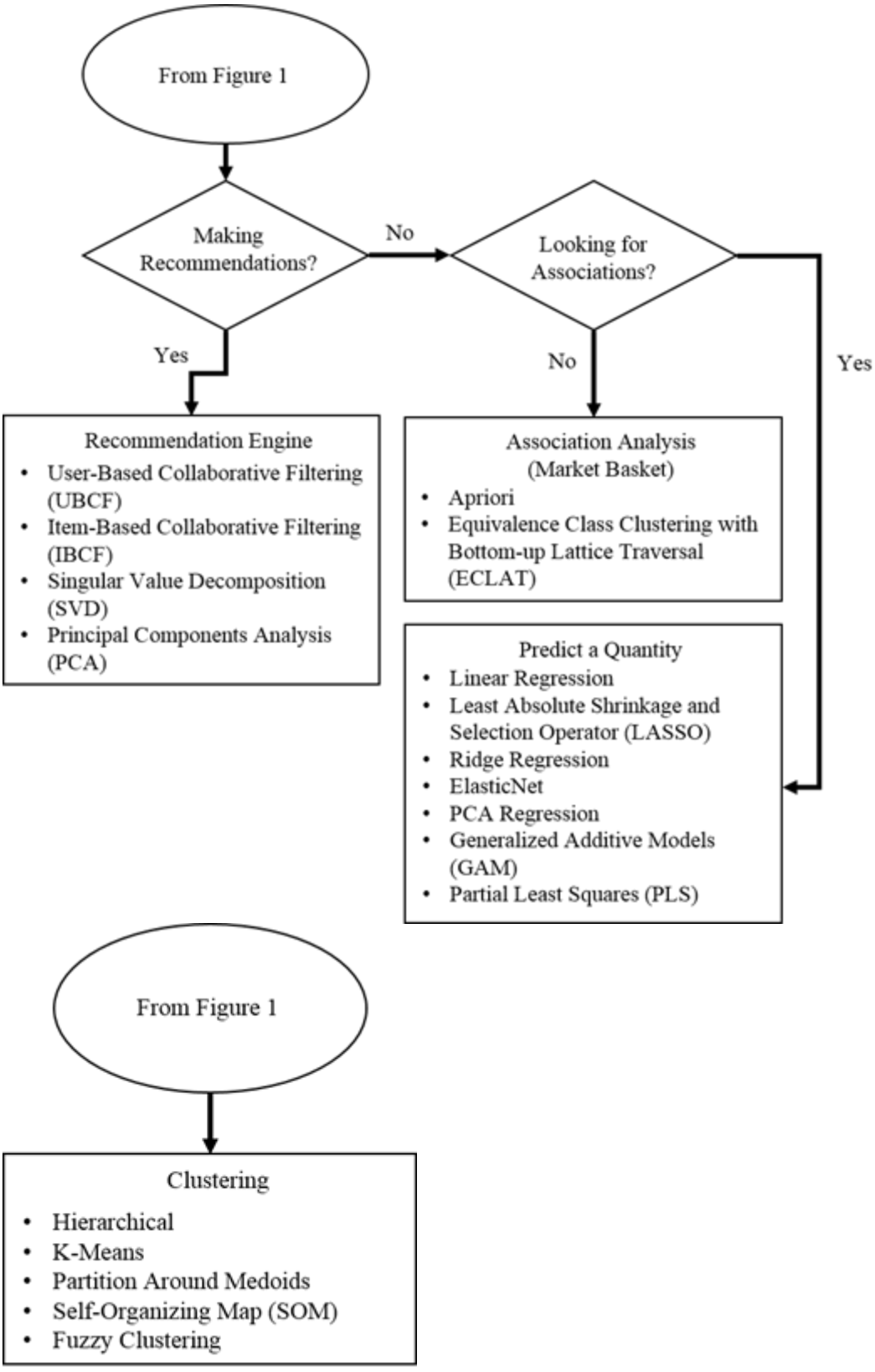
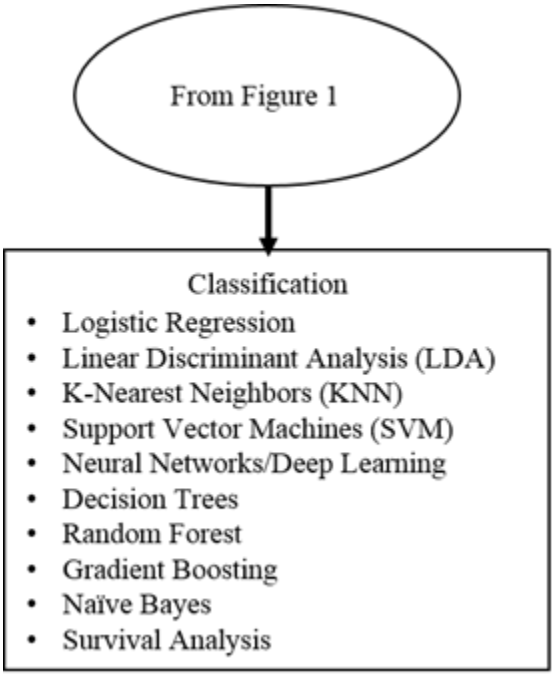
Chapter 1: A Process for Success



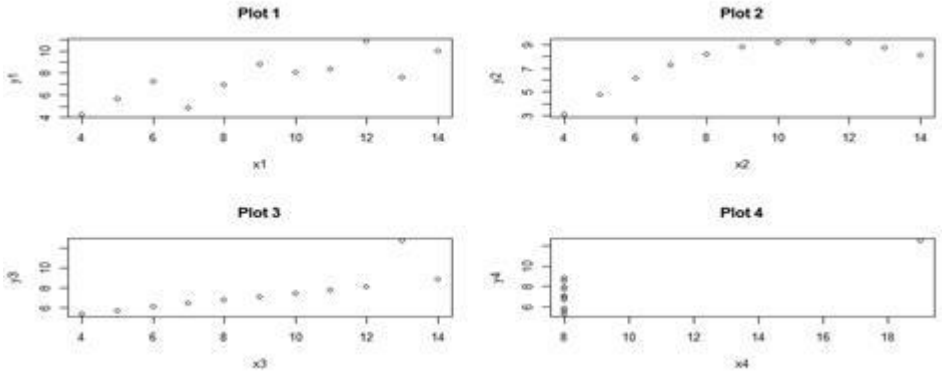

Start
Is the data text or time series?
Yes
See Figure 2
No
Predicting a Category?
No
See Figure 3
Yes
Does the data have labels?
No
See Clustering Figure 4
Yes
See Classification Figure 5

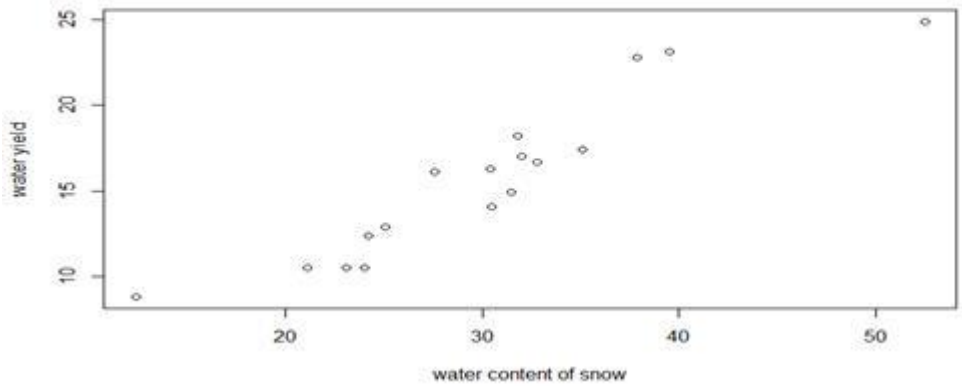

From Figure 1
Is the data text?
No
Is the time series univariate?
Yes
Yes
No
Text-Mining
• Word Frequency
• Wordcloud
• Topic Models (Latent Dirichlet Analysis
• Lexical Analysis, e.g. Polarity, Dispersion
Univariate Forecasting
• Auto-Regressive Integrated Moving Average (ARIMA)
• Exponential Smoothing
• Linear Regression
• ARCH
• GARCH
Multivariate Modeling
• Dynamic Linear Regression
• Vector Autoregression (VAR)
• Granger Causality
• Vector Error Correction Model (VECM)
• Cointegration

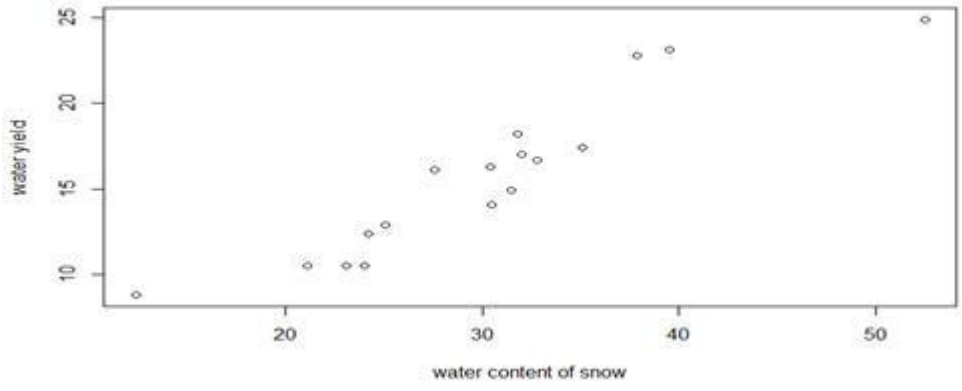

From Figure 1
Making Recommendations?
No
Looking for Associations?
Yes
No
Yes
Recommendation Engine
• User-Based Collaborative Filtering (UBCF)
• Item-Based Collaborative Filtering (IBCF)
• Singular Value Decomposition (SVD)
• Principal Components Analysis (PCA)
Association Analysis (Market Basket)
• Apriori
• Equivalence Class Clustering with Bottom-up Lattice Traversal (ECLAT)
Predict a Quantity
• Linear Regression
• Least Absolute Shrinkage and Selection Operator (LASSO)
• Ridge Regression
• ElasticNet
• PCA Regression
• Generalized Additive Models (GAM)
• Partial Least Squares (PLS)
From Figure 1
Clustering
• Hierarchical
• K-Means
• Partition Around Medoids
• Self-Organizing Map (SOM)
• Fuzzy Clustering

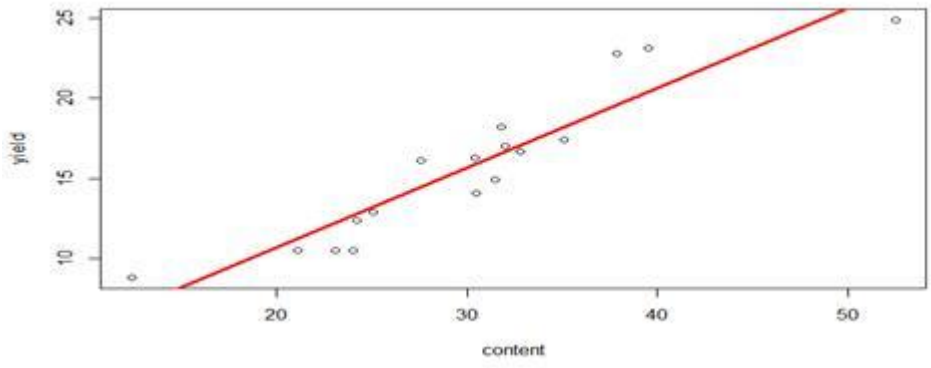


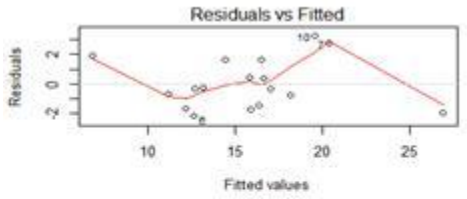
Chapter 2:Linear Regression – The Blocking and Tackling of Machine Learning

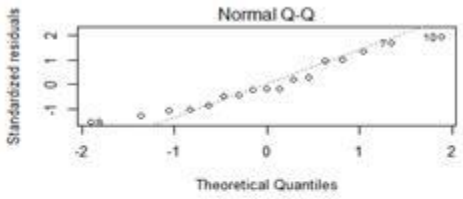


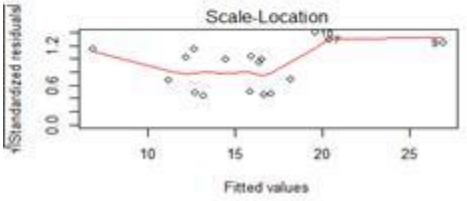

water yield
10
15
20
25
20
30
40
50
water content of snow

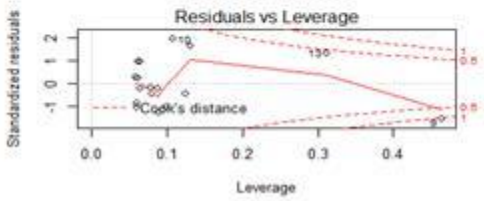

water yield
10
15
20
25
20
30
40
50
water content of snow

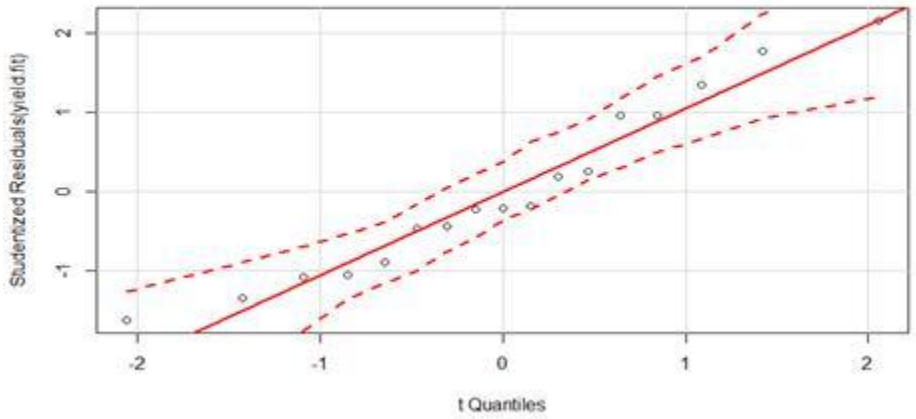

yield
10
15
20
25
20
30
40
50
content

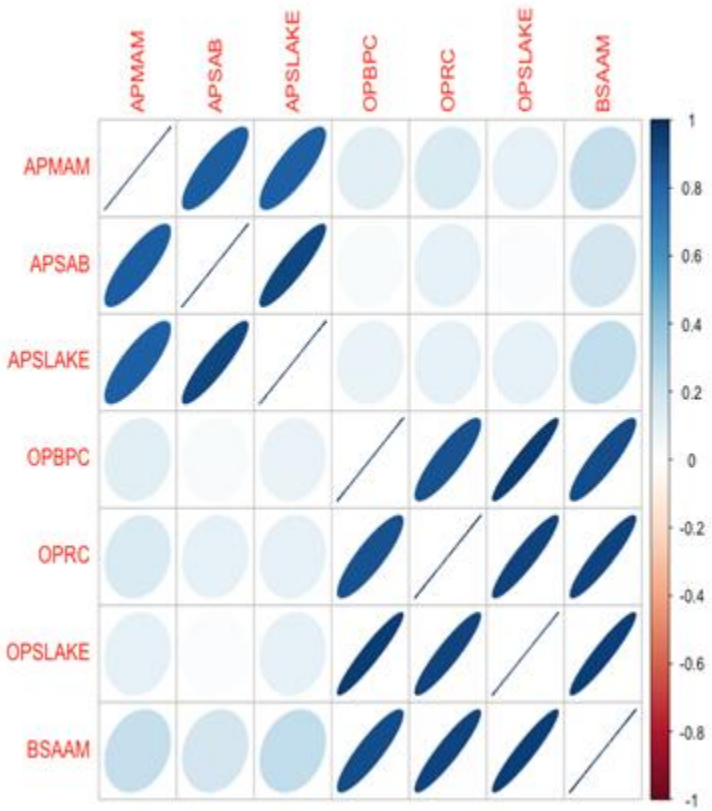

Residuals vs Fitted
Residuals
Fitted values

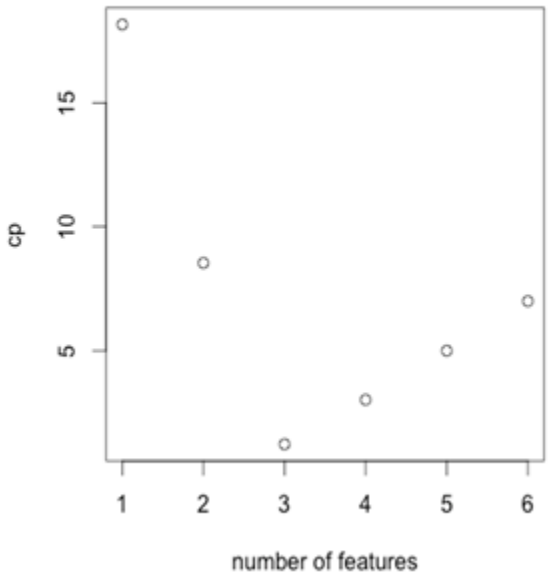

Normal Q-Q
Standardized residuals
Theoretical Quantiles

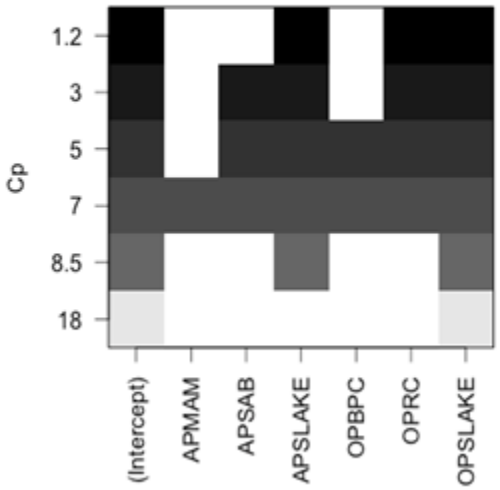

Scale-Location
√|Standardized residuals|
Fitted values

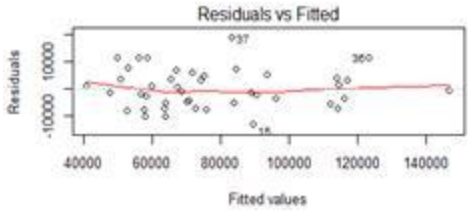

Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage

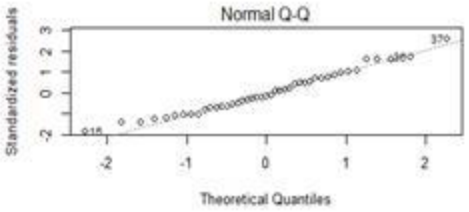

Studentized Residuals(yield.fit)
t Quantiles

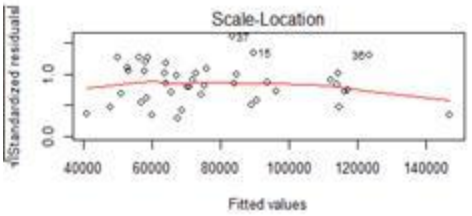

APMAM
APSAB
APSLAKE
OPBPC
OPRC
OPSLAKE
BSAAM
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

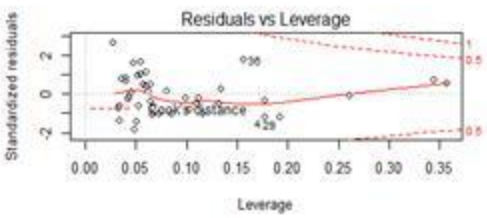

cp
15
10
5
1
2
3
4
5
6
number of features

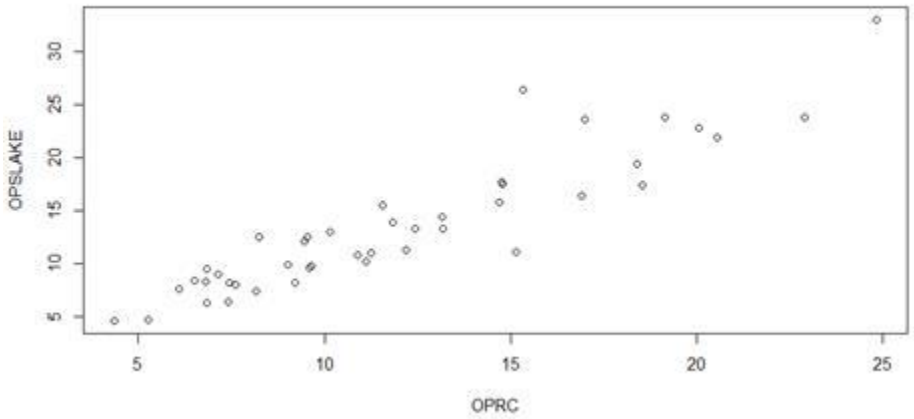

Cp
1.2
3
5
7
8.5
18
(Intercept)
APMAM
APSAB
APSLAKE
OPBPC
OPRC
OPSLAKE

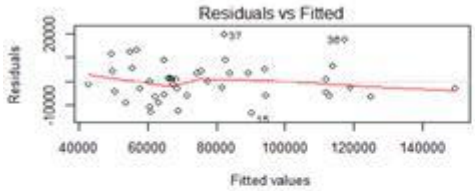

Residuals vs Fitted
Residuals
Fitted values
37
360
15

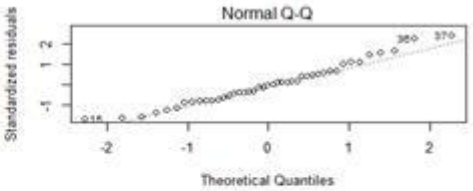

Normal Q-Q
Standardized residuals
Theoretical Quantiles
370
15

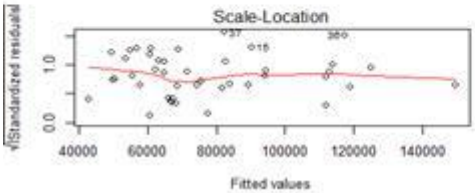

Scale-Location
√|Standardized residuals|
Fitted values
37
15
360

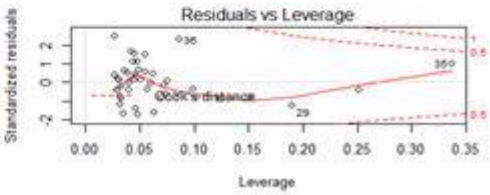

Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage
36
29

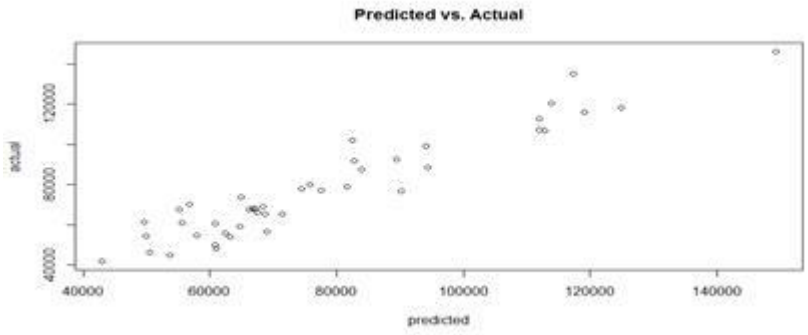

OPSLAKE
OPRC

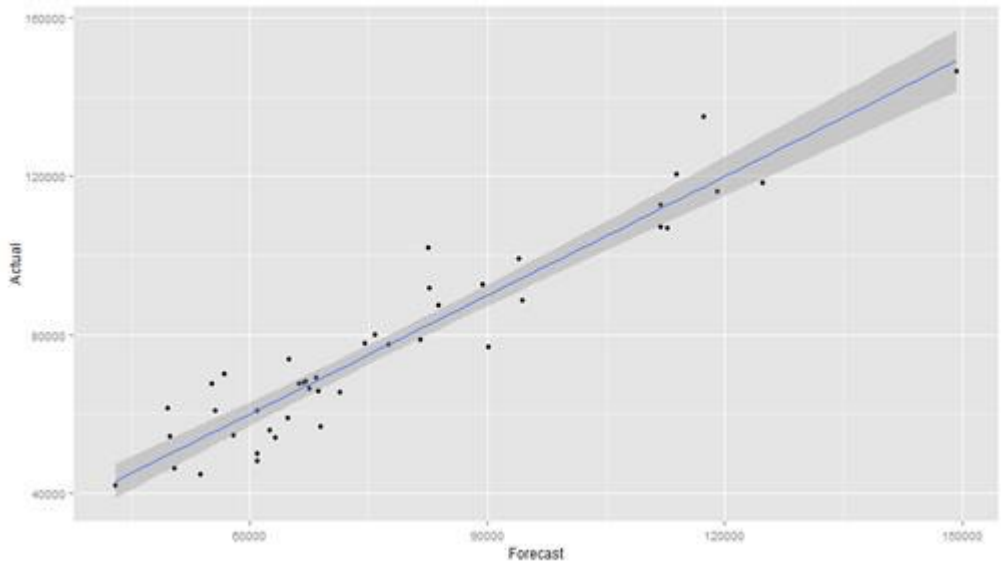

Residuals vs Fitted
Residuals
Fitted values
37
360
15


Normal Q-Q
Standardized residuals
Theoretical Quantiles
360
370
15

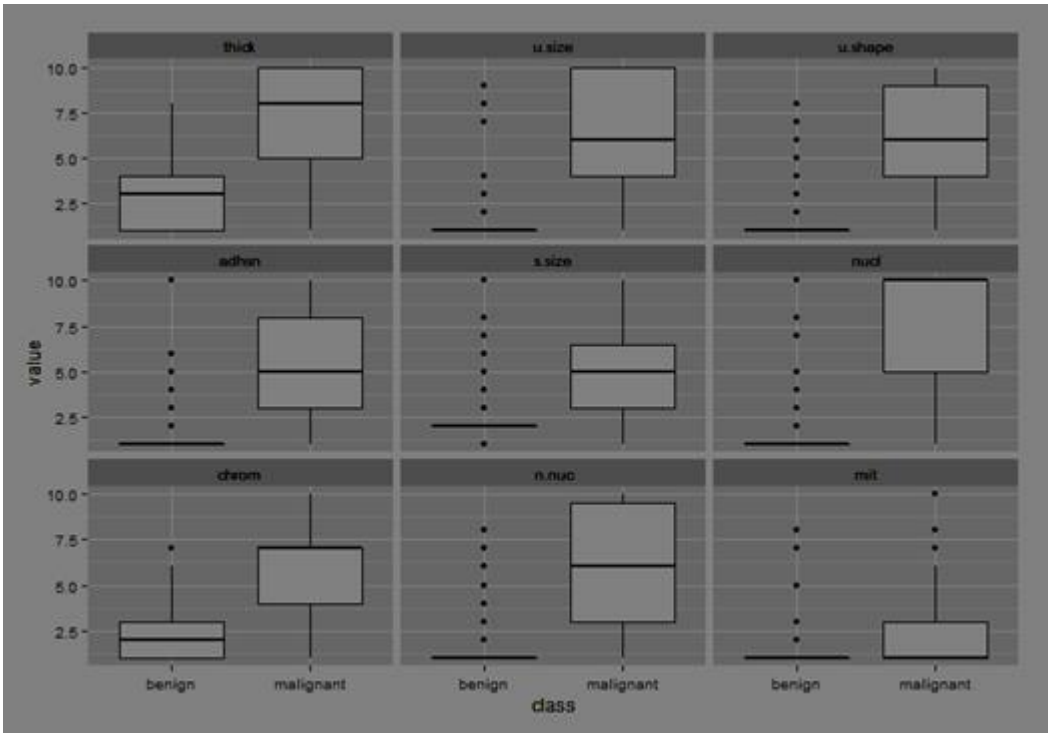

Scale-Location
√|Standardized residuals|
Fitted values
37
15
360

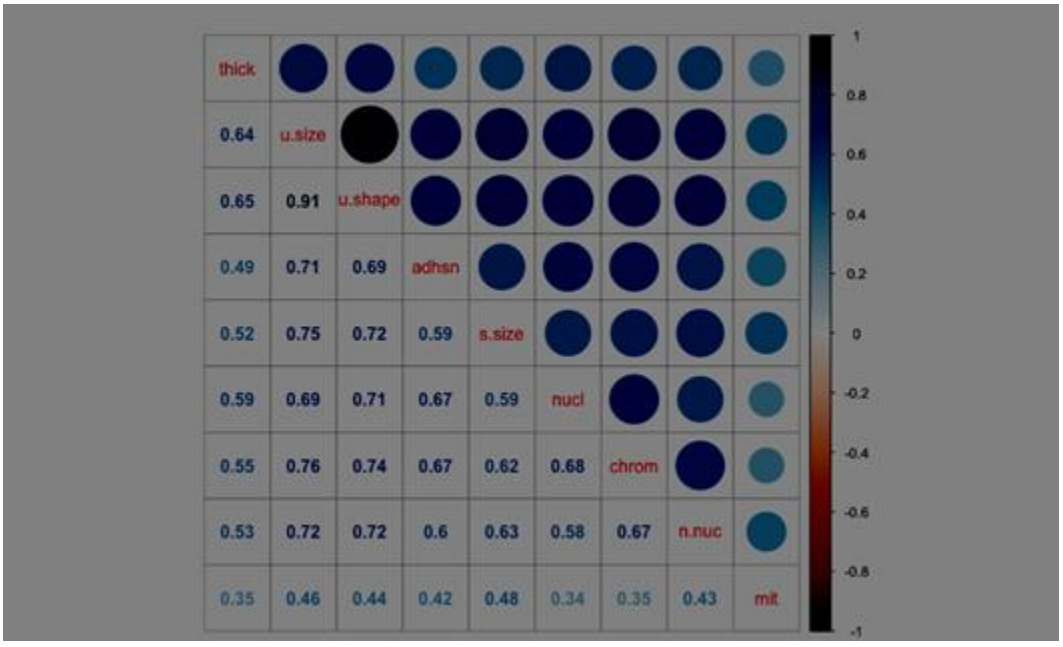

Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage
36
360
29

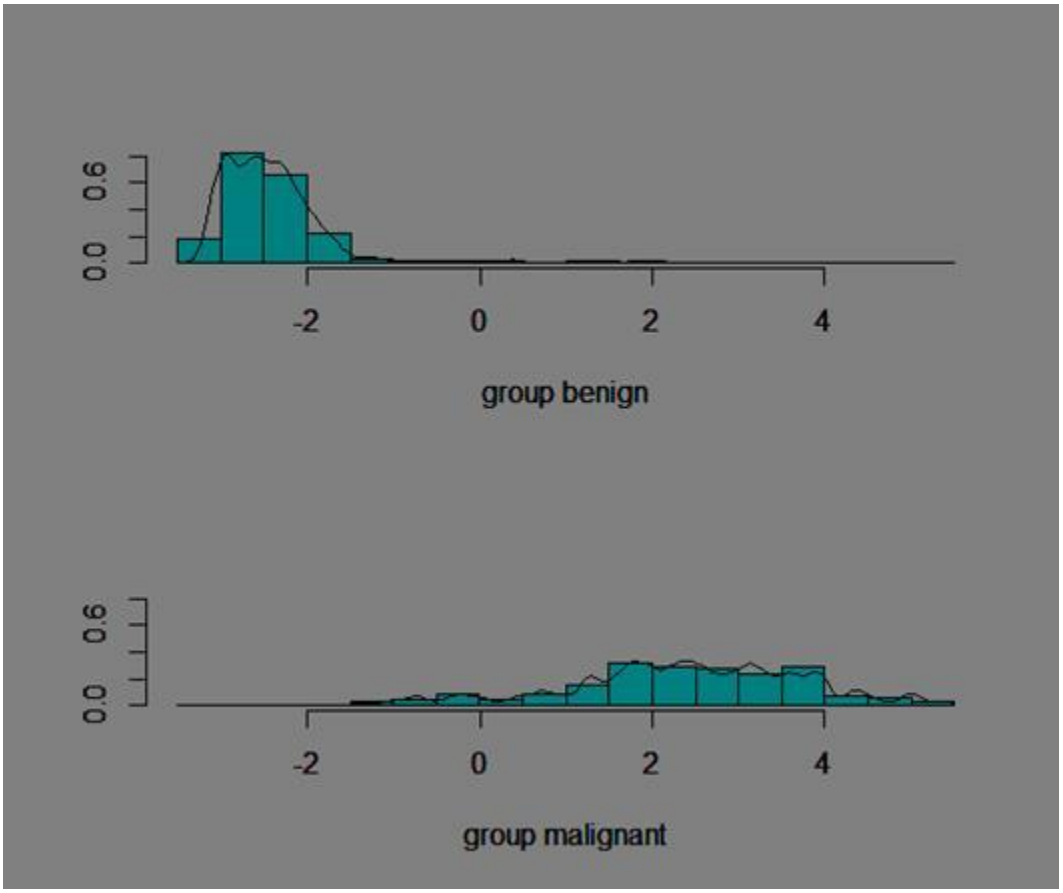

Predicted vs. Actual
actual
predicted
40000
60000
80000
100000
120000
140000

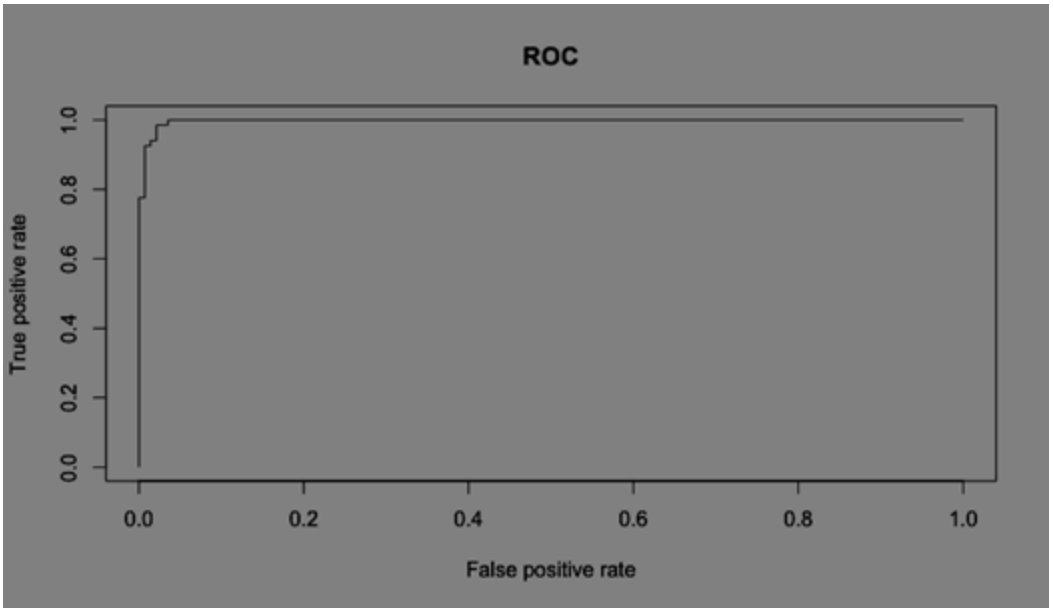

Actual
Forecast
40000
80000
120000
160000

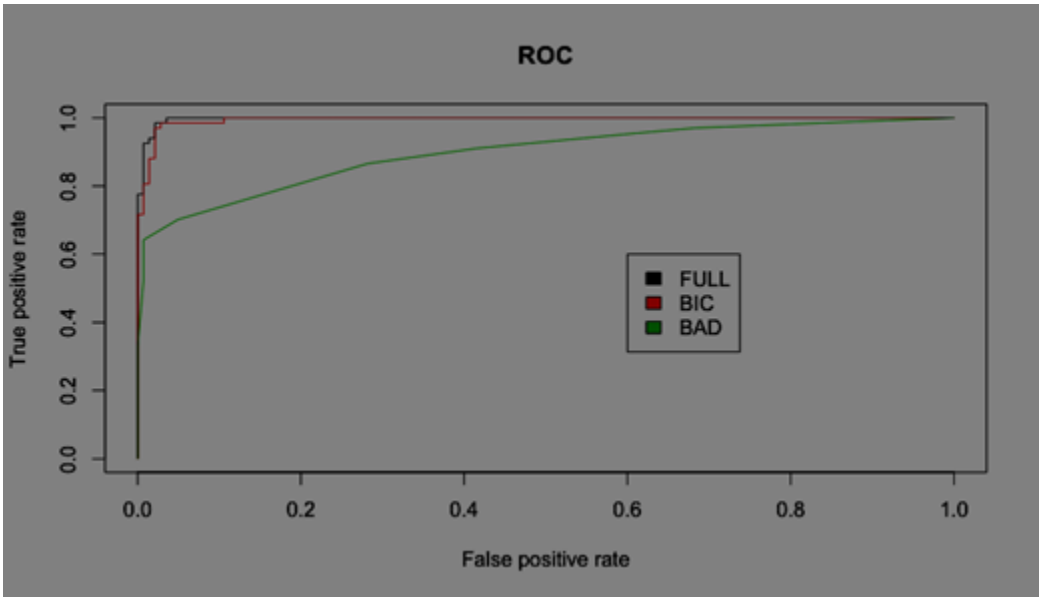
Chapter 3: Logistic Regression and Discriminant Analysis



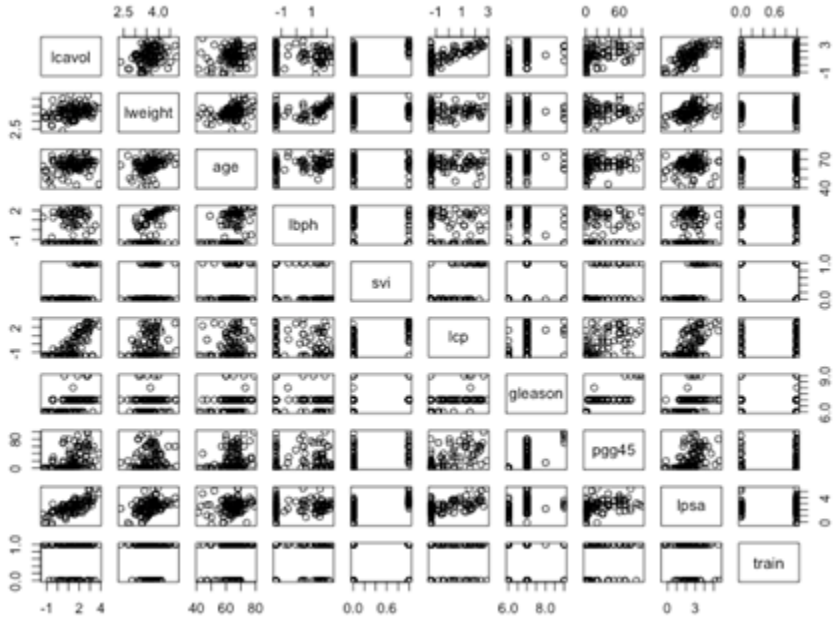


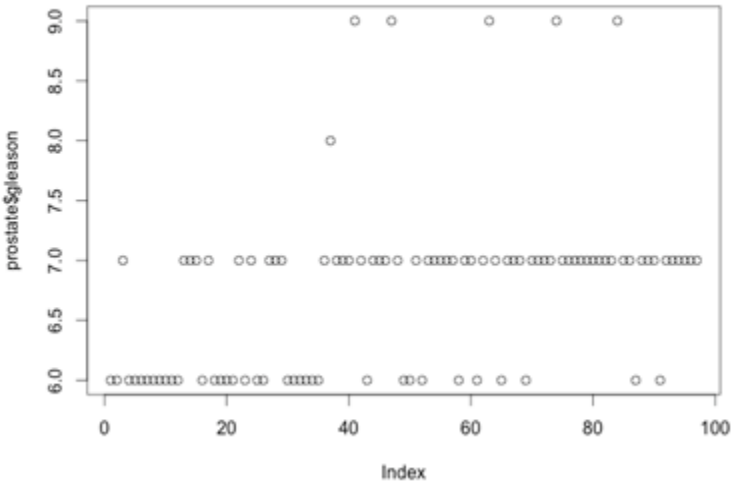

0.0
0.6
-2
0
2
4
group benign
0.0
0.6
-2
0
2
4
group malignant

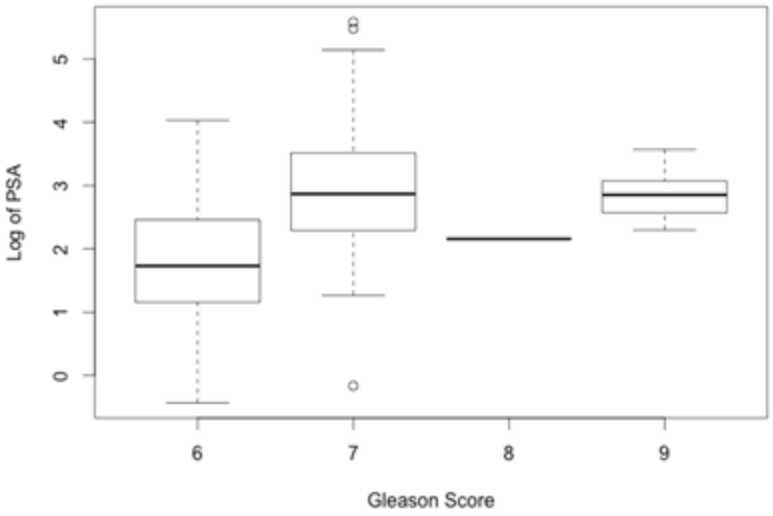

ROC
True positive rate
0.0
0.2
0.4
0.6
0.8
1.0
False positive rate

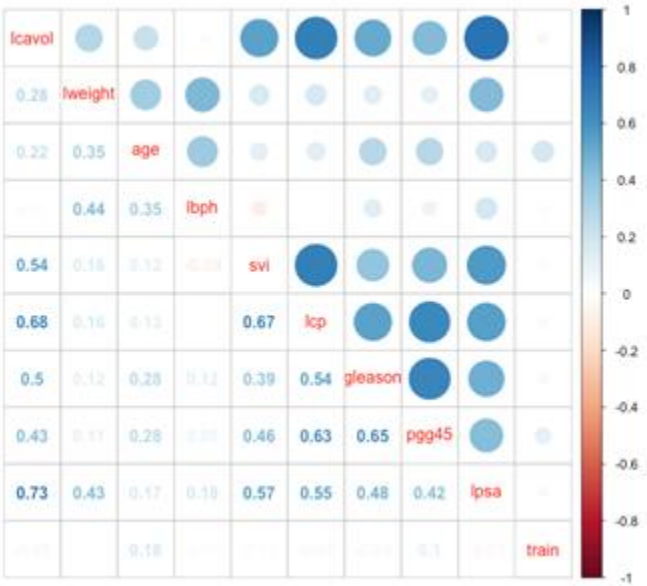

ROC
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
FULL
BIC
BAD

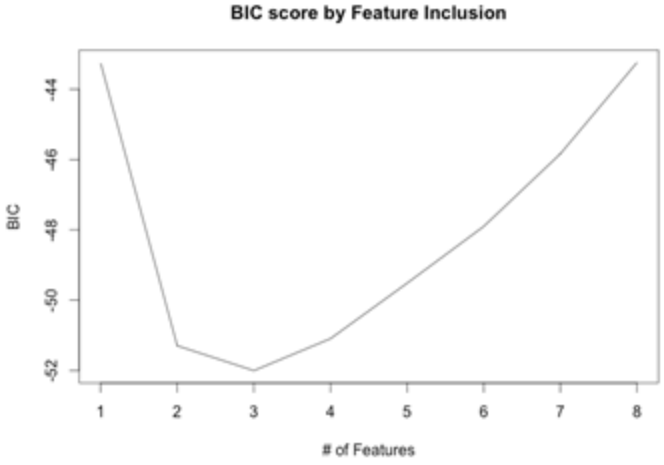
Chapter 4: Advanced Feature Selection in Linear Models

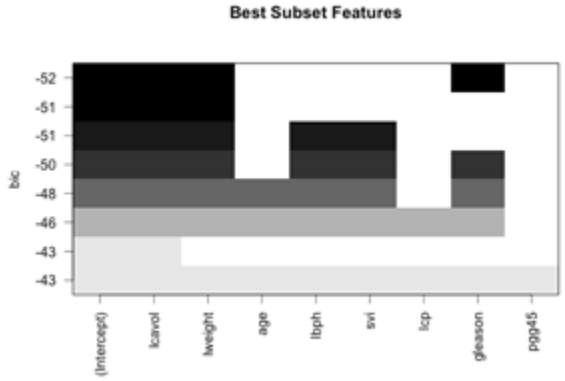


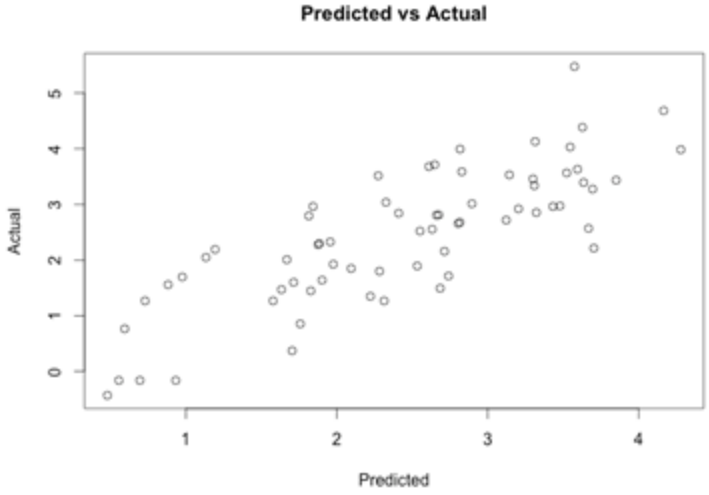


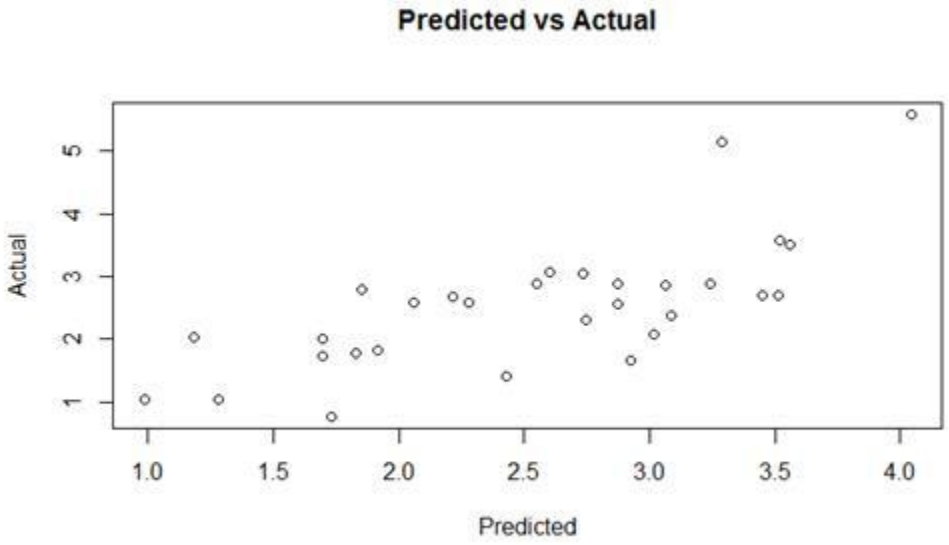

Log of PSA
0
1
2
3
4
5
6
7
8
9
Gleason Score

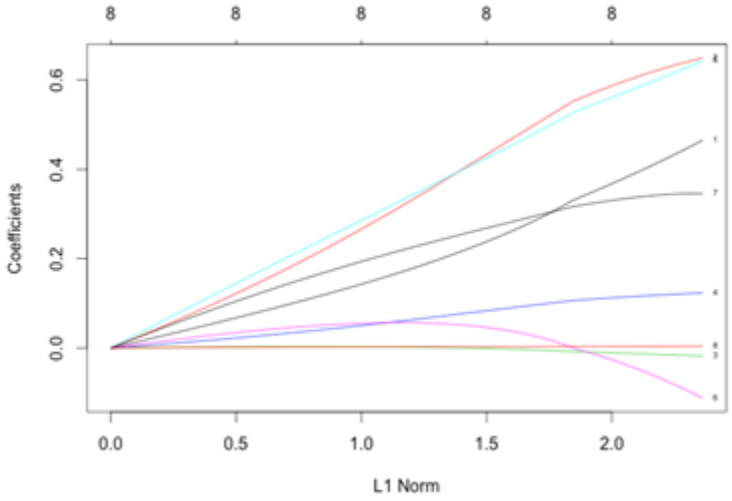

lcavol
lweight
age
lbph
svi
lcp
gleason
pgg45
lpsa
train
0.28
0.22
0.35
0.44
0.35
0.54
0.68
0.67
0.5
0.54
0.43
0.46
0.63
0.65
0.73
0.43
0.57
0.55
0.48
0.42
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

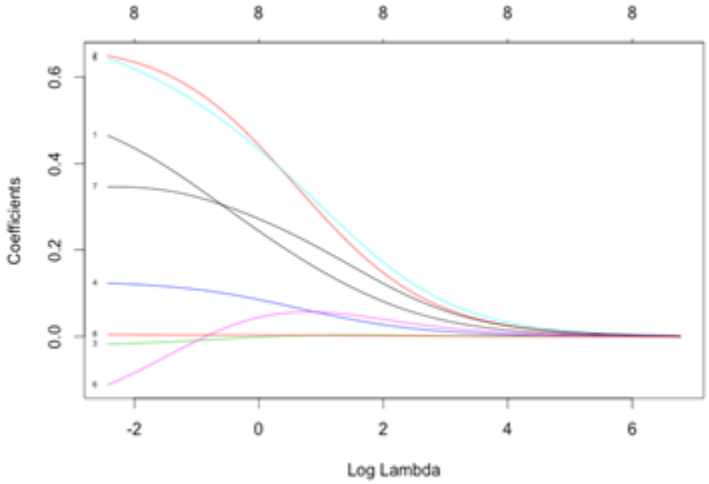

BIC score by Feature Inclusion
BIC
-44
-46
-48
-50
-52
1
2
3
4
5
6
7
8
# of Features

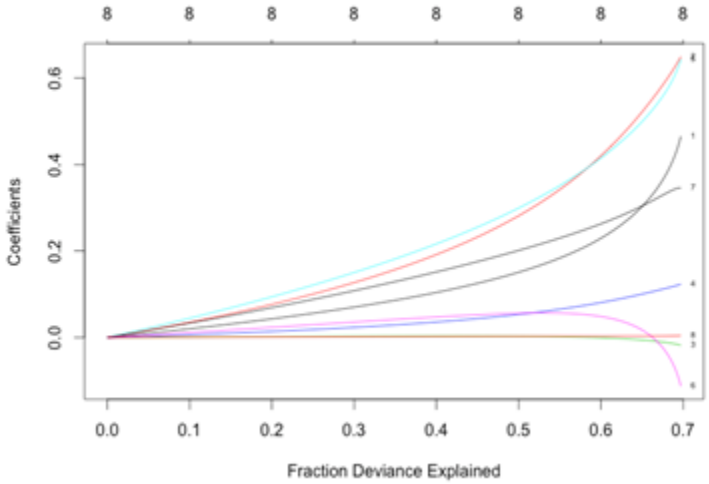

Best Subset Features
bic
-52
-51
-51
-50
-48
-46
-43
-43
(Intercept)
lcavol
lweight
age
lbph
svi
lcp
gleason
pgg45

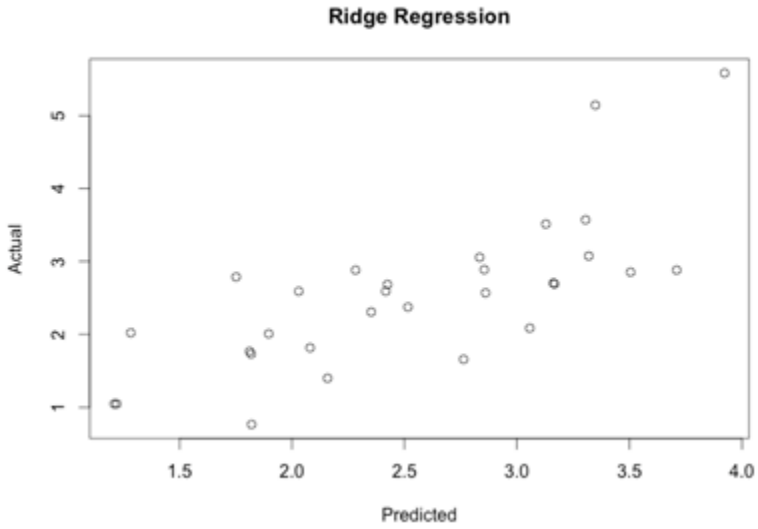

Predicted vs Actual
Actual
Predicted
0
1
2
3
4
5
1
2
3
4

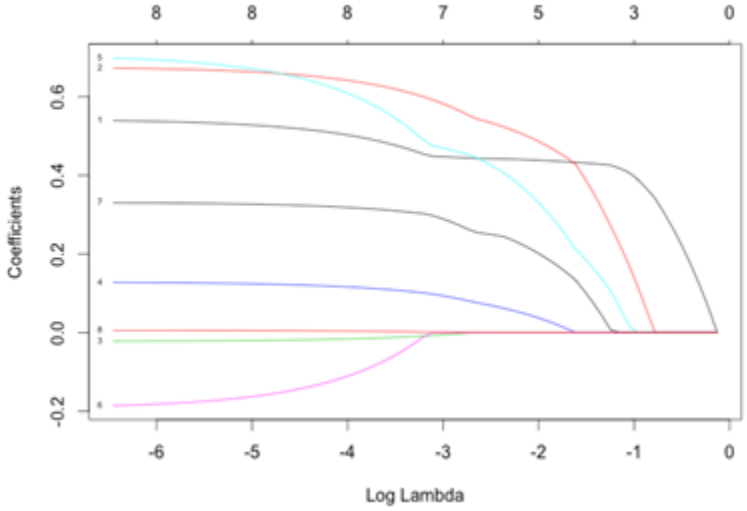

Predicted vs Actual
Actual
Predicted
1
2
3
4
5
1.0
1.5
2.0
2.5
3.0
3.5
4.0

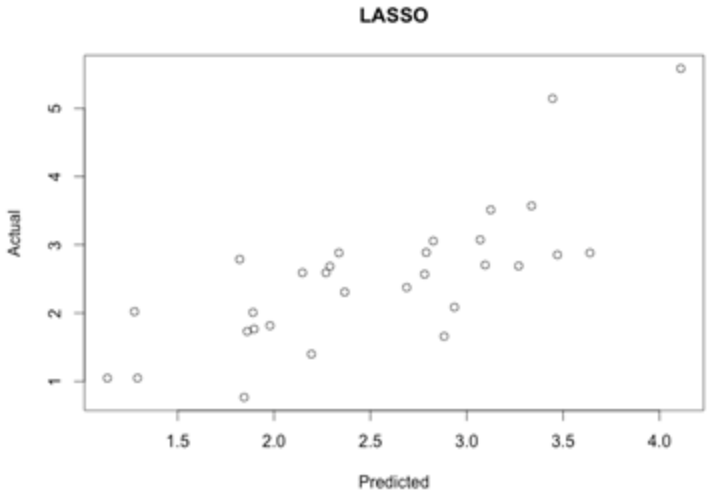

8 8 8 8 8
Coefficients
0.0 0.2 0.4 0.6
0.0 0.5 1.0 1.5 2.0
L1 Norm

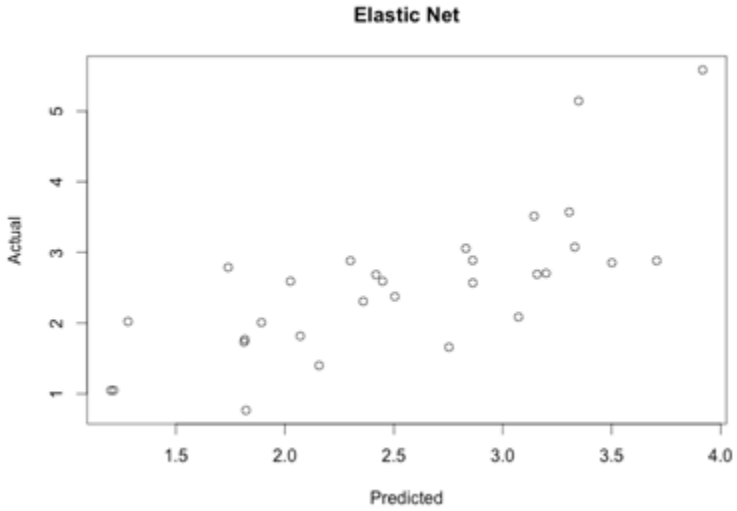

8 8 8 8 8
Coefficients
0.0 0.2 0.4 0.6
-2 0 2 4 6
Log Lambda

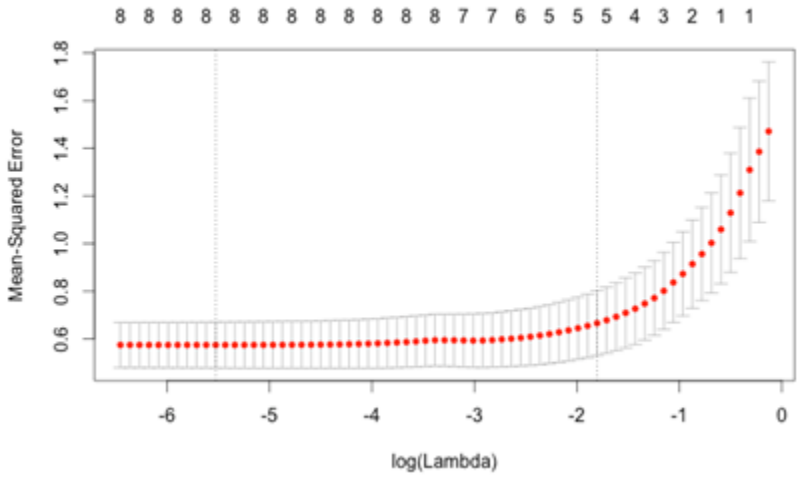

8 8 8 8 8 8 8 8
Coefficients
0.0 0.2 0.4 0.6
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7
Fraction Deviance Explained


Ridge Regression
Actual
Predicted
1
2
3
4
5
1.5
2.0
2.5
3.0
3.5
4.0

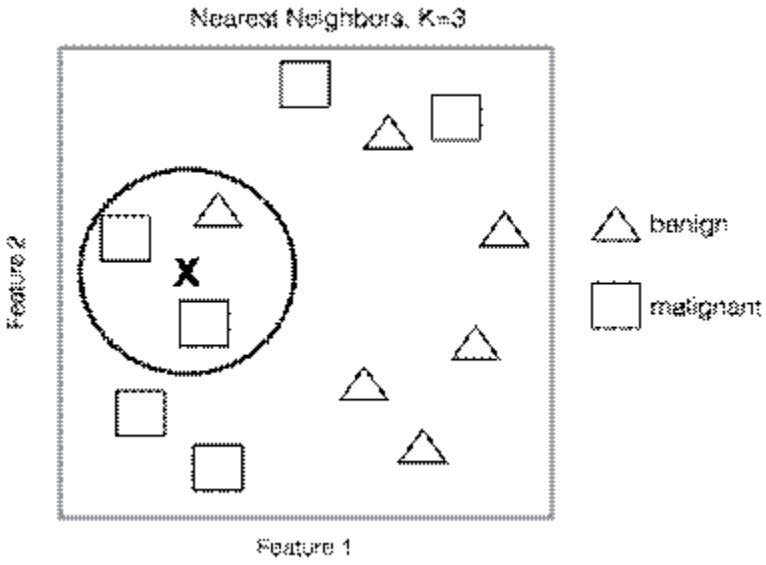

Coefficients
Log Lambda
8
8
8
7
5
3
0
-0.2
0.0
0.2
0.4
0.6
-6
-5
-4
-3
-2
-1
0

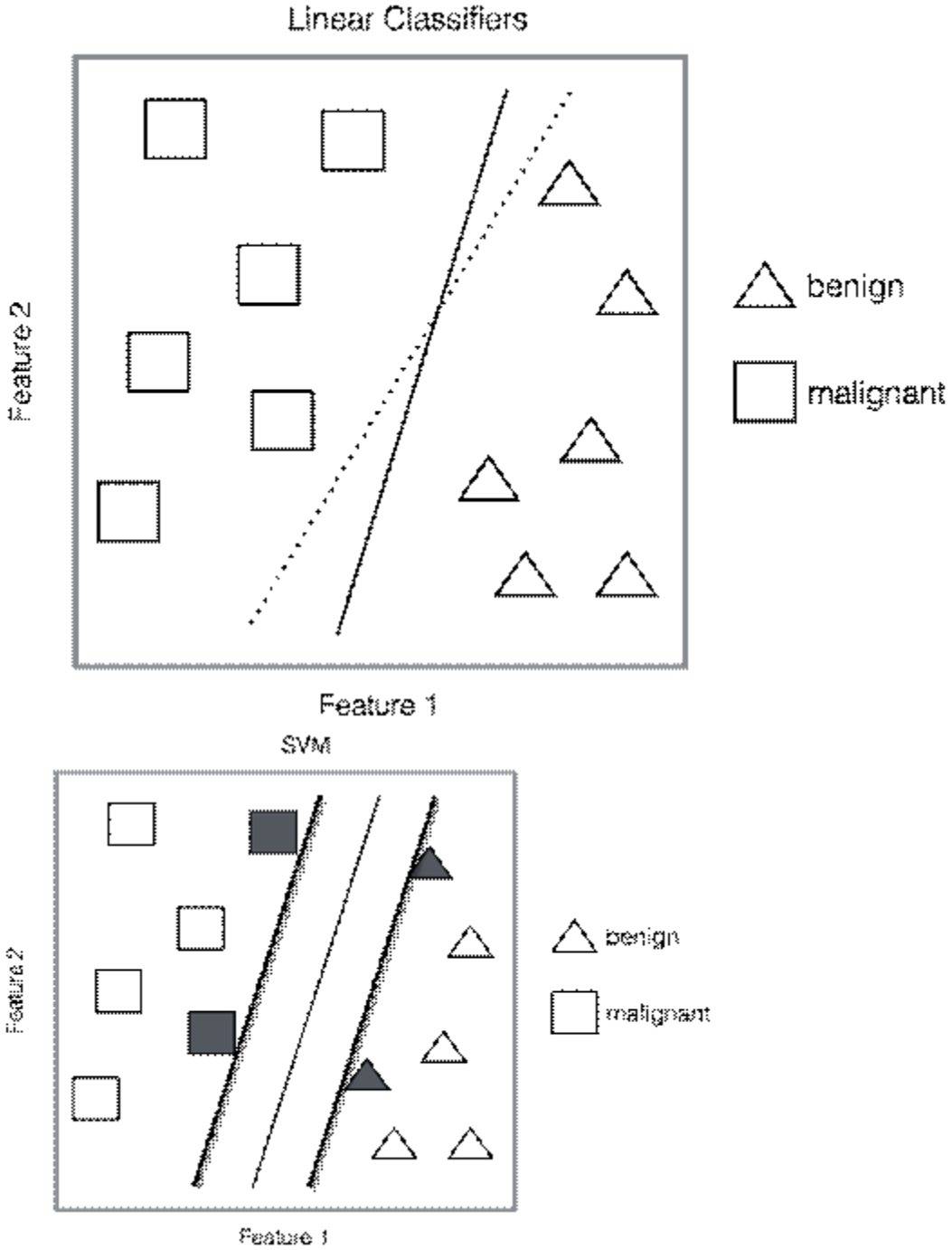

LASSO
Actual
Predicted
1
2
3
4
5
1.5
2.0
2.5
3.0
3.5
4.0

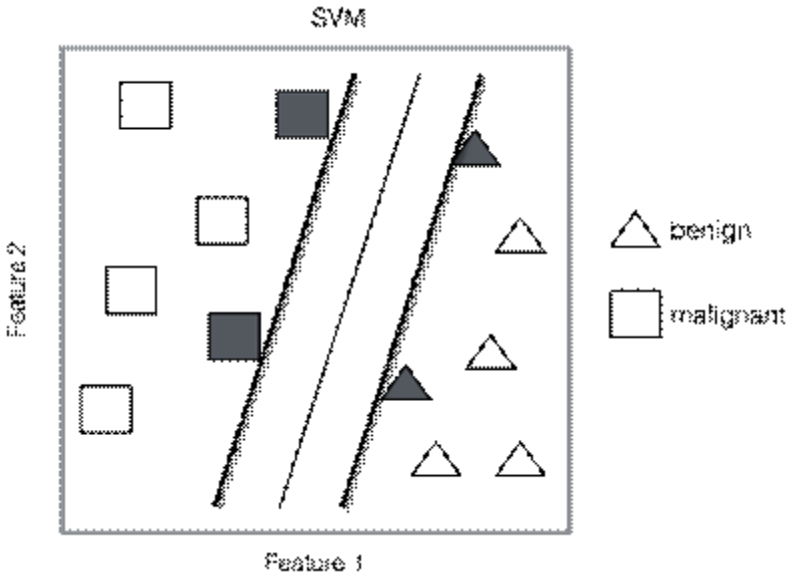

Elastic Net
Actual
Predicted
1
2
3
4
5
1.5
2.0
2.5
3.0
3.5
4.0

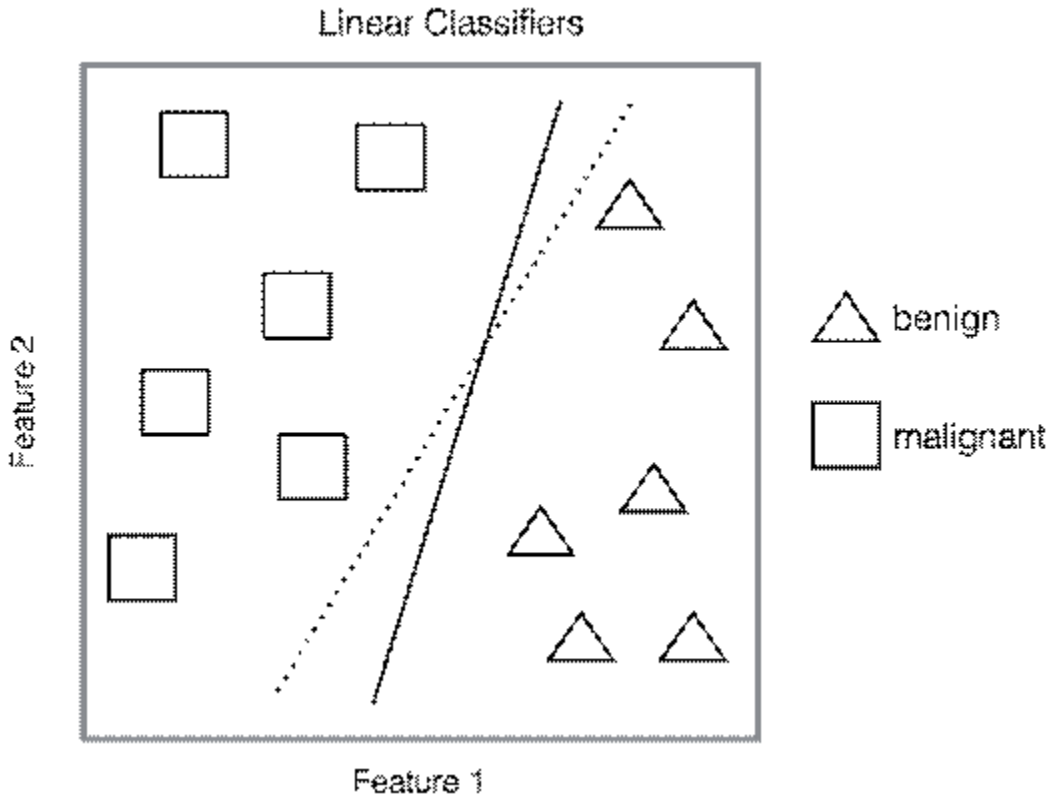

8 8 8 8 8 8 8 8 8 8 8 8 7 7 6 5 5 5 4 3 2 1 1
Mean-Squared Error
0.6
0.8
1.0
1.2
1.4
1.6
1.8
-6
-5
-4
-3
-2
-1
0
log(Lambda)

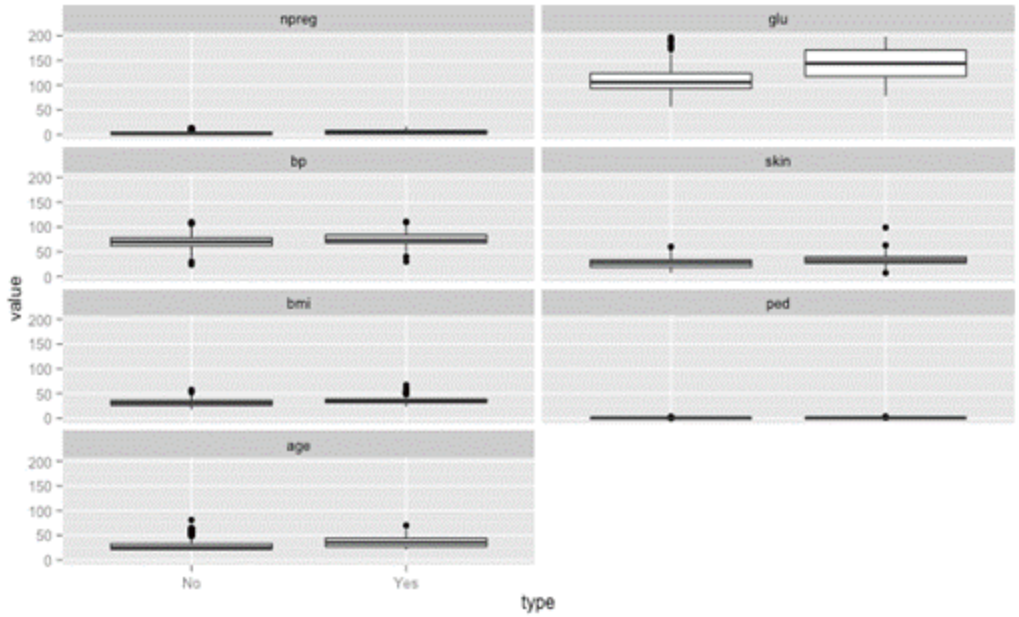
Chapter 5: More Classification Techniques – K-Nearest Neighbors and Support Vector Machines

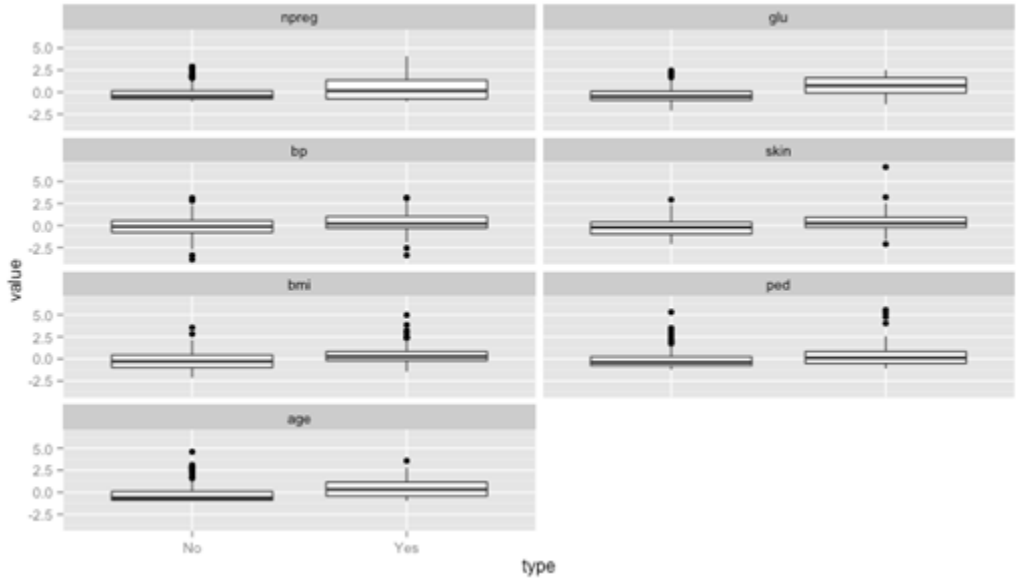


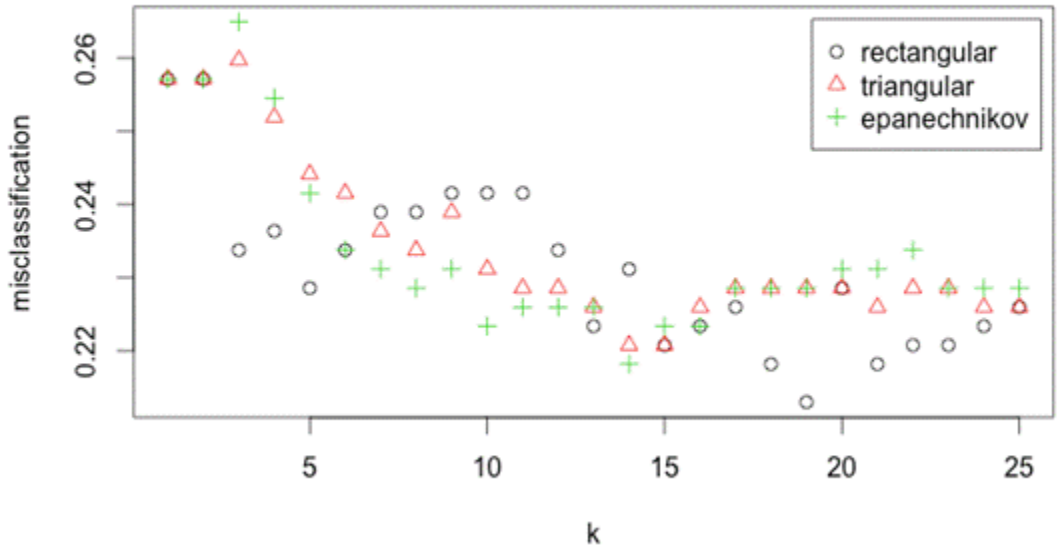

Linear Classifiers
Feature 2
Feature 1
benign
malignant
SVM
Feature 2
Feature 1
benign
malignant


SVM
Feature 2
Feature 1
benign
malignant

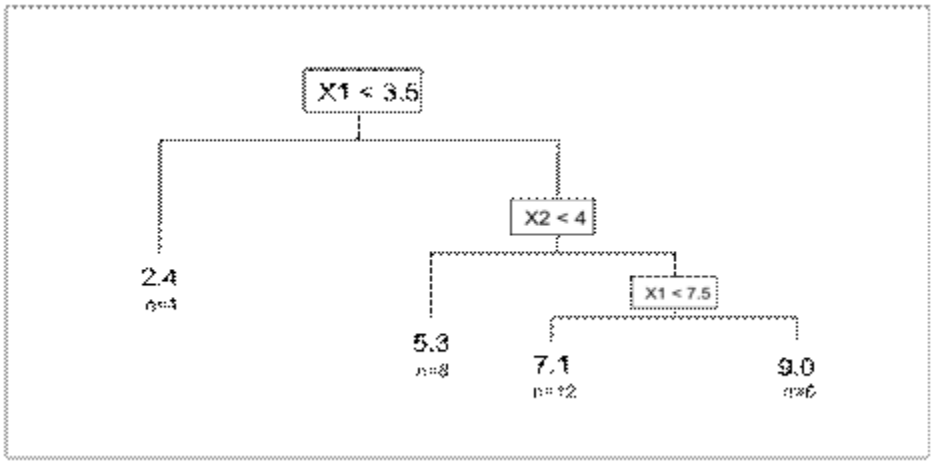

Linear Classifiers
Feature 2
Feature 1
benign
malignant

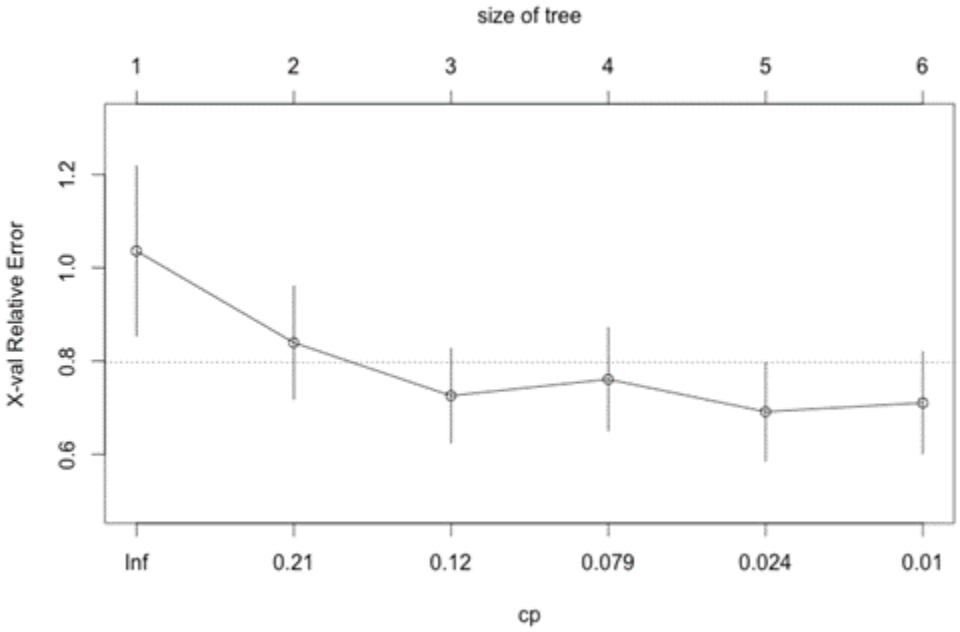

npreg
glu
bp
skin
bmi
ped
age
200
150
100
50
0
value
No
Yes
type

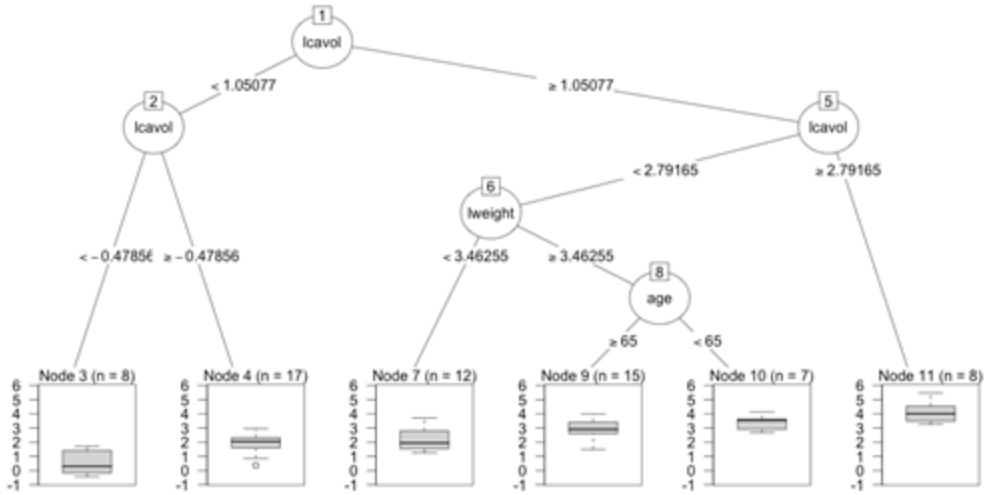

npreg
glu
bp
skin
bmi
ped
age
5.0
2.5
0.0
-2.5
value
No
Yes
type

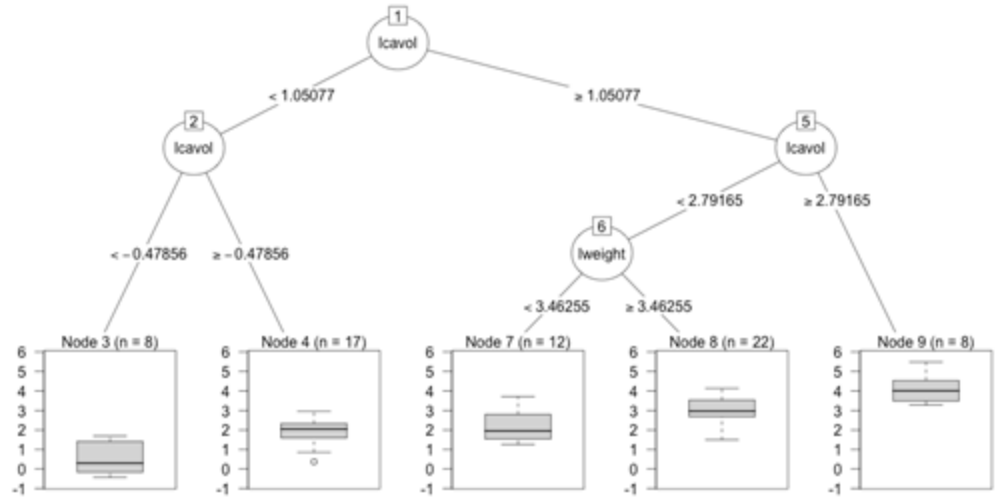


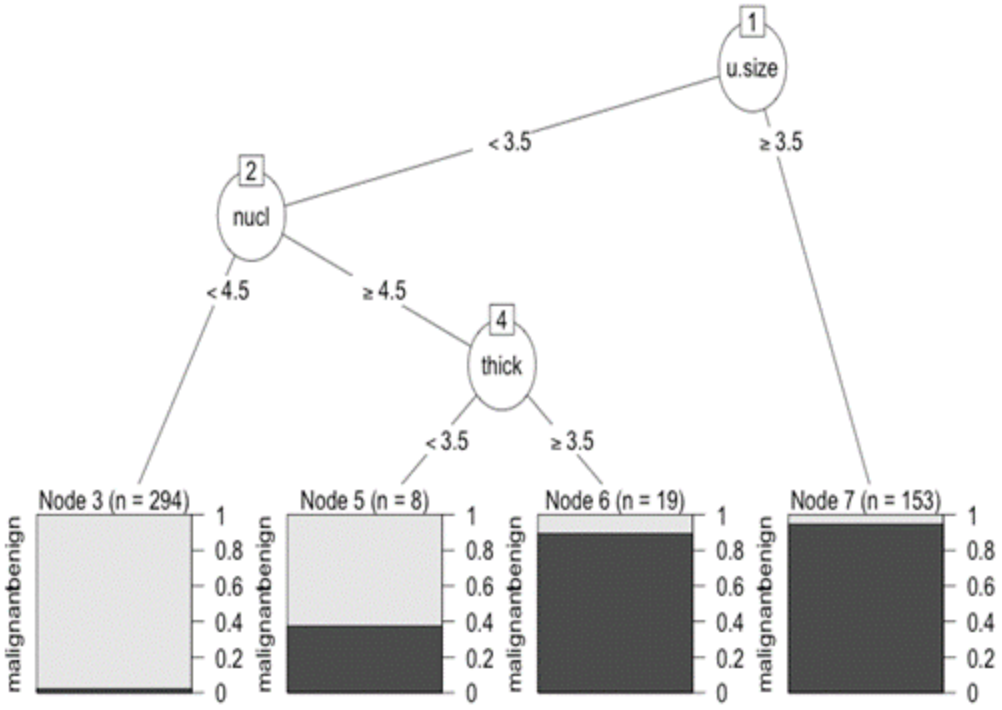
Chapter 6: Classification and Regression Trees

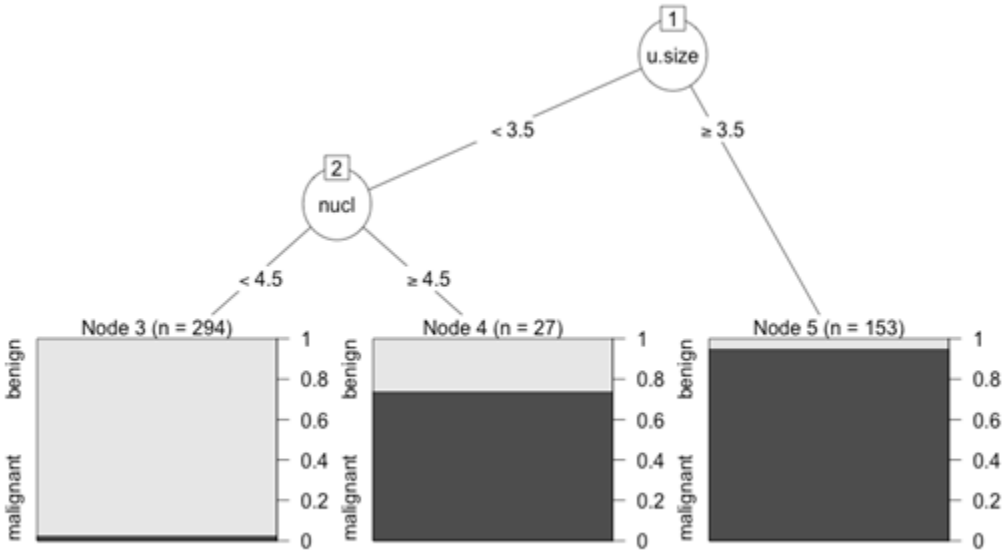


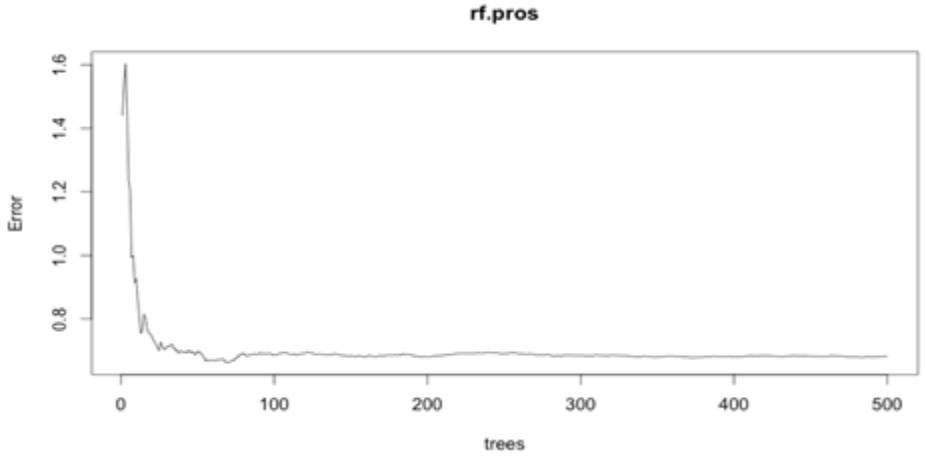

size of tree
1
2
3
4
5
6
X-val Relative Error
0.6
0.8
1.0
1.2
Inf
0.21
0.12
0.079
0.024
0.01
cp

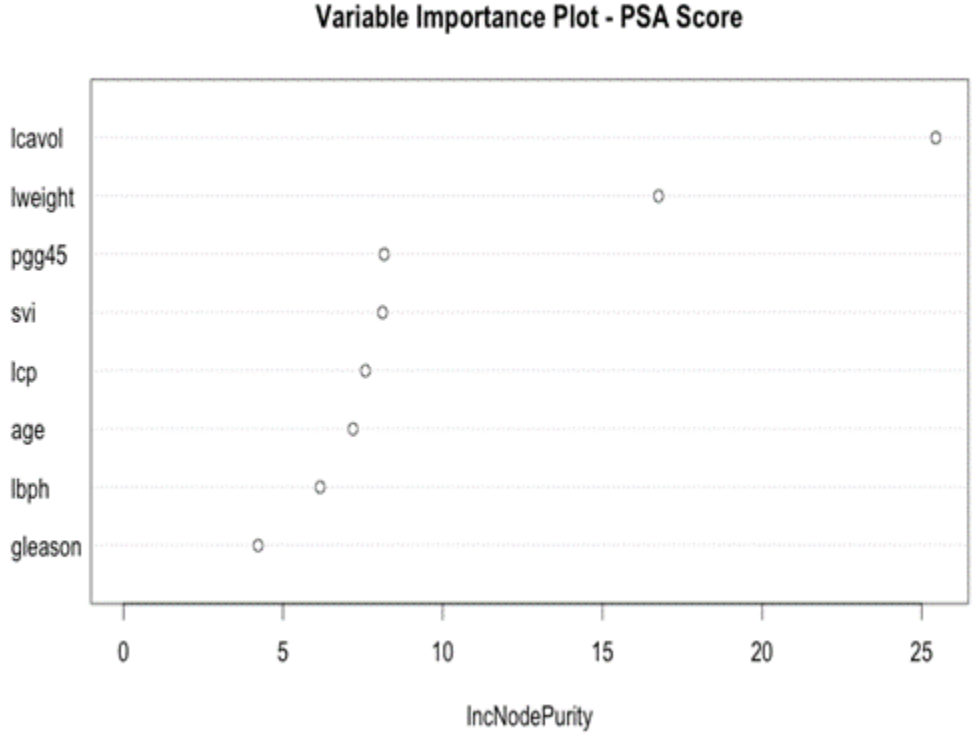

1
lcavol
< 1.05077
≥ 1.05077
2
lcavol
5
lcavol
< 2.79165
≥ 2.79165
6
lweight
< −0.47856
≥ −0.47856
< 3.46255
≥ 3.46255
8
age
≥ 65
< 65
Node 3 (n = 8)
Node 4 (n = 17)
Node 7 (n = 12)
Node 9 (n = 15)
Node 10 (n = 7)
Node 11 (n = 8)

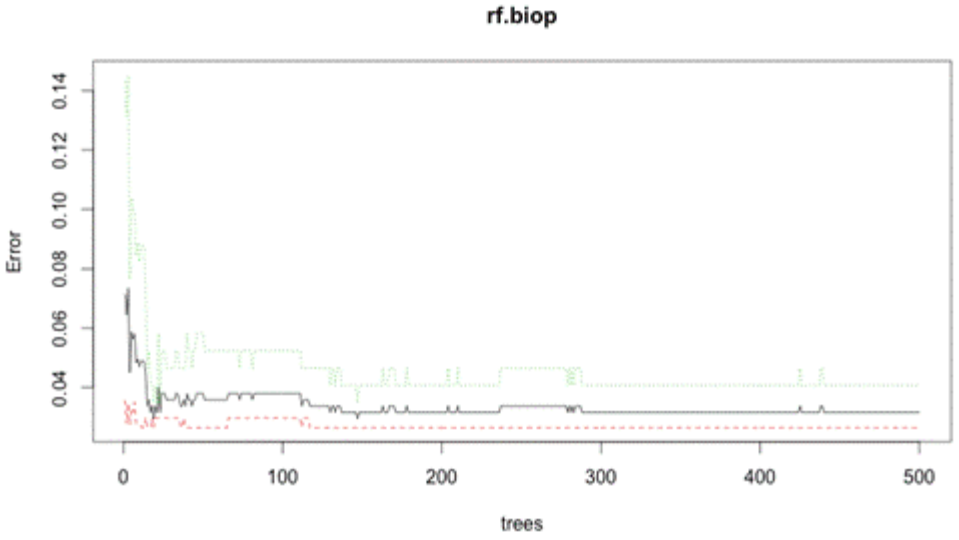

1
lcavol
< 1.05077
≥ 1.05077
2
lcavol
5
lcavol
< −0.47856
≥ −0.47856
< 2.79165
≥ 2.79165
6
lweight
< 3.46255
≥ 3.46255
Node 3 (n = 8)
Node 4 (n = 17)
Node 7 (n = 12)
Node 8 (n = 22)
Node 9 (n = 8)

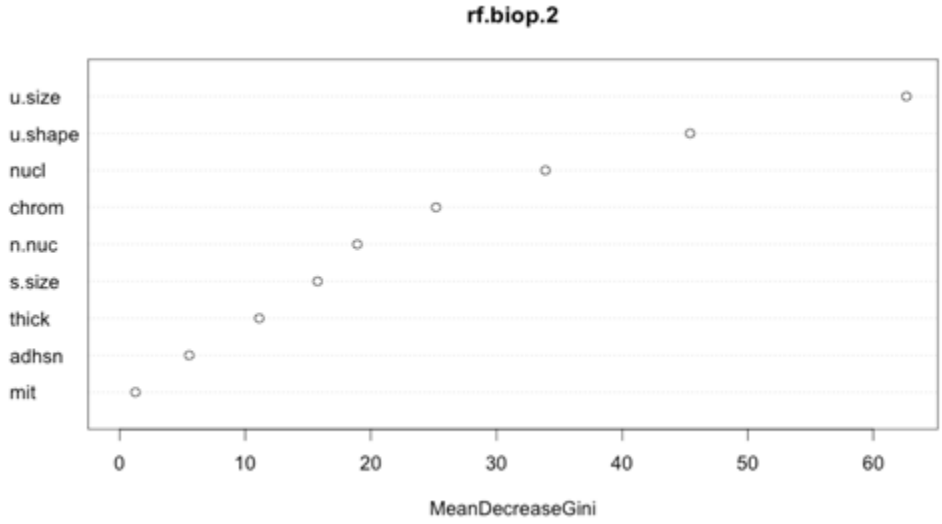

1
u.size
< 3.5
≥ 3.5
2
nucl
< 4.5
≥ 4.5
4
thick
< 3.5
≥ 3.5
Node 3 (n = 294)
Node 5 (n = 8)
Node 6 (n = 19)
Node 7 (n = 153)
malignant
benign

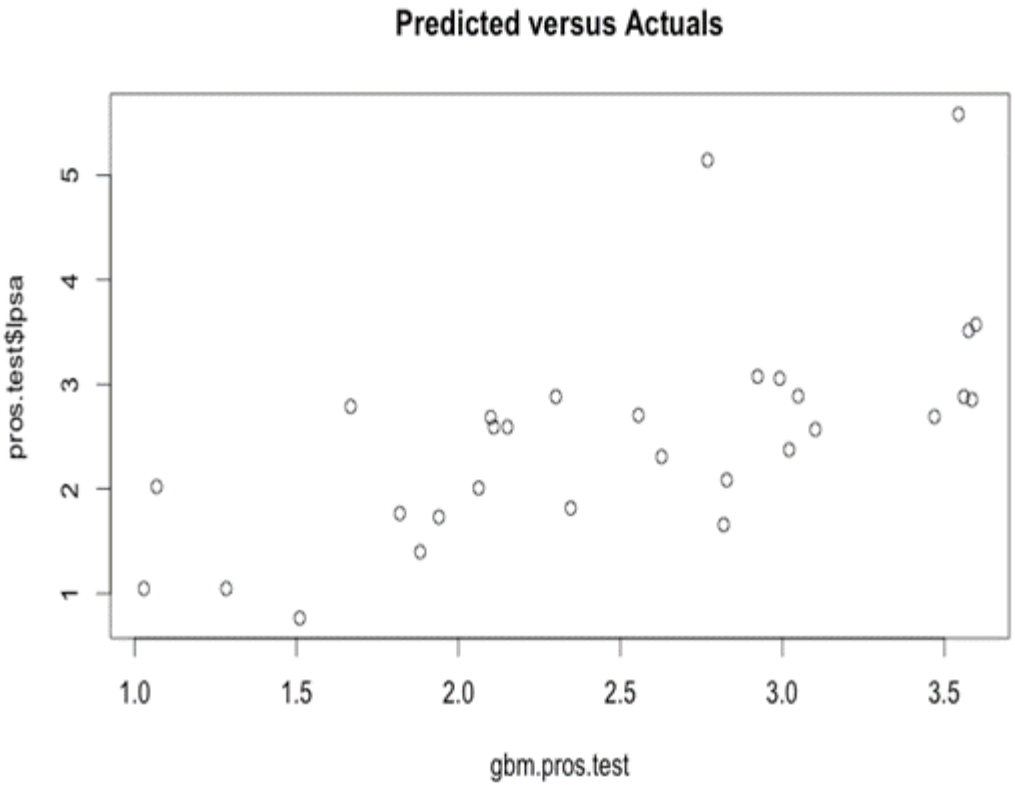

1
u.size
< 3.5
≥ 3.5
2
nucl
< 4.5
≥ 4.5
Node 3 (n = 294)
Node 4 (n = 27)
Node 5 (n = 153)
benign
malignant
1
0.8
0.6
0.4
0.2
0

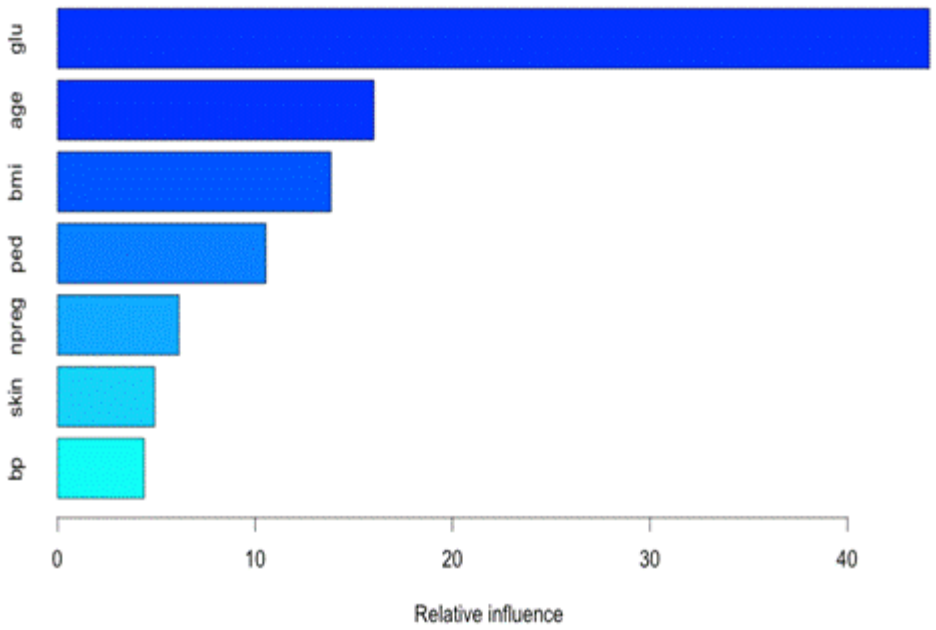

rf.pros
Error
0.8
1.0
1.2
1.4
1.6
0
100
200
300
400
500
trees


Variable Importance Plot - PSA Score
lcavol
lweight
pgg45
svi
lcp
age
lbph
gleason
0
5
10
15
20
25
IncNodePurity

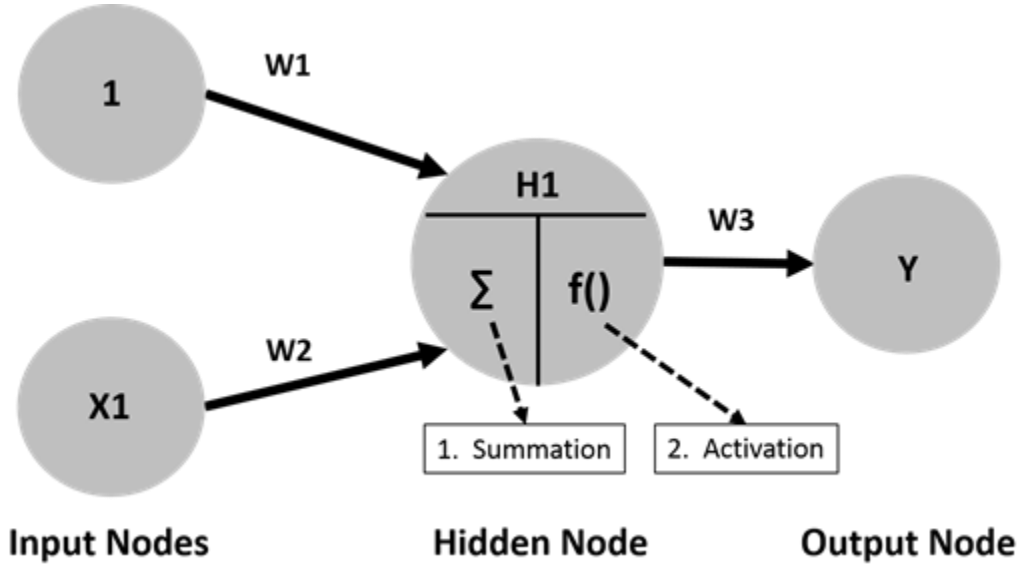

rf.biop
Error
0.04
0.06
0.08
0.10
0.12
0.14
0
100
200
300
400
500
trees

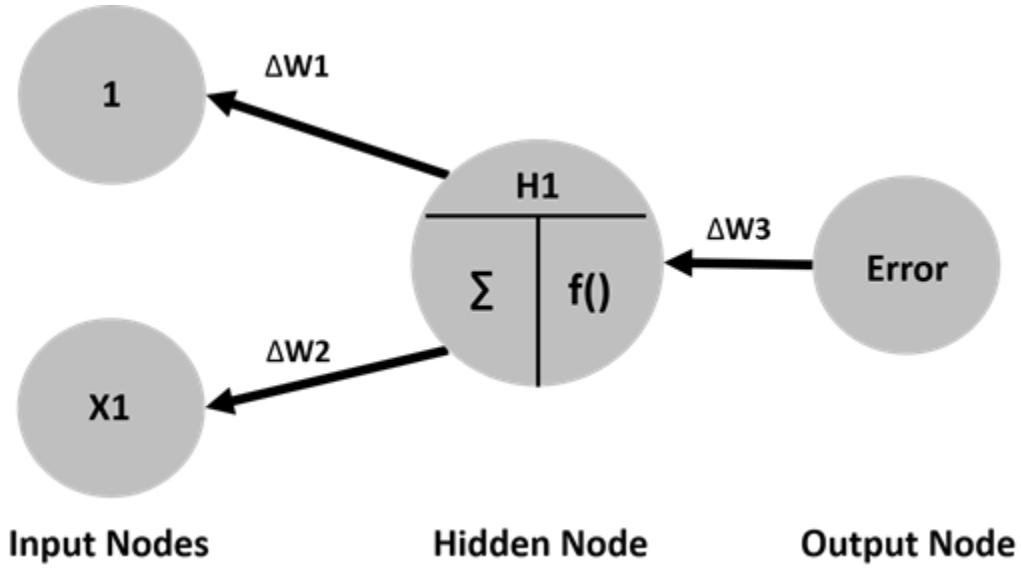

rf.biop.2
u.size
u.shape
nucl
chrom
n.nuc
s.size
thick
adhsn
mit
0
10
20
30
40
50
60
MeanDecreaseGini

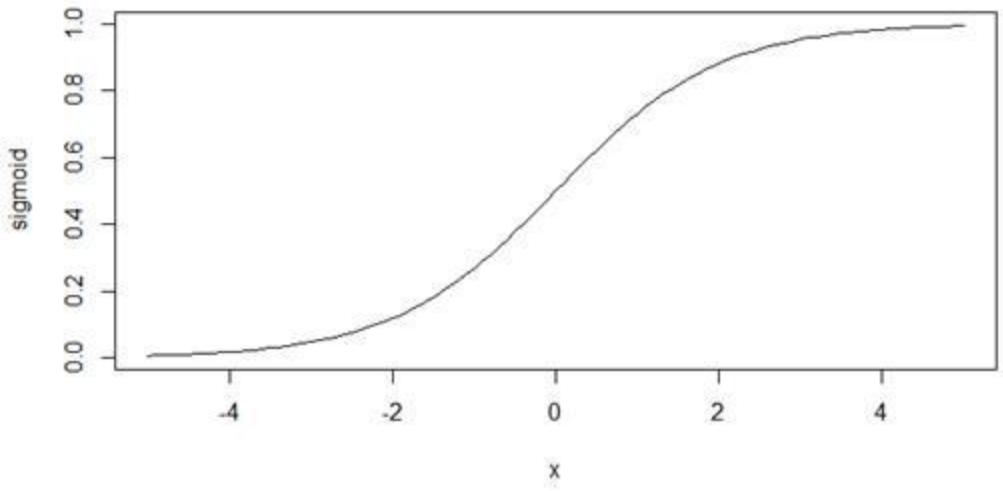

Predicted versus Actuals
pros.test$lpsa
1
2
3
4
5
1.0
1.5
2.0
2.5
3.0
3.5
gbm.pros.test

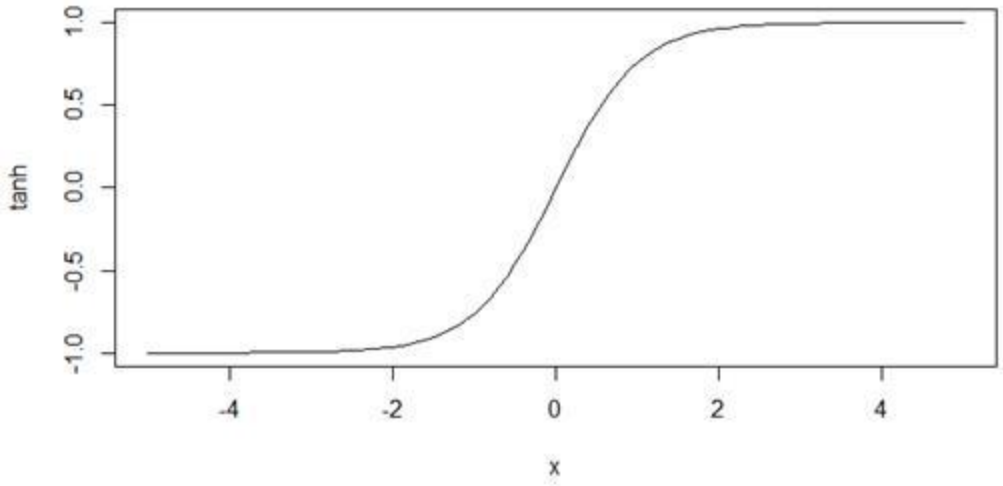

glu
age
bmi
ped
npreg
skin
bp
0
10
20
30
40
Relative influence

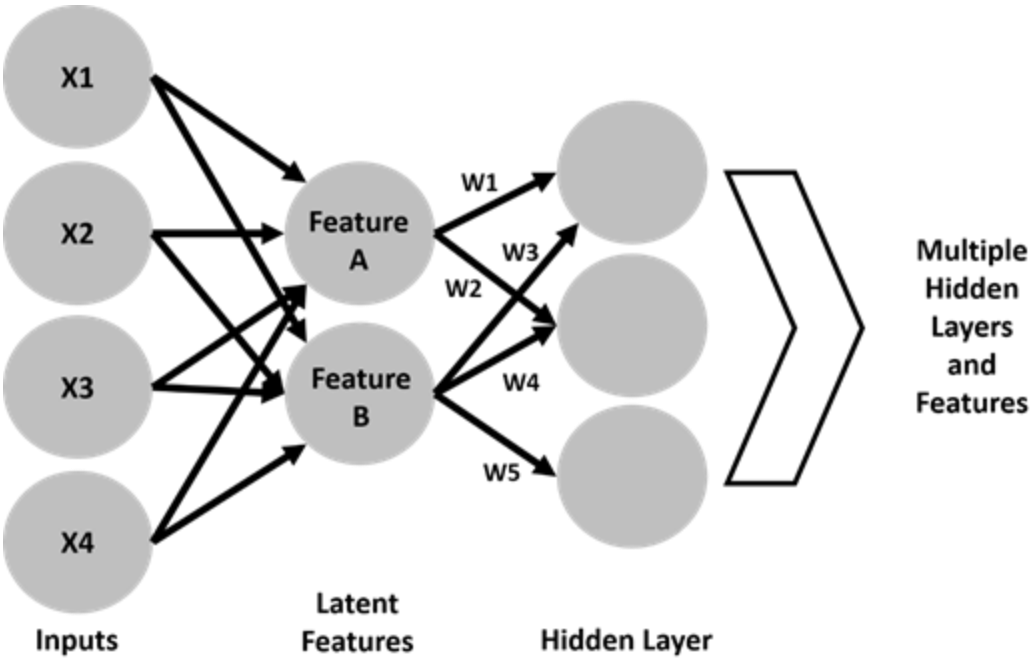
Chapter 7: Neural Networks

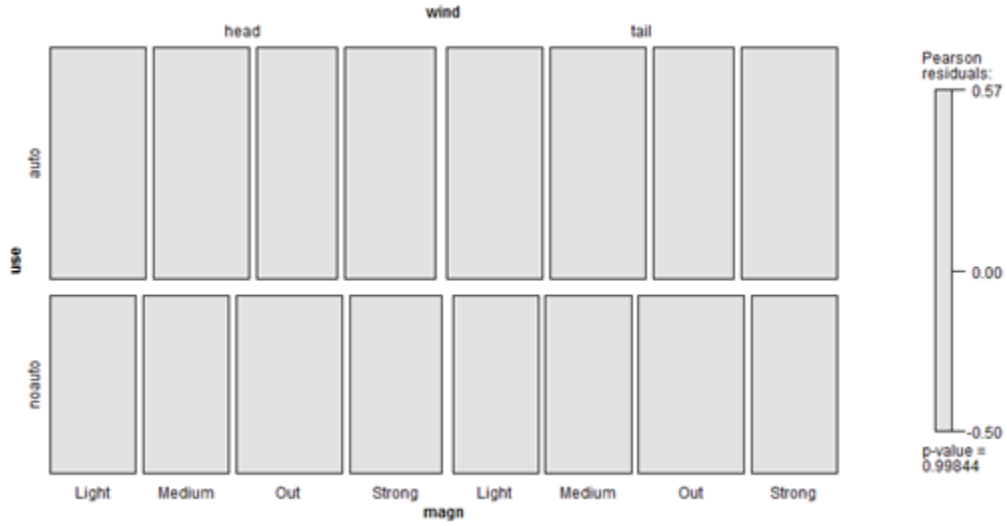


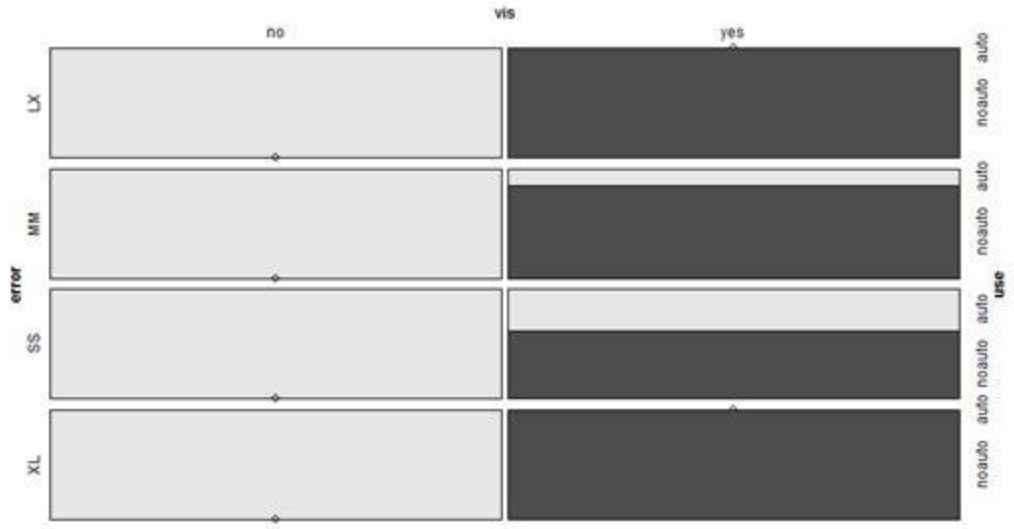


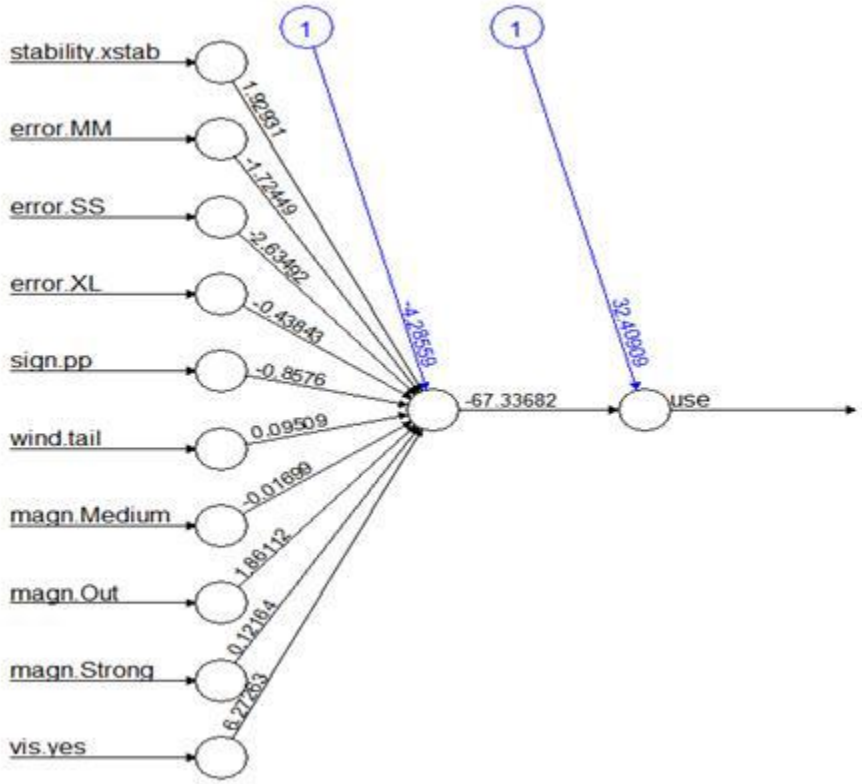

sigmoid
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4
x

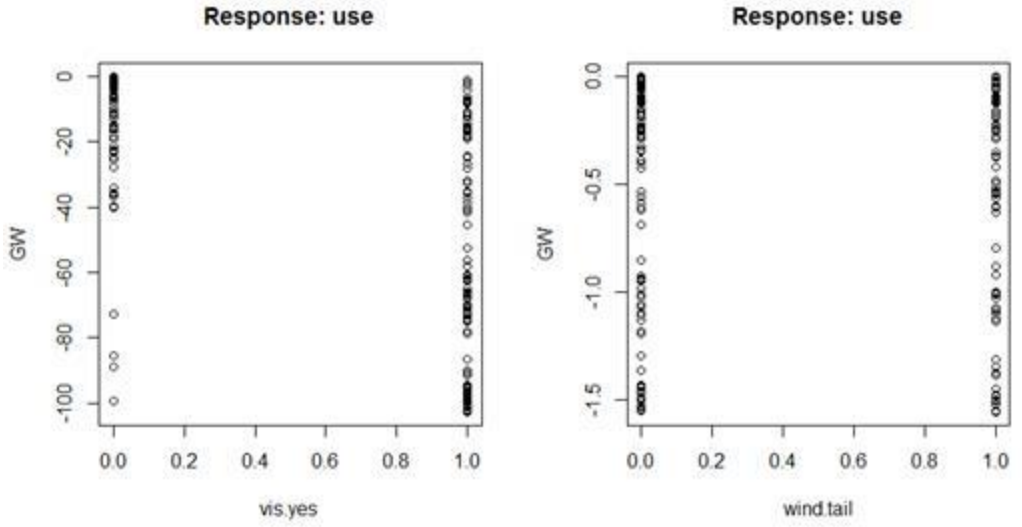

tanh
-1.0
-0.5
0.0
0.5
1.0
-4
-2
0
2
4
x

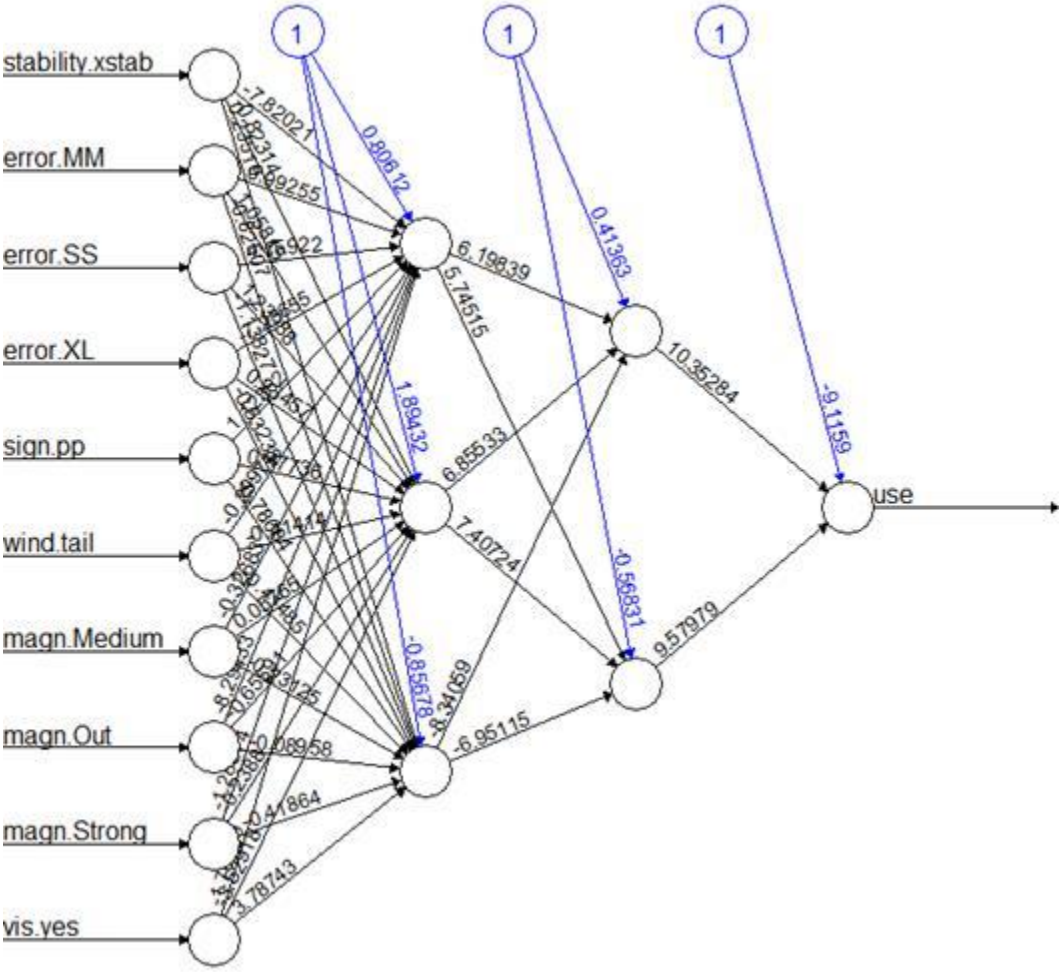

X1
X2
X3
X4
Feature A
Feature B
W1
W2
W3
W4
W5
Multiple Hidden Layers and Features
Inputs
Latent Features
Hidden Layer


wind
head
tail
use
auto
noauto
Light
Medium
Out
Strong
Light
Medium
Out
Strong
magn
Pearson residuals:
0.57
0.00
-0.50
p-value = 0.99844

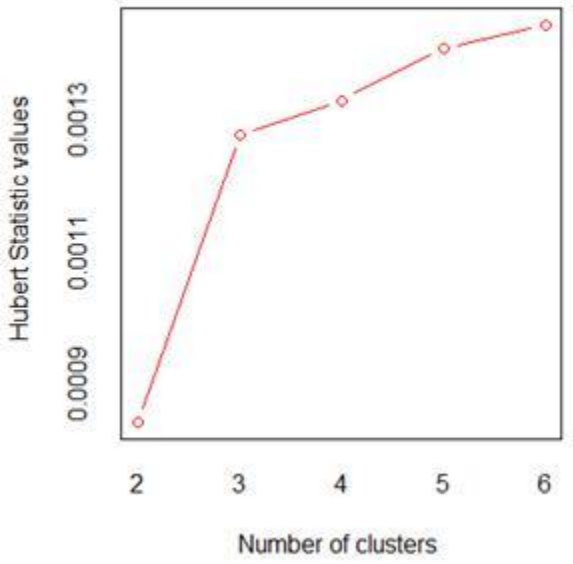

vis
no
yes
error
LX
MM
SS
XL
use
auto
noauto

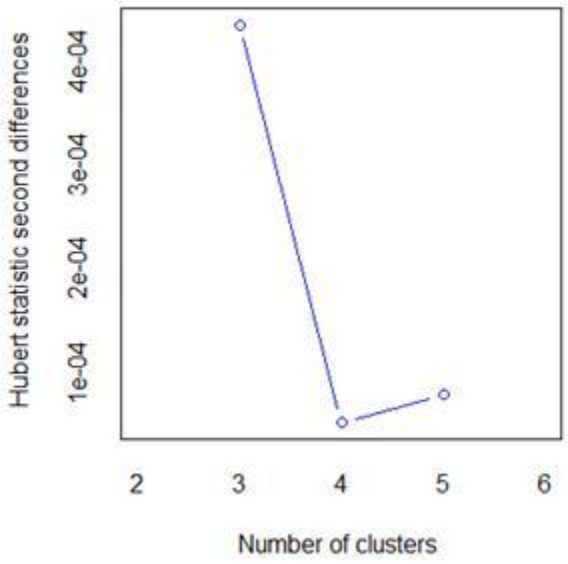

1
1
stability.xstab
error.MM
error.SS
error.XL
sign.pp
wind.tail
magn.Medium
magn.Out
magn.Strong
vis.yes
1.92931
-1.72449
-2.63492
-0.43843
-0.8576
0.09509
-0.01699
1.86112
0.12164
6.27263
-4.28559
32.40909
-67.33682
use

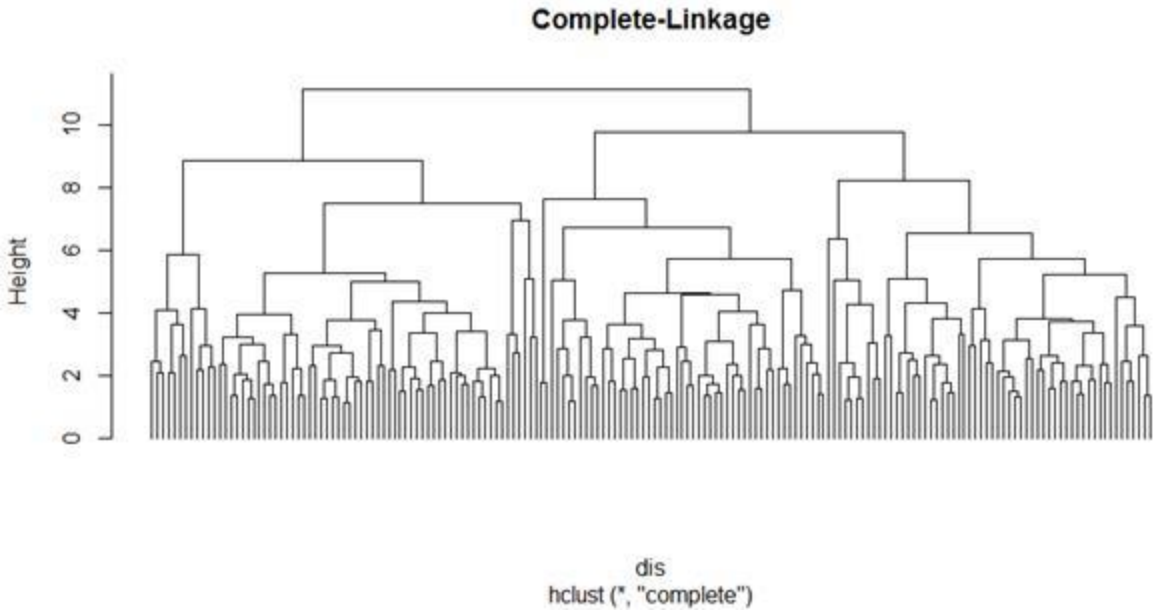

Response: use
GW
vis.yes
Response: use
GW
wind.tail

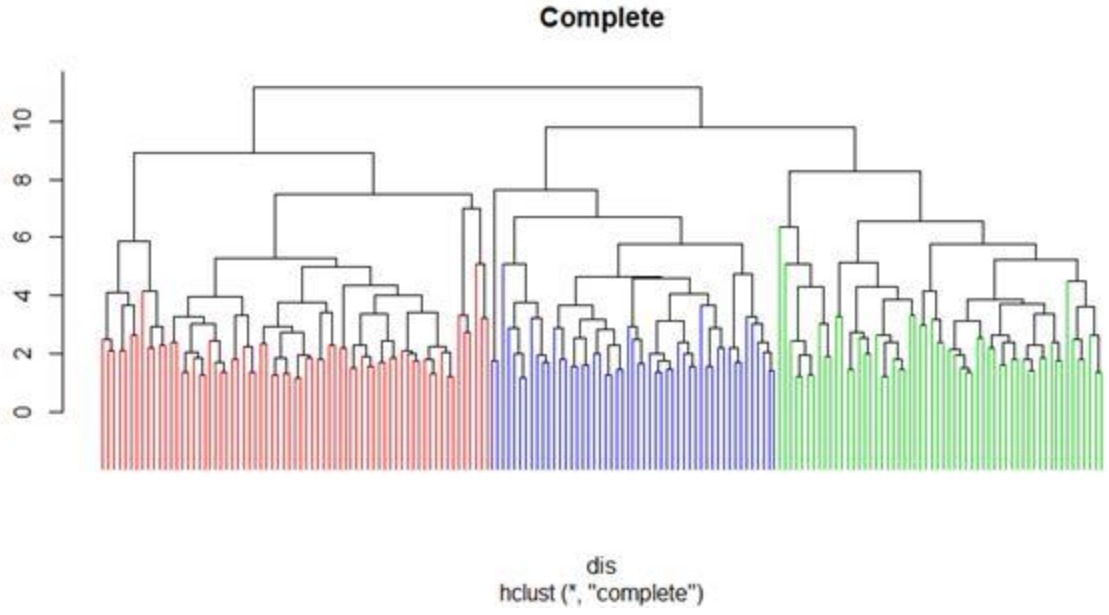

stability.xstab
error.MM
error.SS
error.XL
sign.pp
wind.tail
magn.Medium
magn.Out
magn.Strong
vis.yes
1
1
1
0.80612
0.41363
1.89432
-0.85678
6.19839
5.74515
6.85533
7.40724
-6.95115
10.35284
-0.56831
9.57979
-9.1159
use

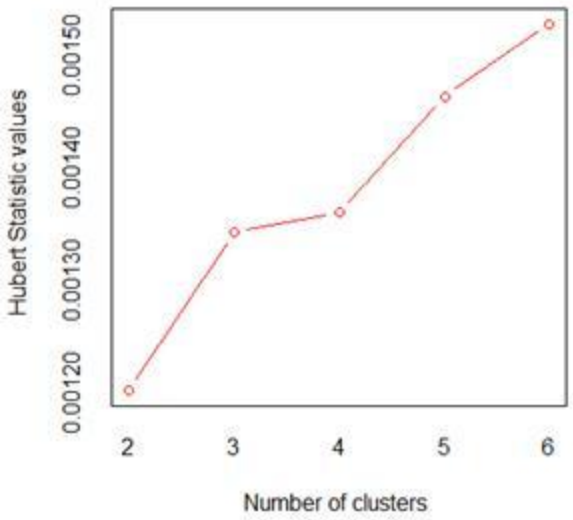
Chapter 8, Cluster Analysis

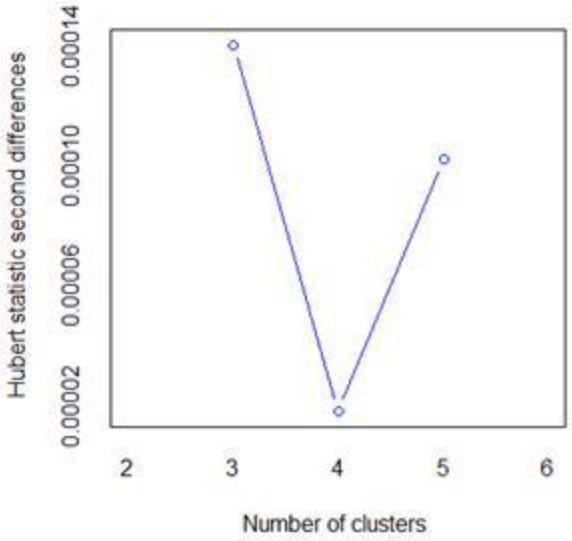


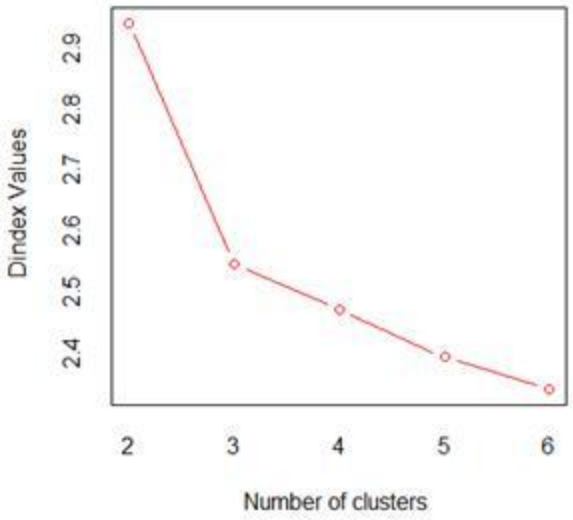


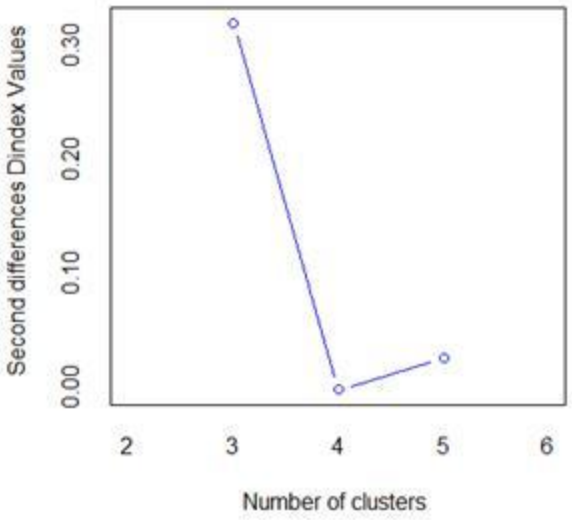

Complete-Linkage
Height
0
2
4
6
8
10
dis
hclust (*, "complete")

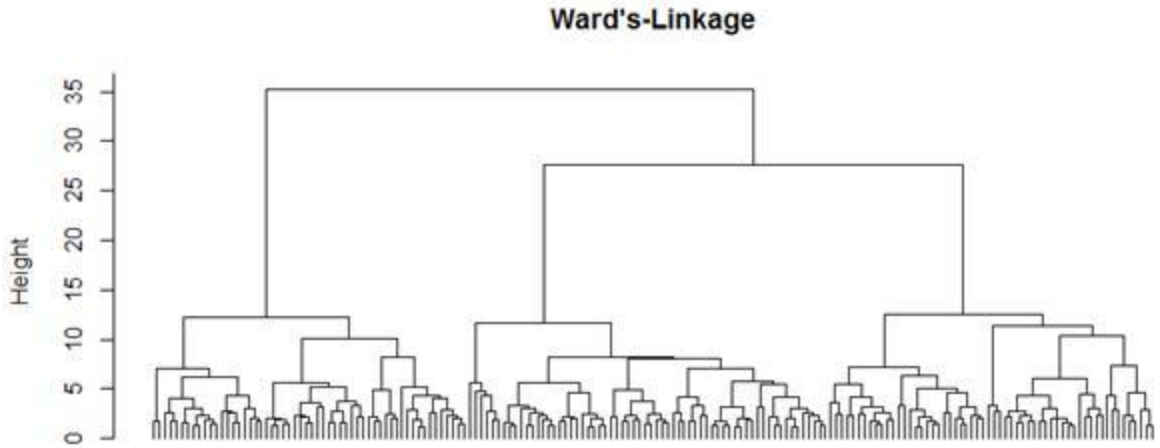

Complete
0
2
4
6
8
10
dis
hclust (*, "complete")


Hubert Statistic values
0.00120
0.00130
0.00140
0.00150
2
3
4
5
6
Number of clusters

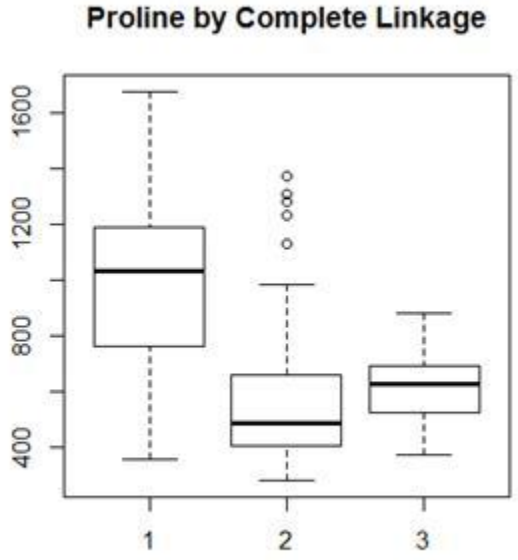

Hubert statistic second differences
0.00002
0.00006
0.00010
0.00014
2
3
4
5
6
Number of clusters

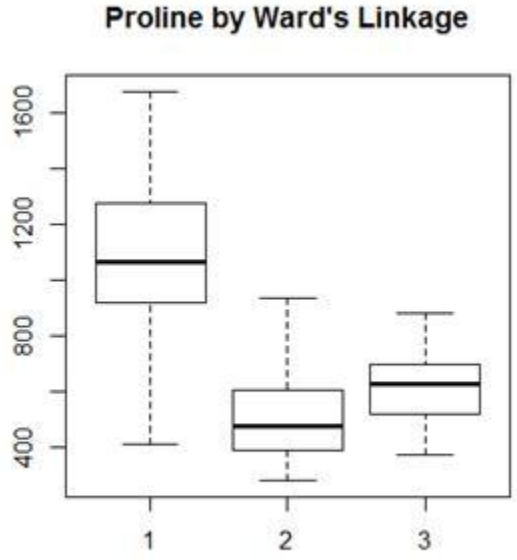

Dindex Values
2.4
2.5
2.6
2.7
2.8
2.9
2
3
4
5
6
Number of clusters

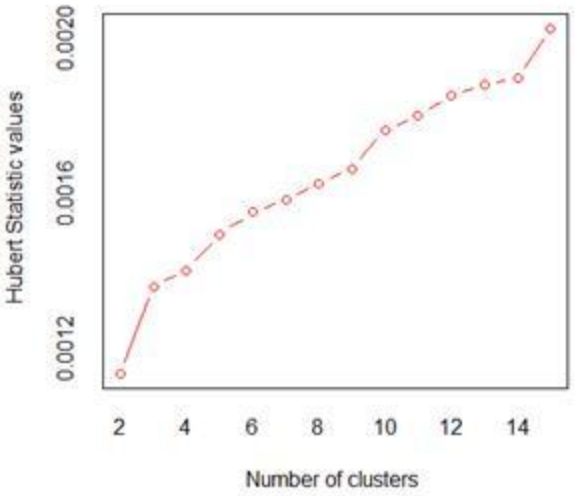

Second differences Dindex Values
0.00
0.10
0.20
0.30
2
3
4
5
6
Number of clusters

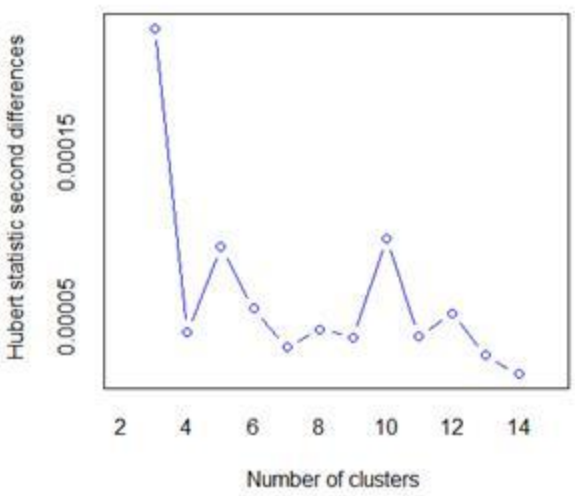


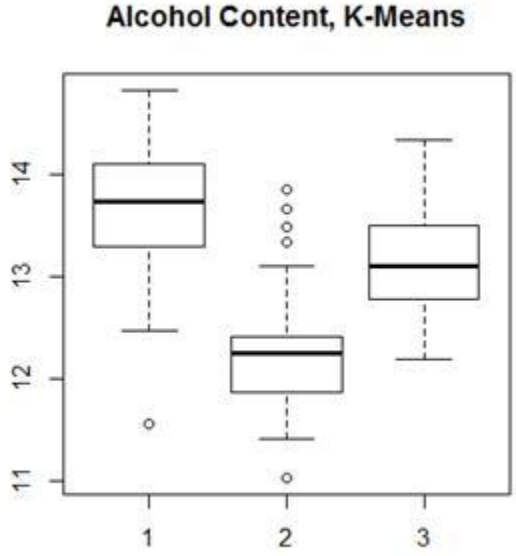
dis
hclust (*, "ward.D2")






Hubert Statistic values
0.0012
0.0016
0.0020
2
4
6
8
10
12
14
Number of clusters

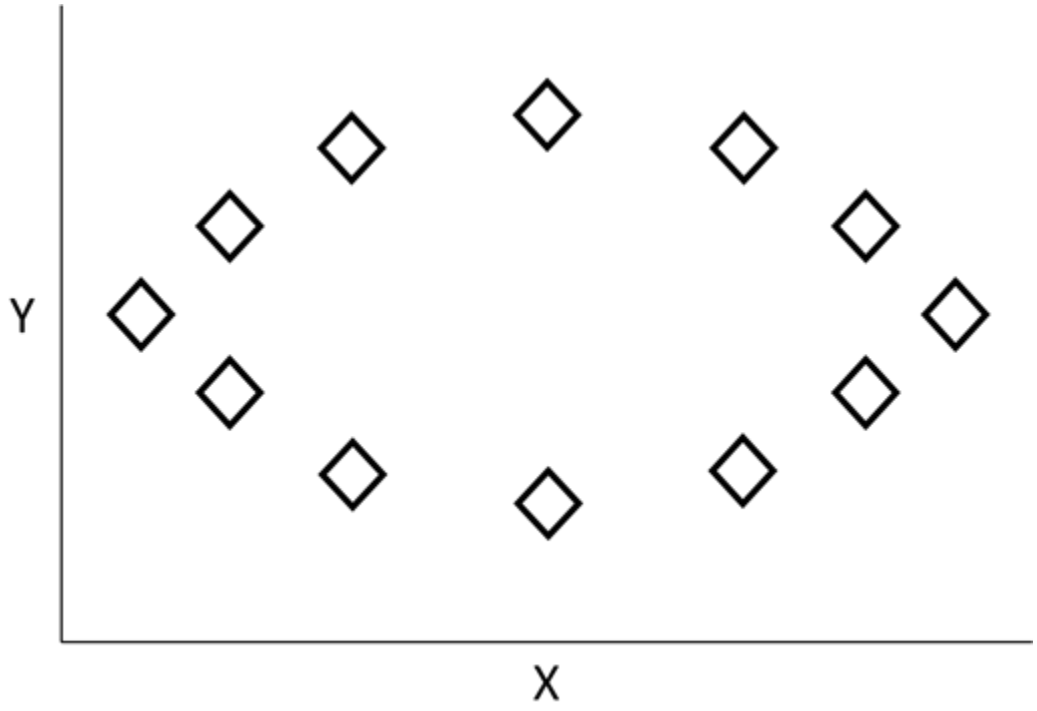

Hubert statistic second differences
0.00005
0.00015
2
4
6
8
10
12
14
Number of clusters

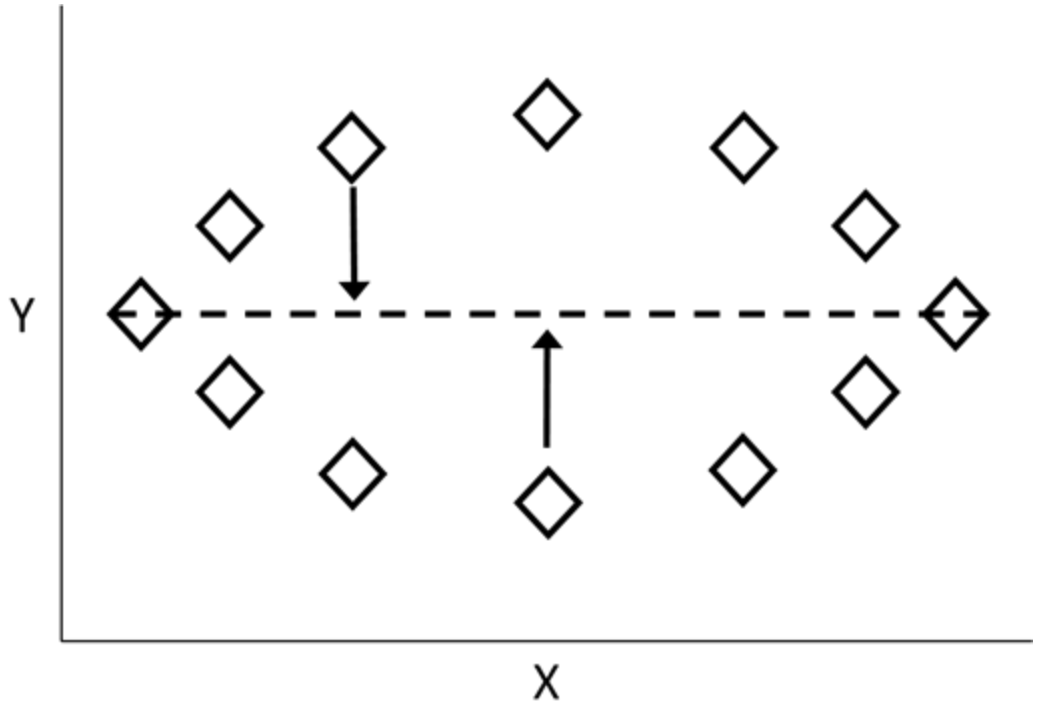

Alcohol Content, K-Means
11
12
13
14
1
2
3

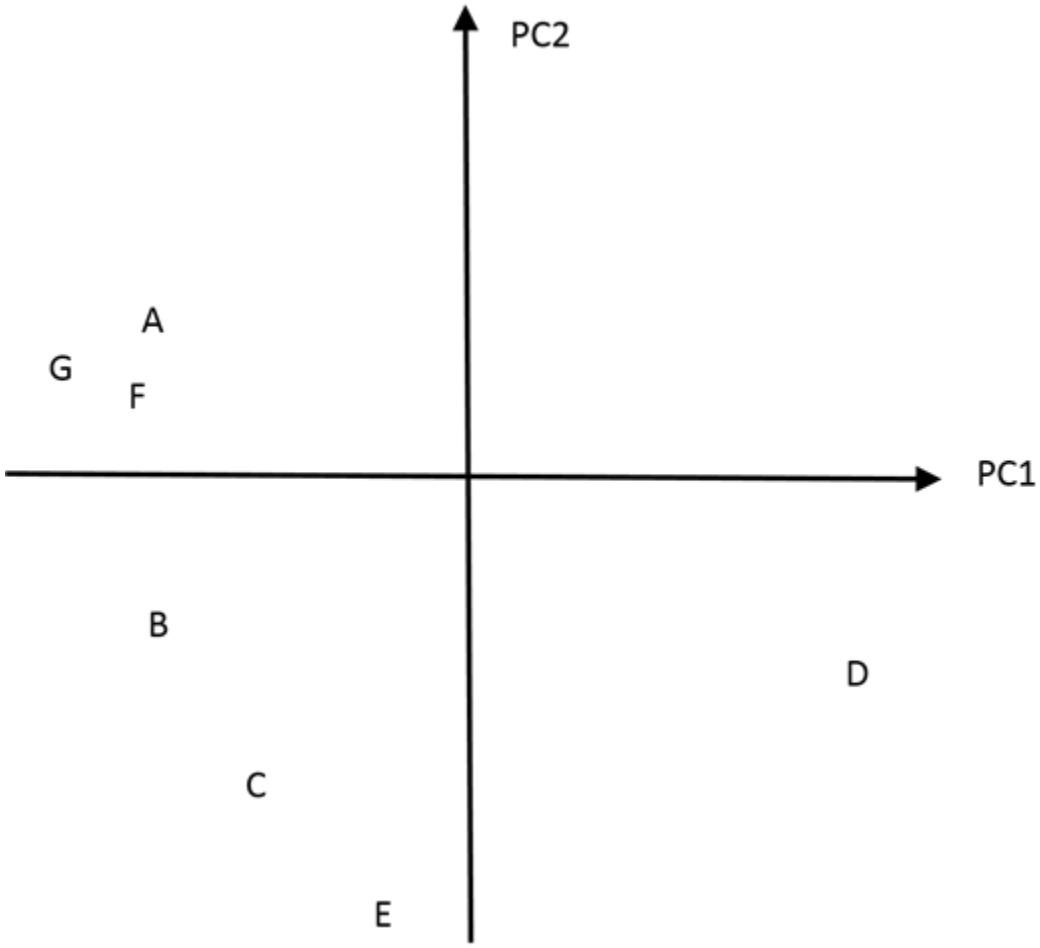
Alcohol Content, Ward's
11
12
13
14
1
2
3

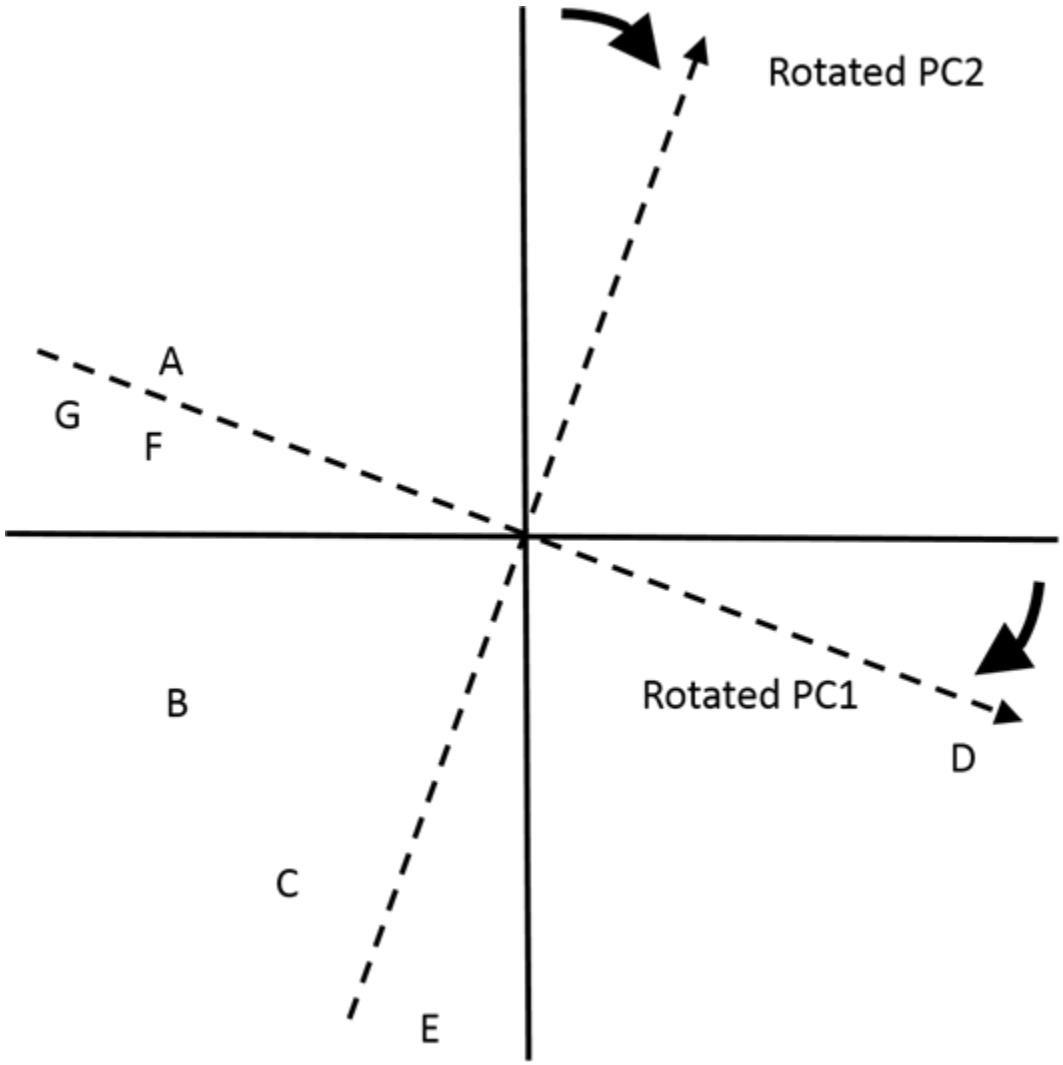
	1	2	3	p.overall
	N=60	N=69	N=49	
Alcohol:				<0.001
High	58 (96.7%)	6 (8.70%)	28 (57.1%)	
Low	2 (3.33%)	63 (91.3%)	21 (42.9%)	
MalicAcid	-0.31 (0.62)	-0.37 (0.89)	0.90 (0.97)	<0.001
Ash	0.28 (0.89)	-0.42 (1.14)	0.25 (0.67)	<0.001
Alk_ash	-0.75 (0.76)	0.24 (1.00)	0.58 (0.67)	<0.001
magnesium	0.43 (0.77)	-0.34 (1.18)	-0.05 (0.77)	<0.001
T_phenols	0.87 (0.54)	-0.06 (0.86)	-0.99 (0.56)	<0.001
Flavanoids	0.96 (0.40)	0.04 (0.70)	-1.23 (0.31)	<0.001
Non_flav	-0.58 (0.56)	0.00 (0.98)	0.71 (1.00)	<0.001
Proantho	0.55 (0.72)	0.05 (1.06)	-0.75 (0.72)	<0.001
C_Intensity	0.20 (0.53)	-0.87 (0.38)	0.99 (1.00)	<0.001
Hue	0.46 (0.51)	0.44 (0.89)	-1.19 (0.51)	<0.001
OD280_315	0.77 (0.50)	0.25 (0.69)	-1.30 (0.38)	<0.001
Proline	1.14 (0.74)	-0.72 (0.51)	-0.38 (0.37)	<0.001
comp_cluster	-0.94 (0.42)	-0.14 (0.59)	1.35 (0.00)	<0.001
ward_cluster	-1.16 (0.00)	0.11 (0.49)	1.27 (0.00)	<0.001
km_cluster	-1.16 (0.16)	0.06 (0.34)	1.33 (0.00)	<0.001
class:				<0.001
1	59 (98.3%)	0 (0.00%)	0 (0.00%)	
2	1 (1.67%)	69 (100%)	1 (2.04%)	
3	0 (0.00%)	0 (0.00%)	48 (98.0%)	

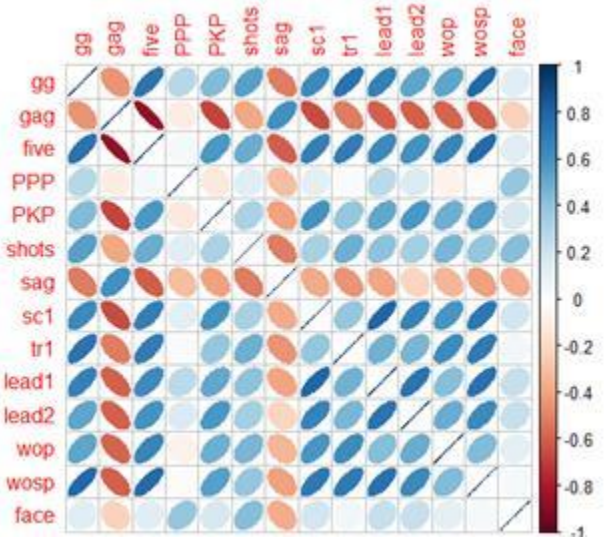
Chapter 9: Principal Components Analysis

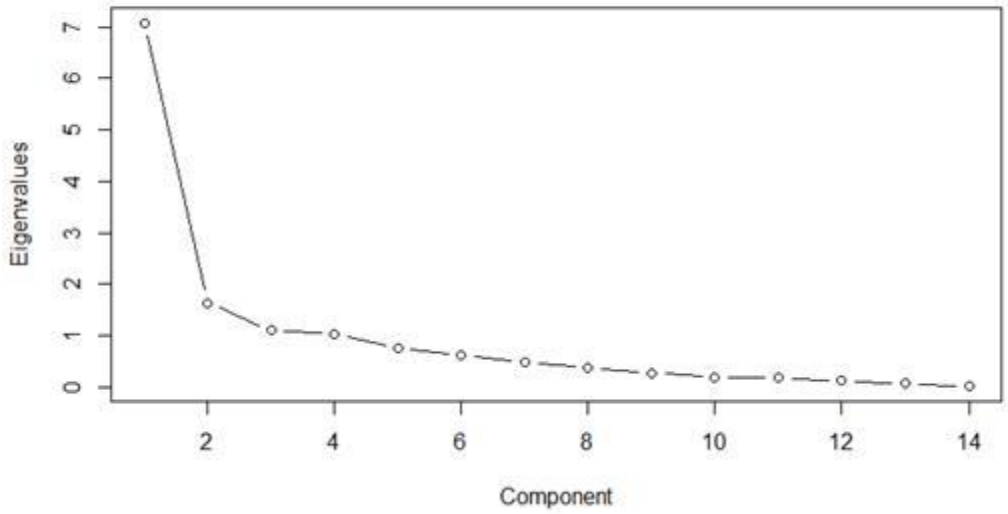


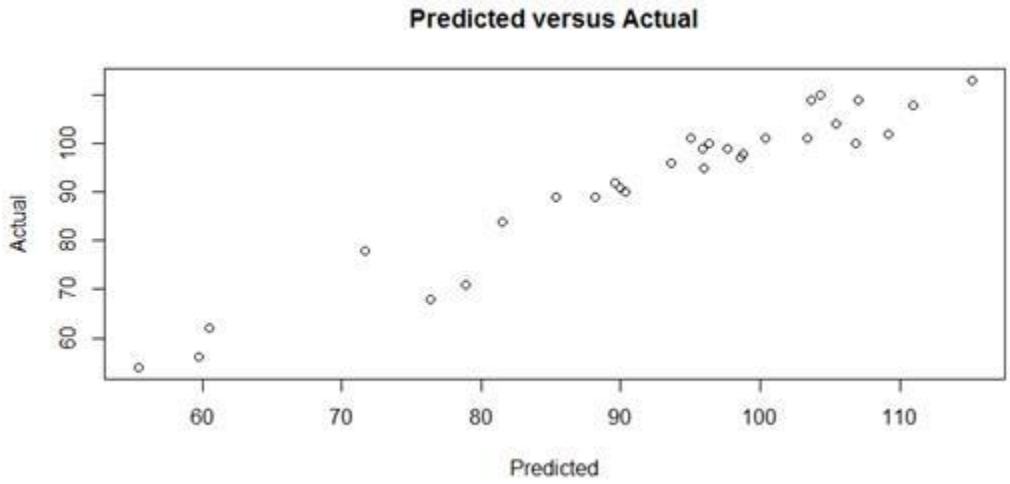


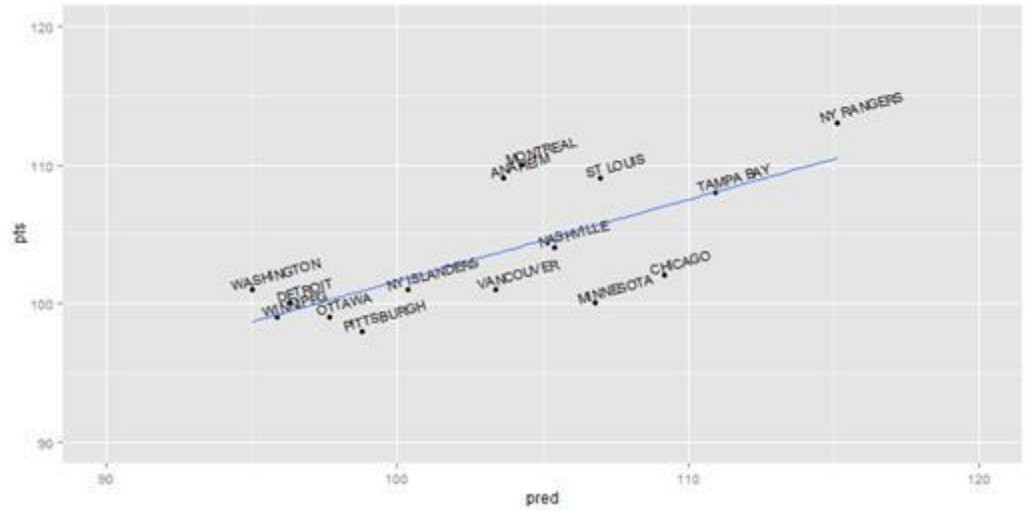

PC2
PC1
A
G
F
B
D
C
E

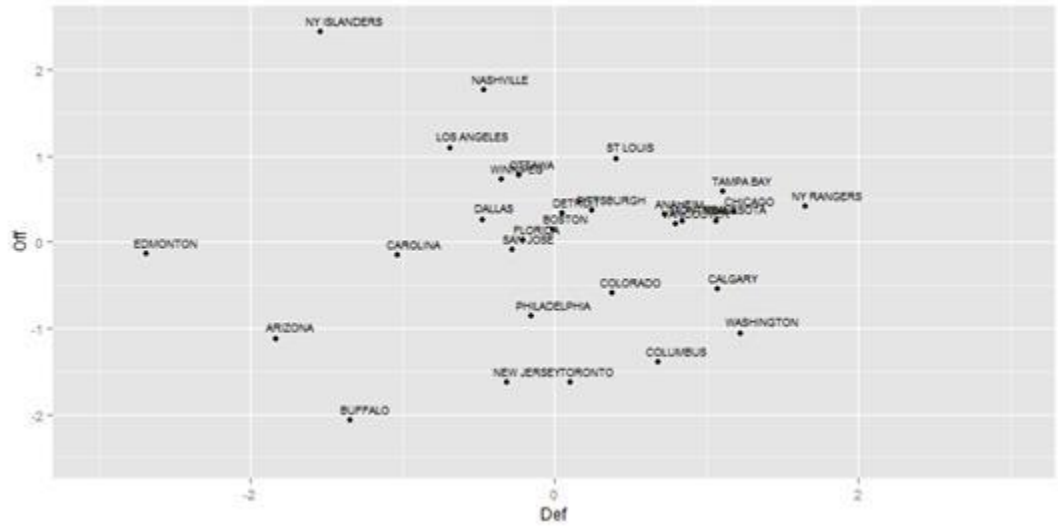

Rotated PC2
A
G
F
B
Rotated PC1
D
C
E


gg
gag
five
PPP
PKP
shots
sag
sc1
tr1
lead1
lead2
wop
wosp
face
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

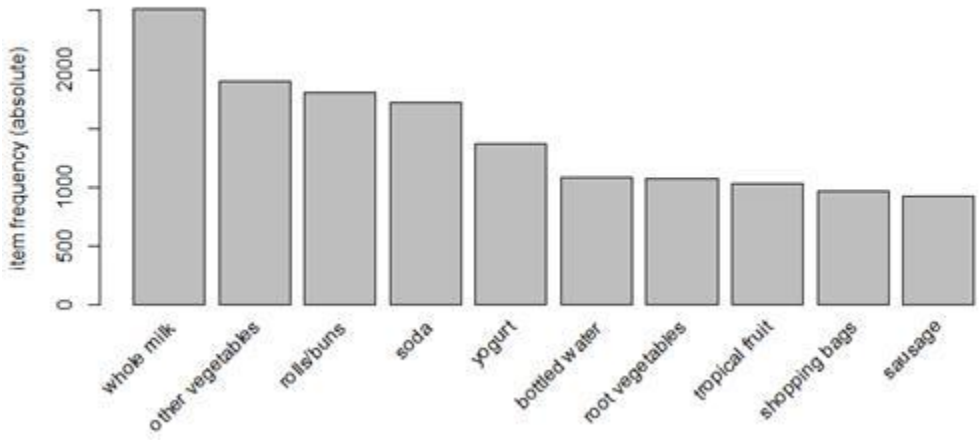

Eigenvalues
0
1
2
3
4
5
6
7
2
4
6
8
10
12
14
Component

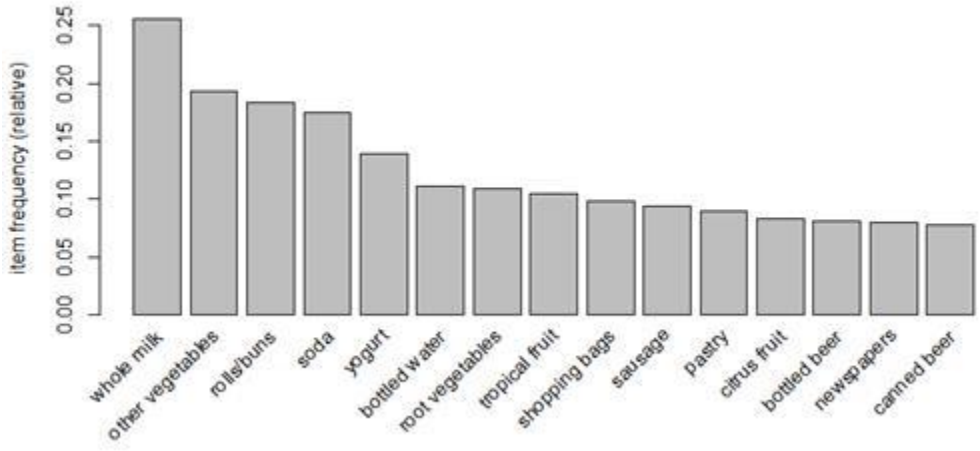

Predicted versus Actual
Actual
60
70
80
90
100
60
70
80
90
100
110
Predicted

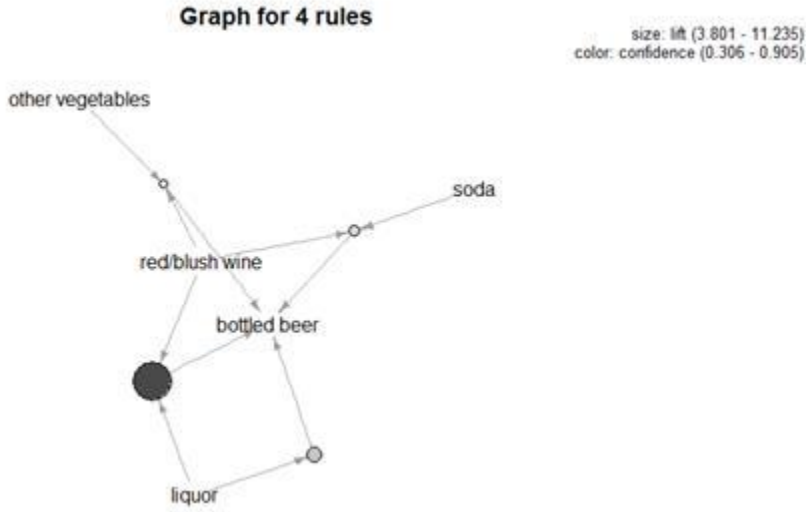

120
110
100
90
pts
90
100
110
120
pred
NY RANGERS
MONTREAL
ANAHEIM
ST LOUIS
TAMPA BAY
NASHVILLE
WASHINGTON
DETROIT
WINNIPEG
NY ISLANDERS
VANCOUVER
CHICAGO
MINNESOTA
OTTAWA
PITTSBURGH


2
1
0
-1
-2
Off
-2
0
2
Def
NY ISLANDERS
NASHVILLE
LOS ANGELES
ST LOUIS
TAMPA BAY
NY RANGERS
DALLAS
BOSTON
EDMONTON
CAROLINA
SAN JOSE
COLORADO
CALGARY
PHILADELPHIA
WASHINGTON
ARIZONA
COLUMBUS
NEW JERSEY
TORONTO
BUFFALO

Chapter 10: Market Basket Analysis and Recommendation Engines

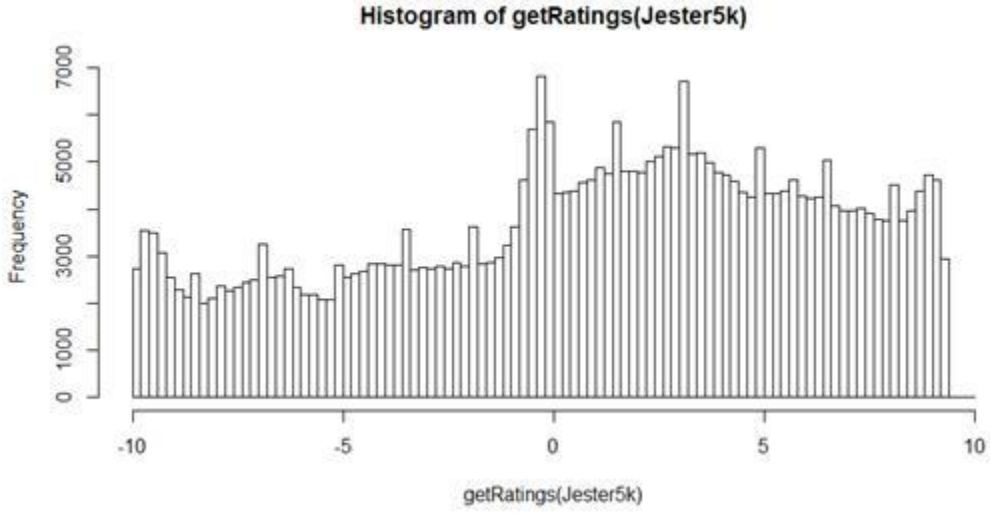


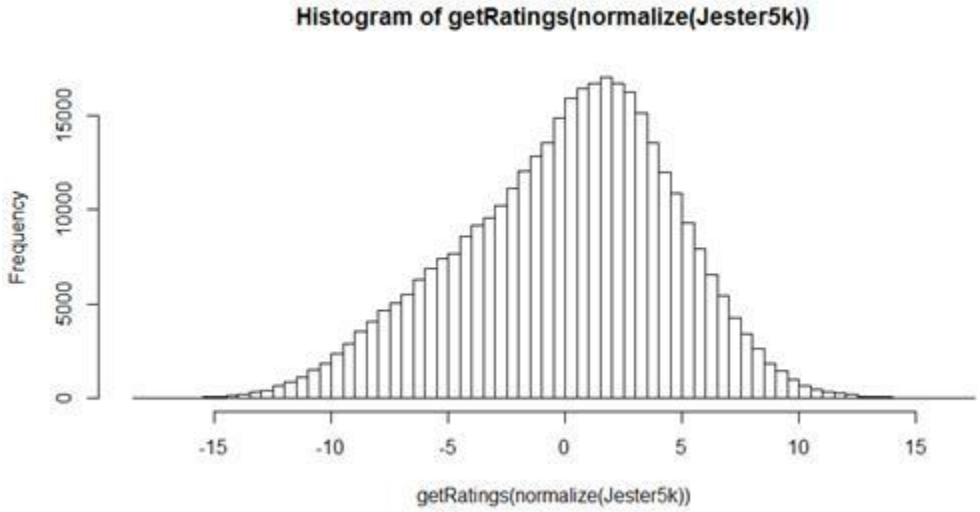


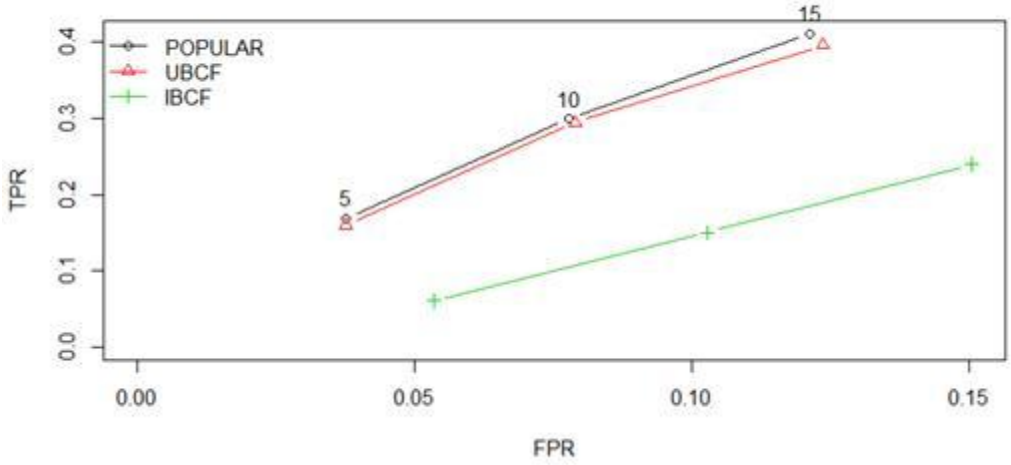


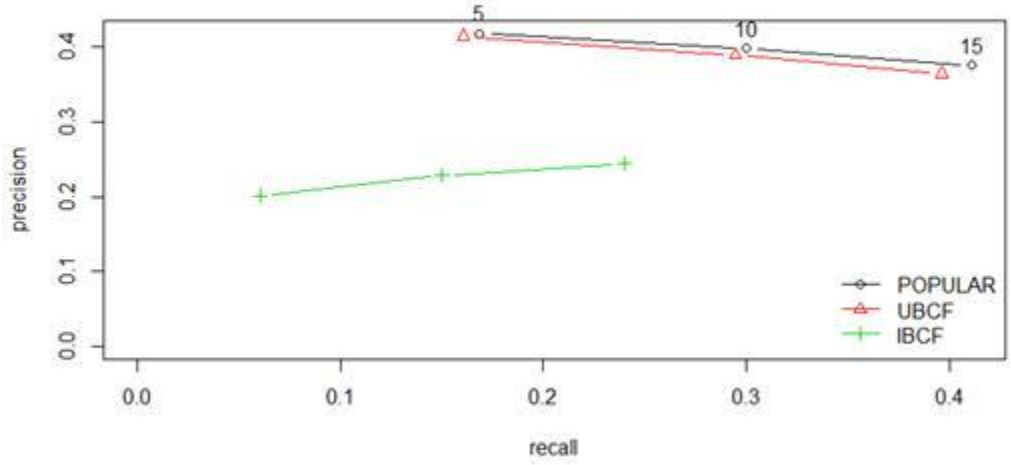
	Avengers	American Sniper	Les Miserable	Mad Max
Homer	3	5	3	4
Marge	5	2	5	3
Bart	5	5	1	4
Lisa	5	1	5	2
Flanders	1	1	4	1
Me	1	5	2	?

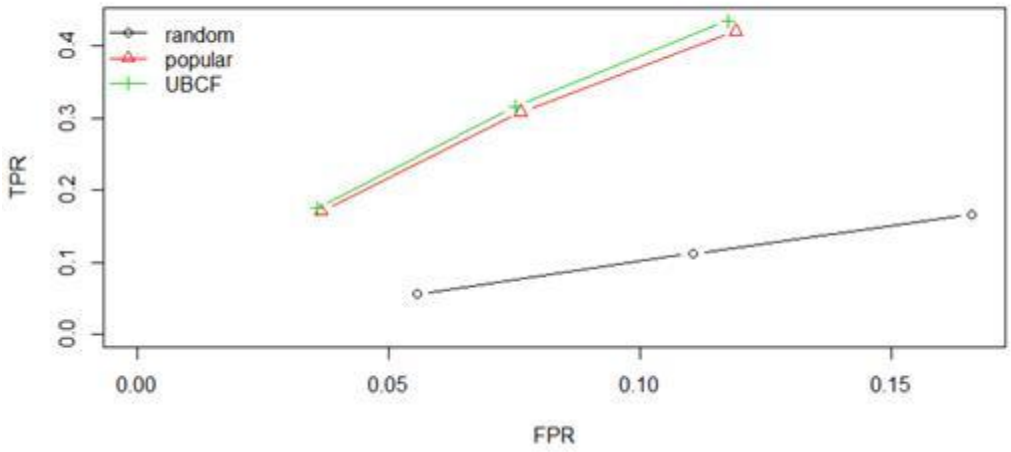
	Avengers	American Sniper	Les Miserable	Mad Max
Homer	3	5	3	4
Marge	5	2	5	3
Bart	5	5	1	4
Lisa	5	1	5	2
Flanders	1	1	4	1
Me	1	5	2	?

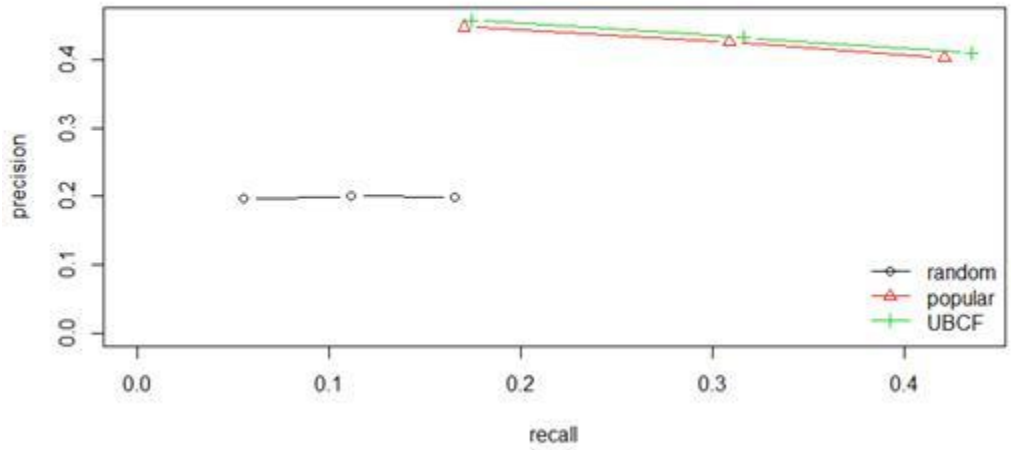

Histogram of getRatings(Jester5k)
Frequency
0
1000
3000
5000
7000
-10
-5
0
5
10
getRatings(Jester5k)


Histogram of getRatings(normalize(Jester5k))
Frequency
0
5000
10000
15000
-15
-10
-5
0
5
10
15
getRatings(normalize(Jester5k))

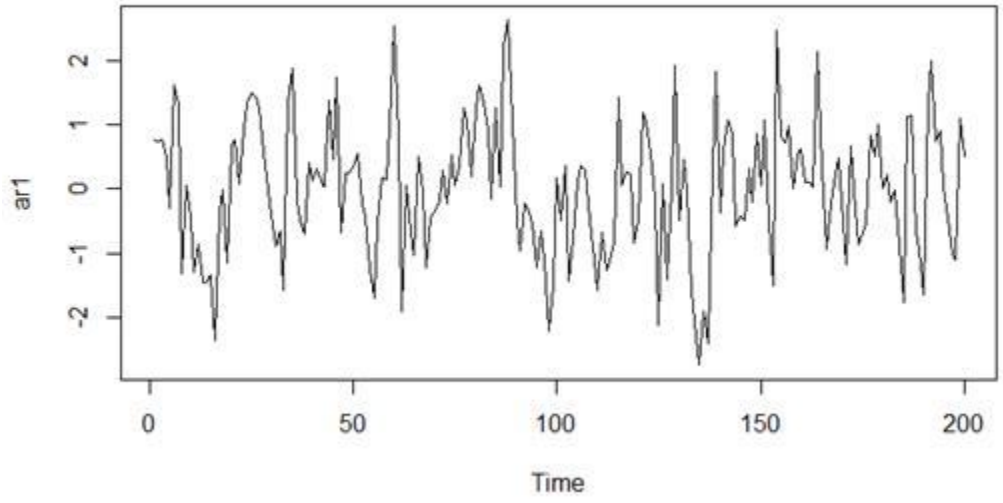

POPULAR
UBCF
IBCF
TPR
FPR
0.0
0.1
0.2
0.3
0.4
0.00
0.05
0.10
0.15
5
10
15

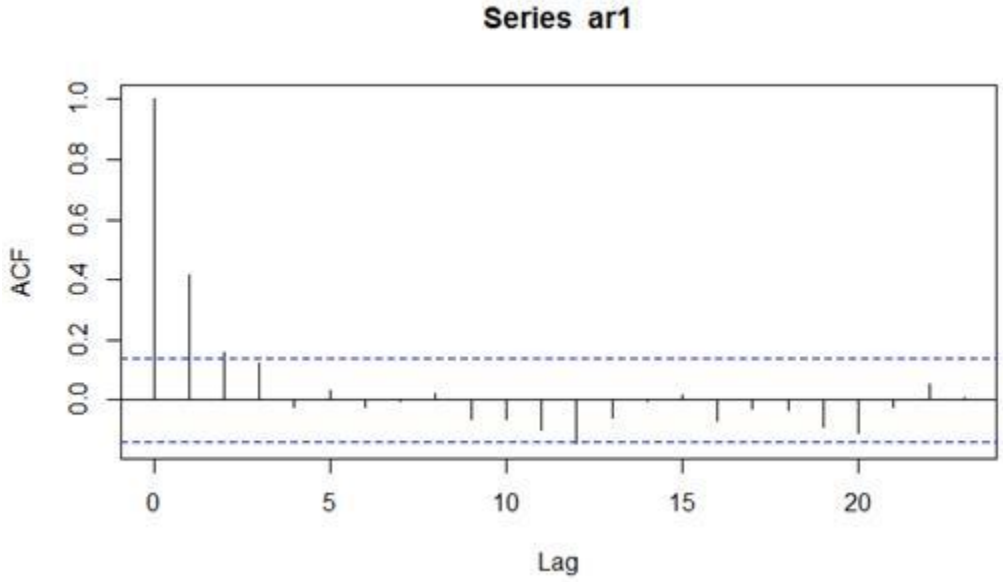

precision
recall
0.0
0.1
0.2
0.3
0.4
5
10
15
POPULAR
UBCF
IBCF

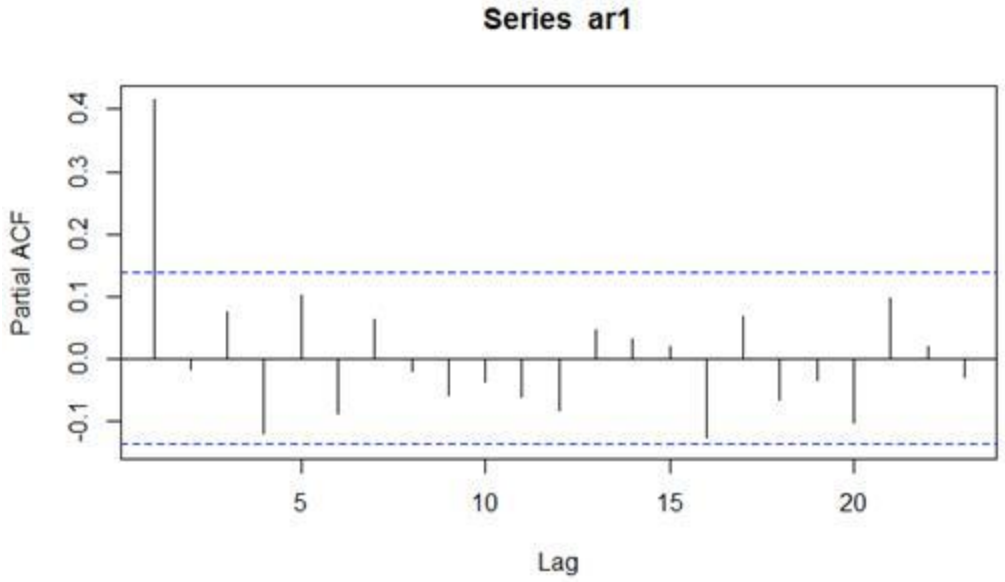

random
popular
UBCF
TPR
FPR
0.0
0.1
0.2
0.3
0.4
0.00
0.05
0.10
0.15

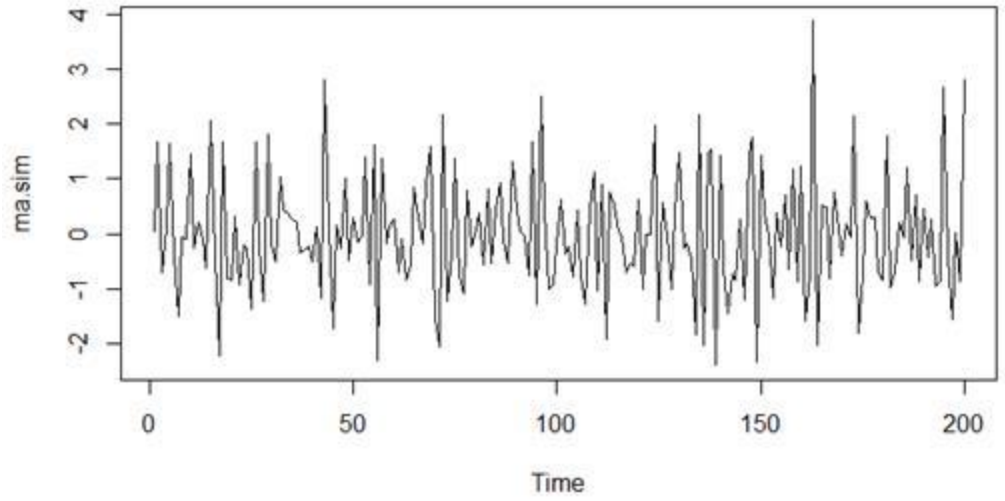

precision
recall
random
popular
UBCF
0.0
0.1
0.2
0.3
0.4

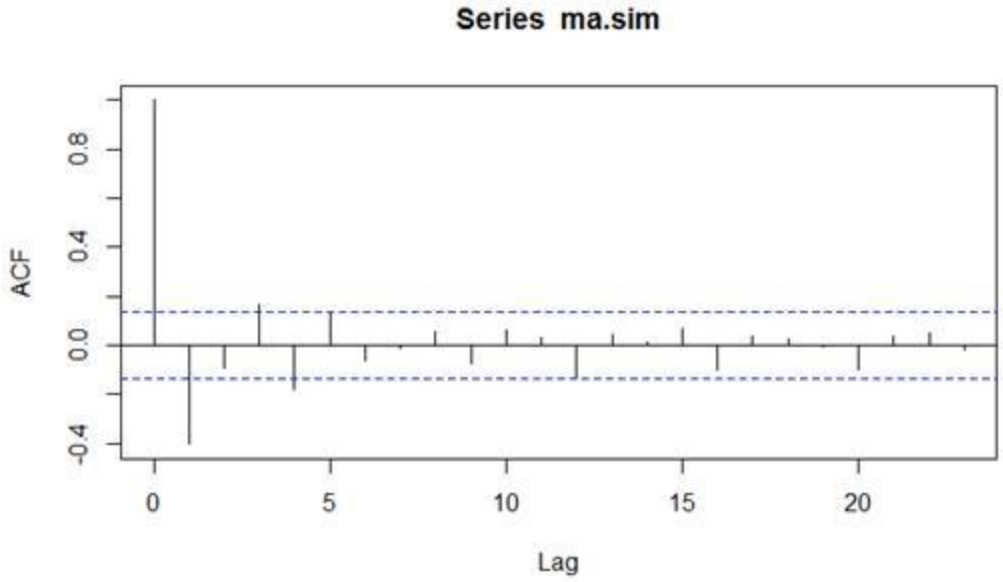
Chapter 11: Time Series and Causality

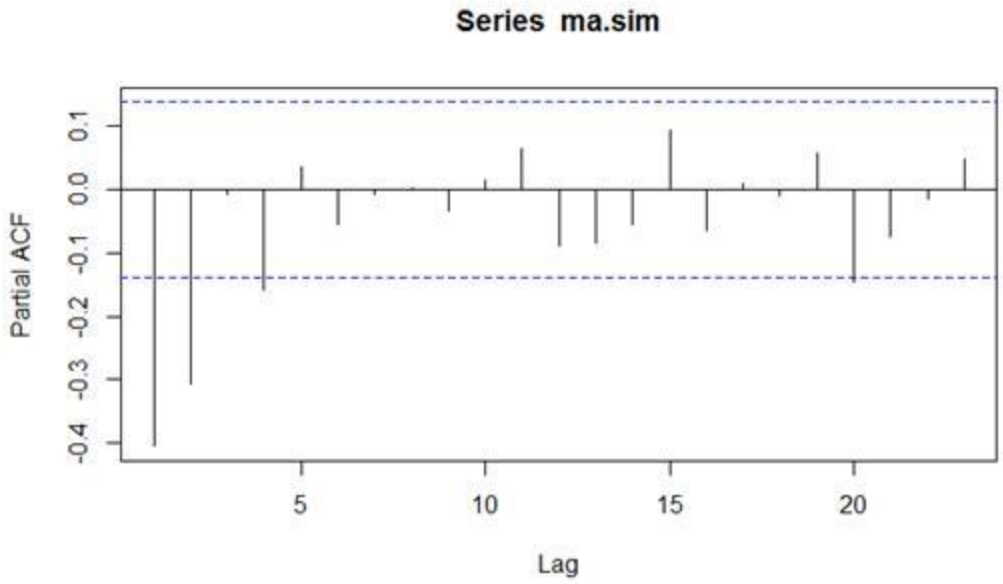


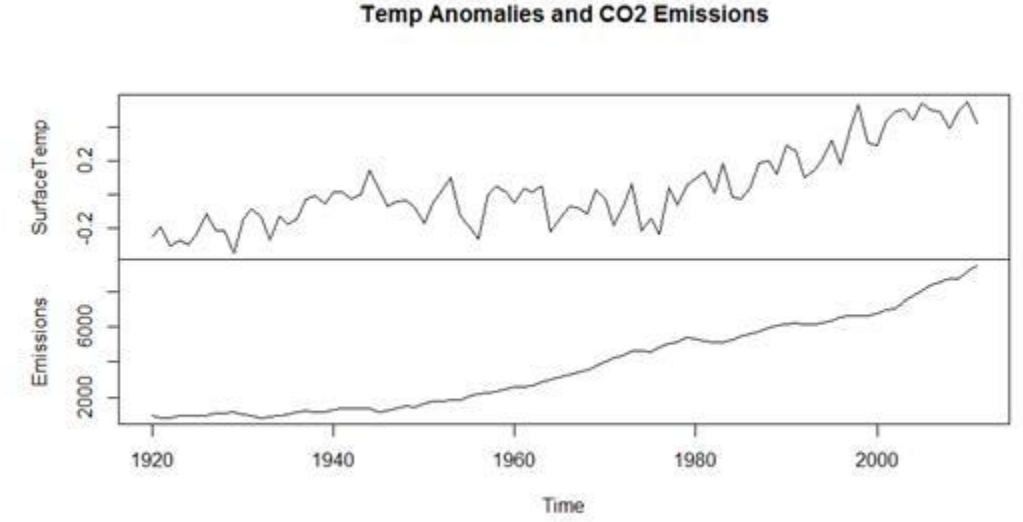


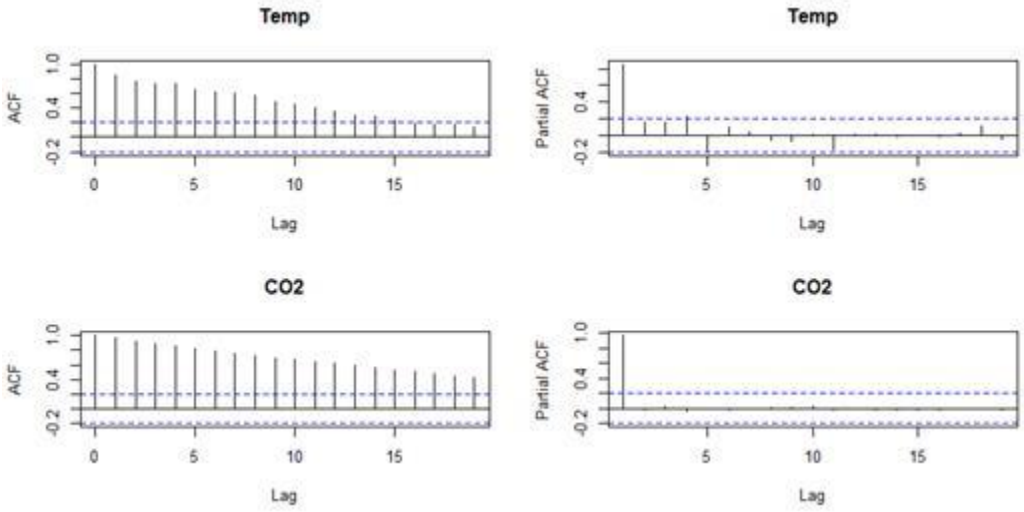

Series ar1
Partial ACF
0.4
0.3
0.2
0.1
0.0
-0.1
5
10
15
20
Lag

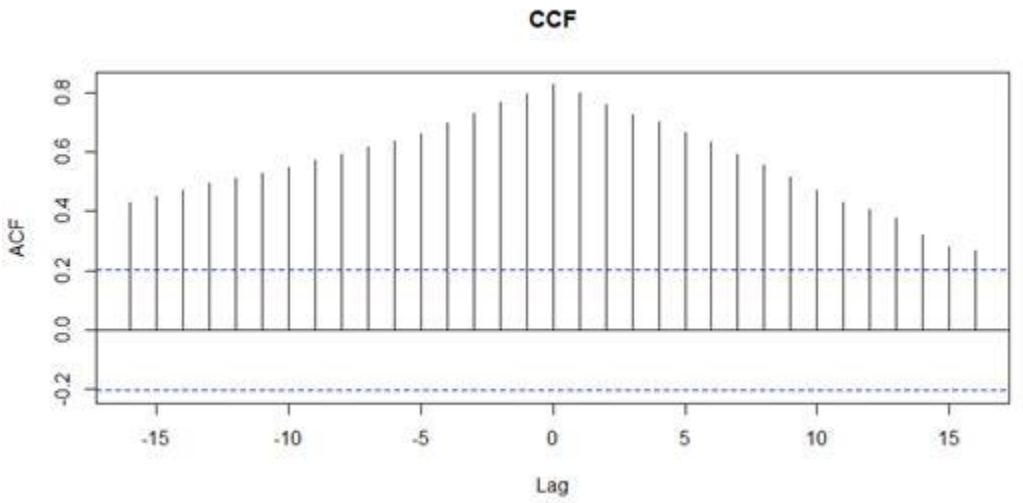

ma.sim
4
3
2
1
0
-1
-2
0
50
100
150
200
Time

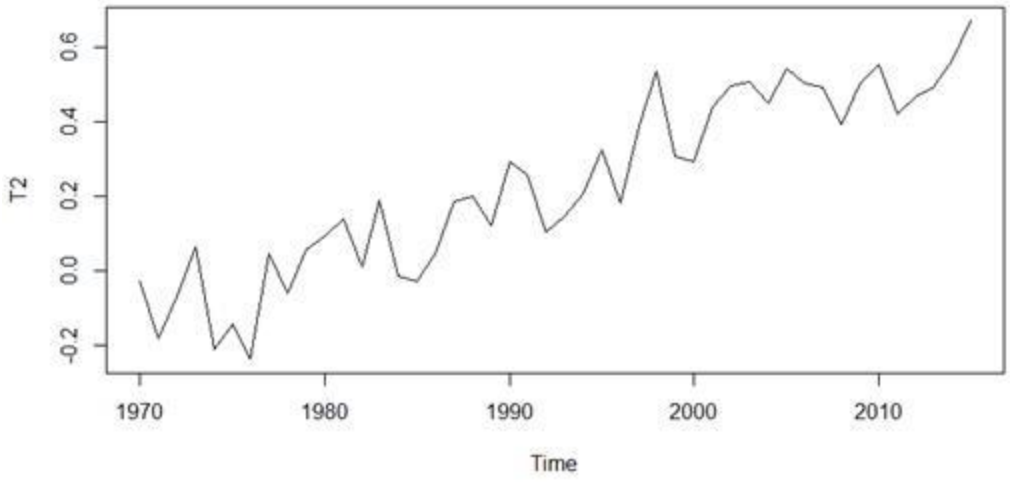

Series ma.sim
ACF
0.8
0.4
0.0
-0.4
0
5
10
15
20
Lag

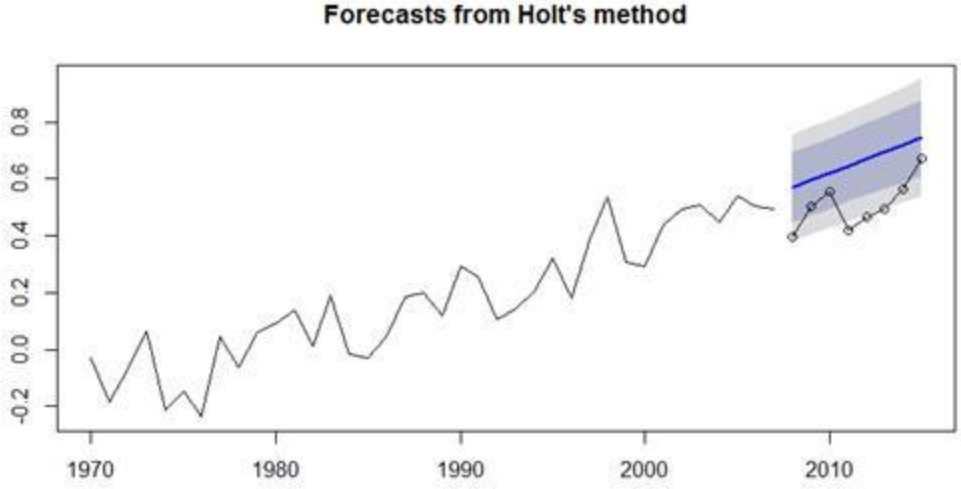

Series ma.sim
Partial ACF
0.1
0.0
-0.1
-0.2
-0.3
-0.4
5
10
15
20
Lag

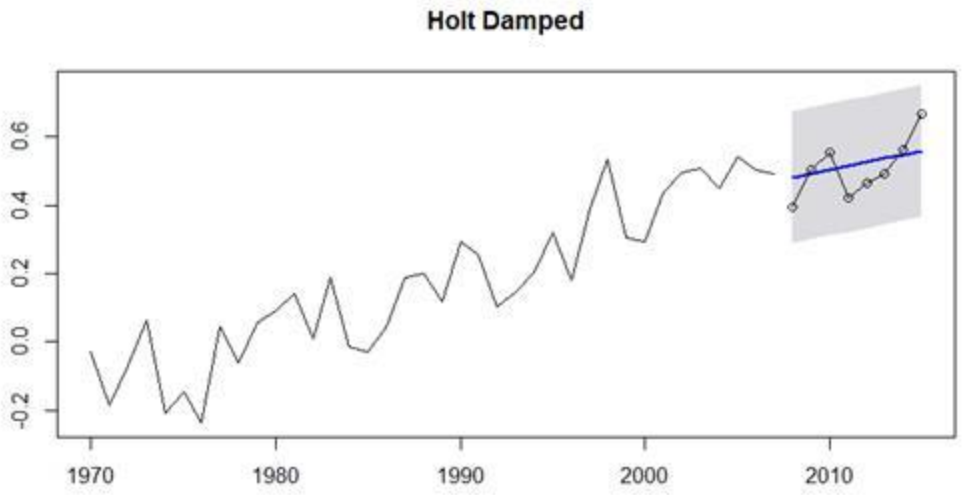

Temp Anomalies and CO2 Emissions
SurfaceTemp
-0.2
0.2
Emissions
2000
6000
1920
1940
1960
1980
2000
Time

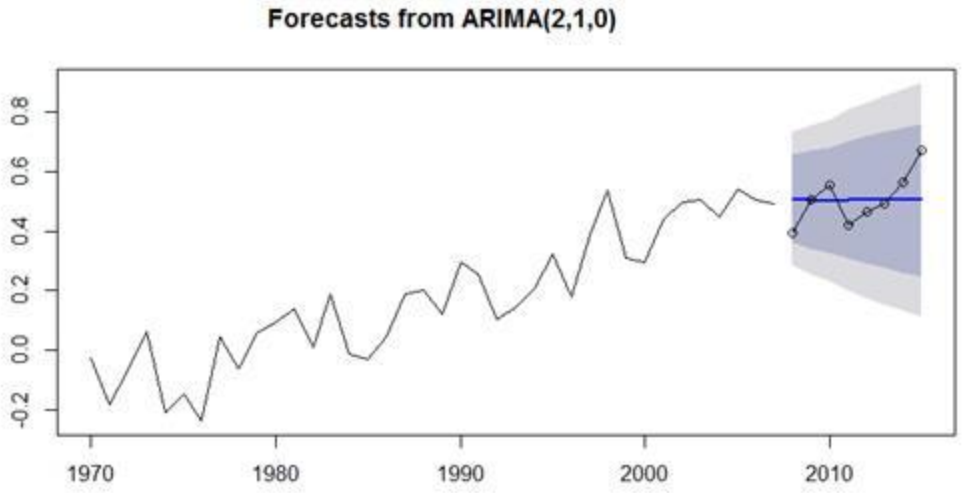

Temp
ACF
1.0
0.4
-0.2
0
5
10
15
Lag
Temp
Partial ACF
0.4
-0.2
5
10
15
Lag
CO2
ACF
1.0
0.4
-0.2
0
5
10
15
Lag
CO2
Partial ACF
1.0
0.4
-0.2
5
10
15
Lag

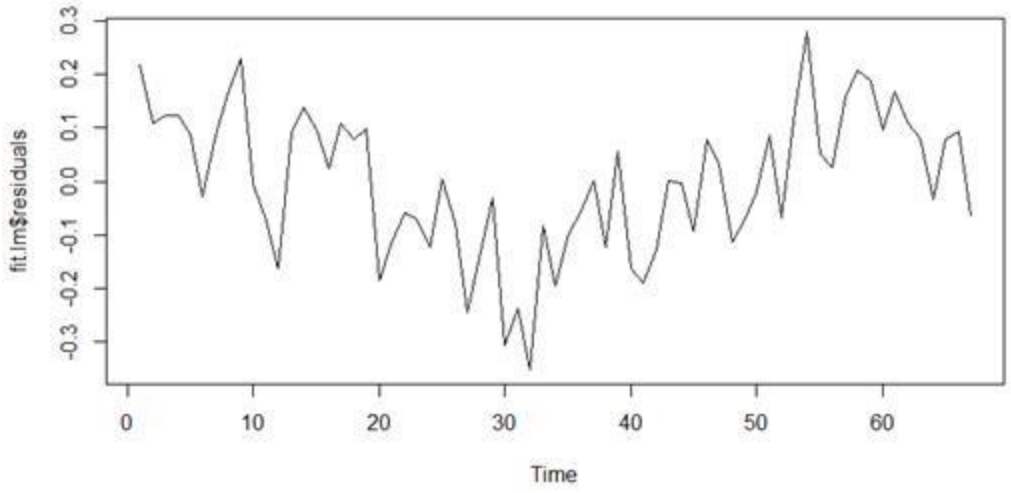

CCF
ACF
0.8
0.6
0.4
0.2
0.0
-0.2
-15
-10
-5
0
5
10
15
Lag

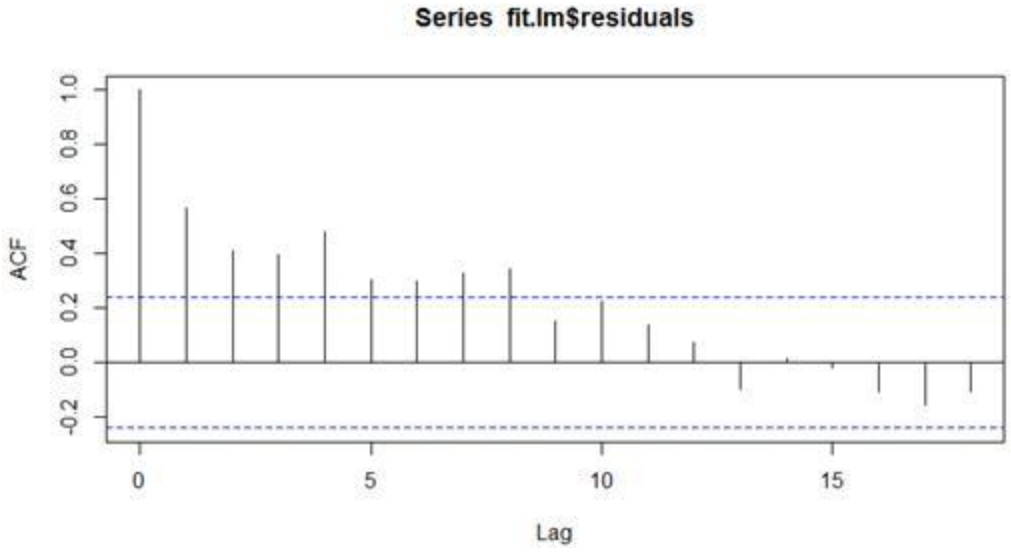

T2
-0.2
0.0
0.2
0.4
0.6
1970
1980
1990
2000
2010
Time

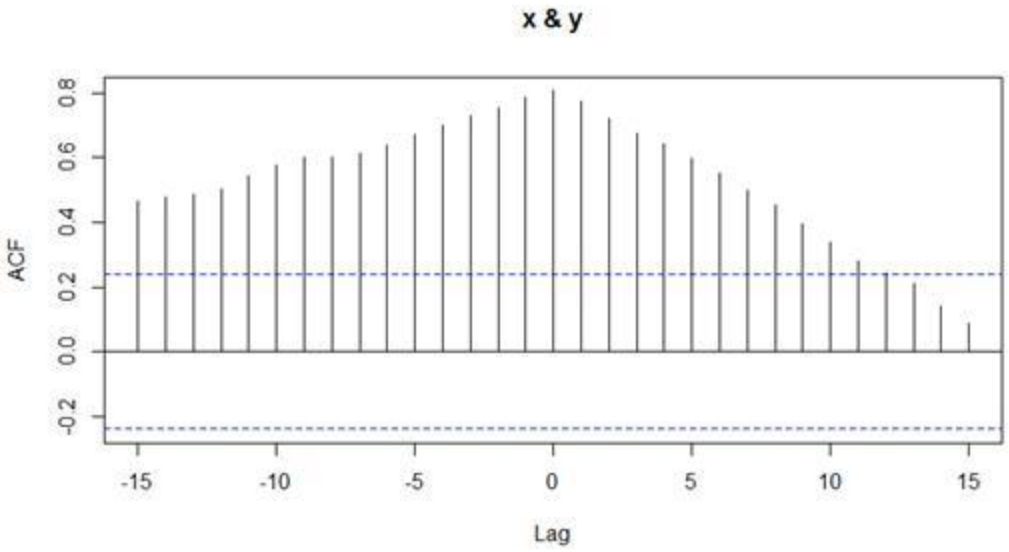

Forecasts from Holt's method
-0.2
0.0
0.2
0.4
0.6
0.8
1970
1980
1990
2000
2010

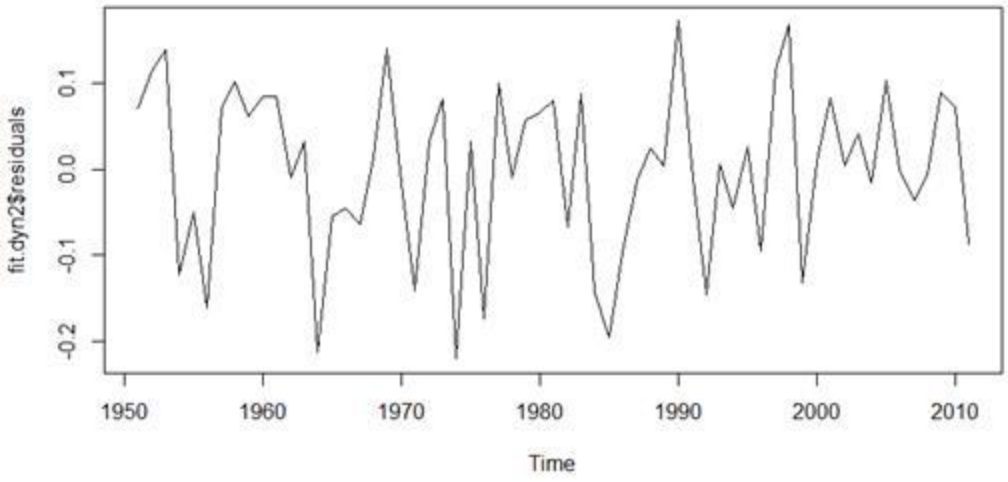

Holt Damped
-0.2
0.0
0.2
0.4
0.6
1970
1980
1990
2000
2010

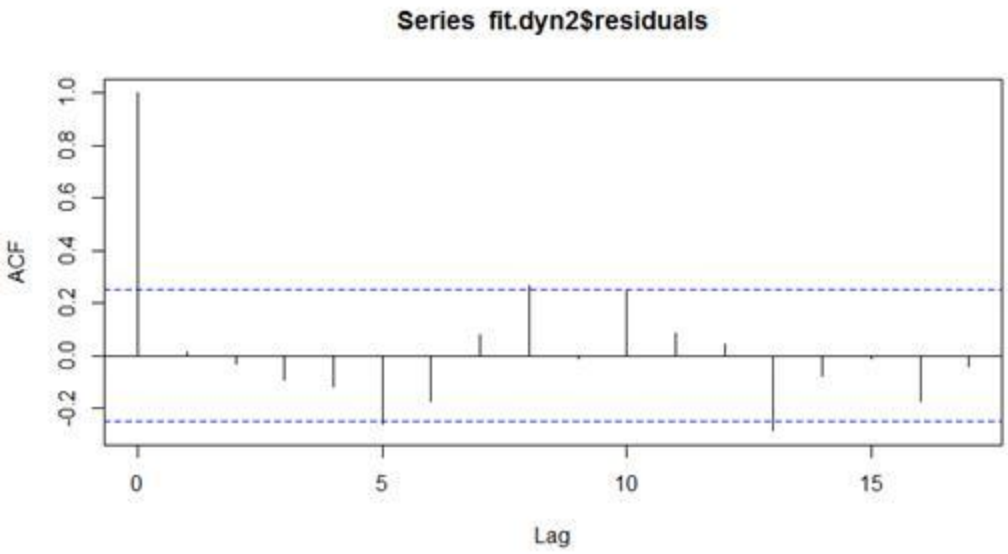

Forecasts from ARIMA(2,1,0)
-0.2
0.0
0.2
0.4
0.6
0.8
1970
1980
1990
2000
2010

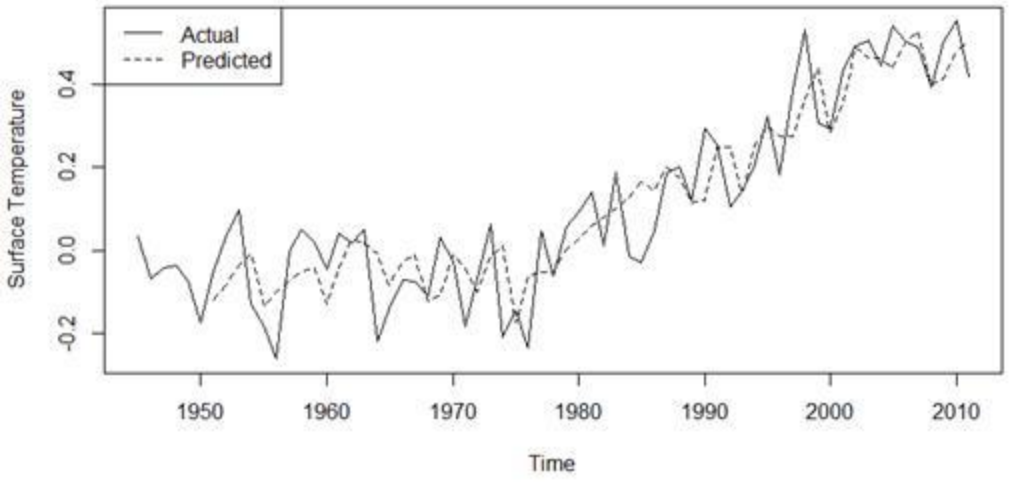

fit.lm$residuals
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
0
10
20
30
40
50
60
Time

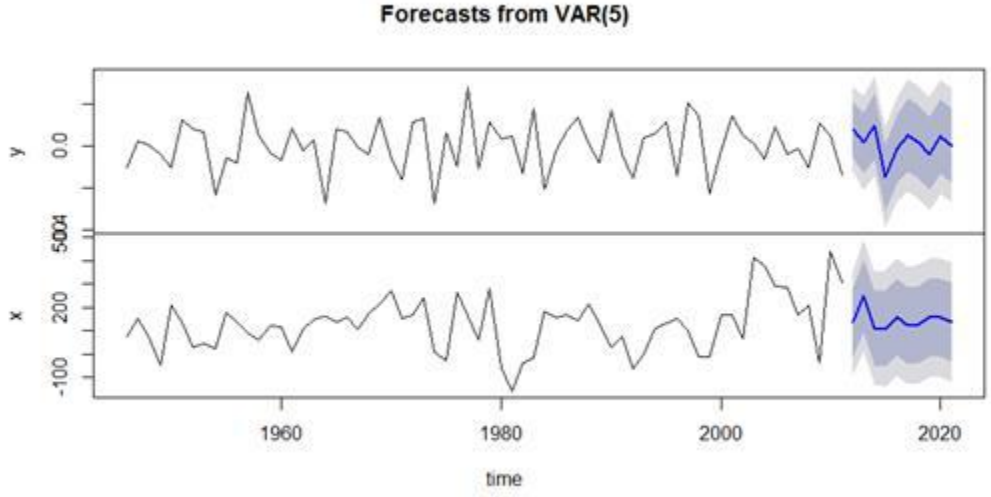

Series fit.lm$residuals
ACF
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0
5
10
15
Lag


x & y
ACF
0.8
0.6
0.4
0.2
0.0
-0.2
-15
-10
-5
0
5
10
15
Lag

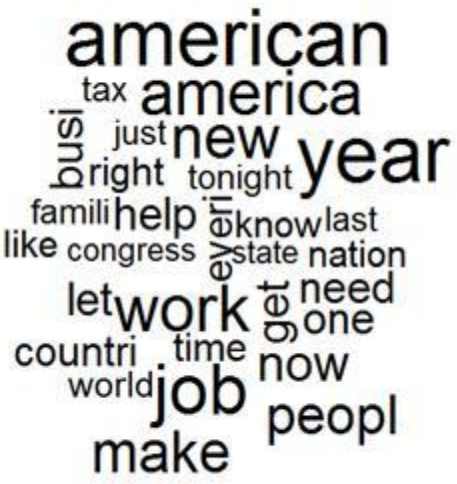

fit.dyn2$residuals
0.1
0.0
-0.1
-0.2
1950
1960
1970
1980
1990
2000
2010
Time

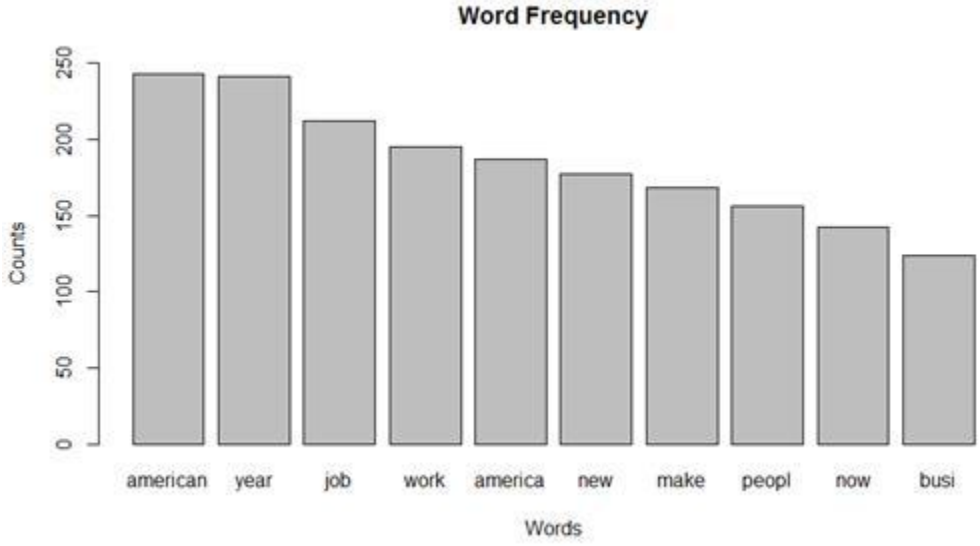

Series fit.dyn2$residuals
ACF
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0
5
10
15
Lag

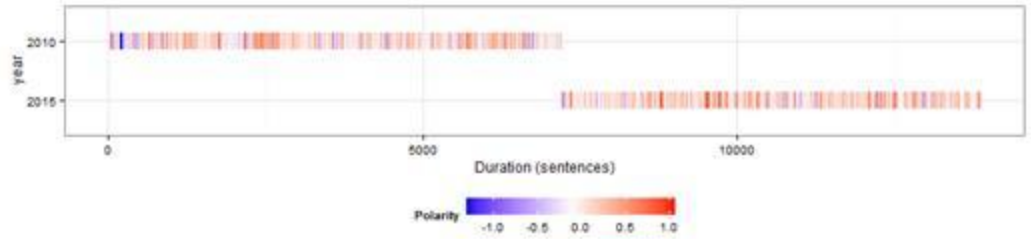

Actual
Predicted
Surface Temperature
-0.2
0.0
0.2
0.4
1950
1960
1970
1980
1990
2000
2010
Time

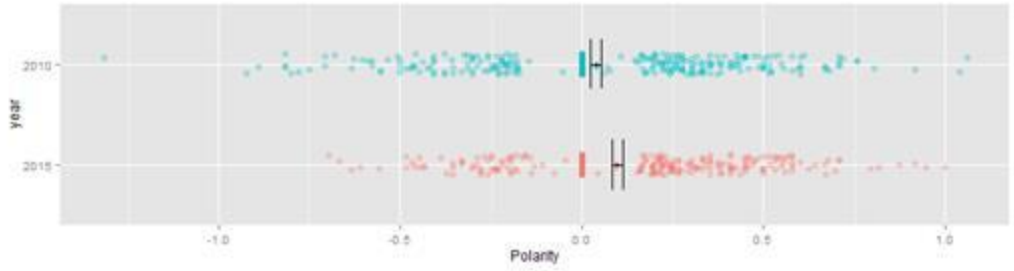

Forecasts from VAR(5)
y
0.0
x
-100
200
500
1960
1980
2000
2020
time

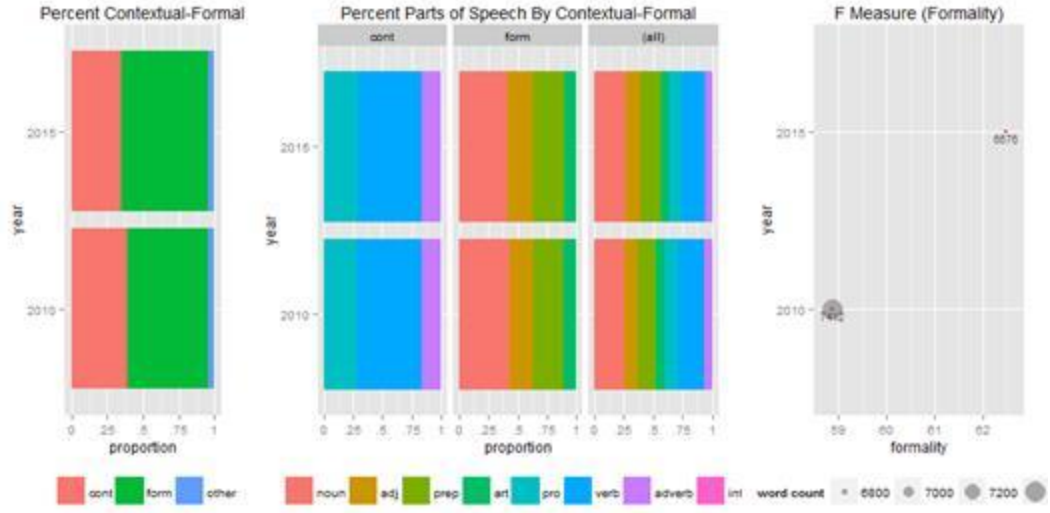
Chapter 12: Text Mining

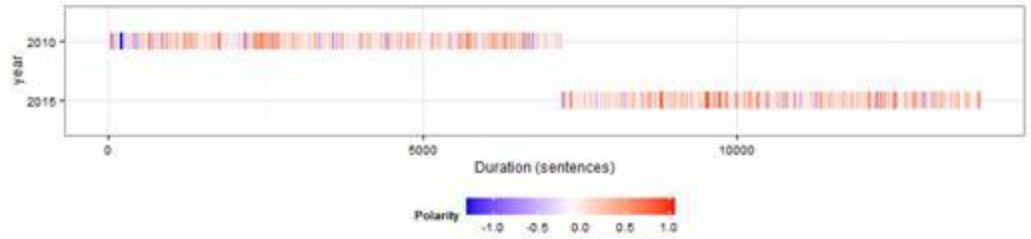

american
tax
america
busi
just
new
year
right
tonight
famili
help
everi
know
last
like
congress
state
nation
let
work
get
need
one
countri
time
now
world
job
peopl
make


Word Frequency
Counts
0
50
100
150
200
250
american
year
job
work
america
new
make
peopl
now
busi
Words

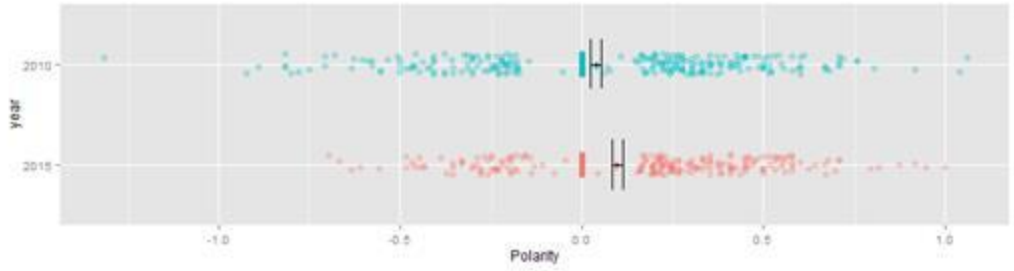

year
2010
2015
0
5000
10000
Duration (sentences)
Polarity
-1.0
-0.5
0.0
0.5
1.0

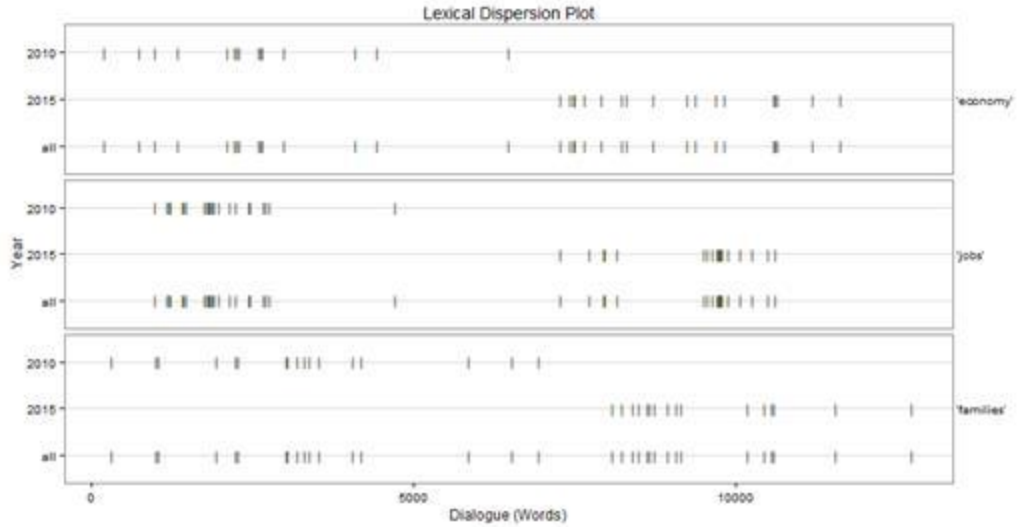

year
2010
2015
-1.0
-0.5
0.0
0.5
1.0
Polarity


Percent Contextual-Formal
year
2015
2010
0
.25
.5
.75
1
proportion
cont
form
other
Percent Parts of Speech By Contextual-Formal
cont
form
(all)
noun
adj
prep
art
pro
verb
adverb
int
F Measure (Formality)
6676
6612
59
60
61
62
formality
word count
6600
7000
7200


year
2010
2015
0
5000
10000
Duration (sentences)

Polarity
-1.0
-0.5
0.0
0.5
1.0


year
2010
2015
-1.0
-0.5
0.0
0.5
1.0
Polarity


Lexical Dispersion Plot
Year
2010
2015
all
'economy'
'jobs'
'families'
0
5000
10000
Dialogue (Words)

Appendix: R Fundamentals

[Home]

Download

CRAN

R Project

About R
Contributors
What's New?
Mailing Lists
Bug Tracking
Conferences
Search

R Foundation

Foundation
Board
Members
Donors
Donate

The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To **download R**, please choose your preferred CRAN mirror.

If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

News

- **R version 3.2.2 (Fire Safety)** has been released on 2015-08-14.
- **The R Journal Volume 7/1** is available.
- **R version 3.1.3 (Smooth Sidewalk)** has been released on 2015-03-09.
- **useR! 2015**, will take place at the University of Aalborg, Denmark, June 30 - July 3, 2015.
- **useR! 2014**, took place at the University of California, Los Angeles, USA June 30 - July 3, 2014.

CRAN Mirrors

The Comprehensive R Archive Network is available at the following URLs, please choose a location close to you. Some statistics on the status of the mirrors can be found here: main page, windows release, windows old release.

0-Cloud
https://cran.rstudio.com/ — Rstudio, automatic redirection to servers worldwide
http://cran.rstudio.com/ — Rstudio, automatic redirection to servers worldwide

Algeria
http://cran.usthb.dz/ — University of Science and Technology Houari Boumediene

Argentina
http://mirror.fcaglp.unlp.edu.ar/CRAN/ — Universidad Nacional de La Plata

Australia
http://cran.csiro.au/ — CSIRO
http://cran.ms.unimelb.edu.au/ — University of Melbourne

USA

https://cran.cnr.Berkeley.edu/	University of California, Berkeley, CA
http://cran.cnr.Berkeley.edu/	University of California, Berkeley, CA
http://cran.stat.ucla.edu/	University of California, Los Angeles, CA
http://cran.mirrorcatalogs.com/	Qarea Inc.
http://mirror.las.iastate.edu/CRAN/	Iowa State University, Ames, IA
http://ftp.ussg.iu.edu/CRAN/	Indiana University
https://rweb.crmda.ku.edu/cran/	University of Kansas, Lawrence, KS
http://rweb.crmda.ku.edu/cran/	University of Kansas, Lawrence, KS
http://watson.nci.nih.gov/cran_mirror/	National Cancer Institute, Bethesda, MD
https://cran.mtu.edu/	Michigan Technological University, Houghton, MI

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- Download R for Linux
- Download R for (Mac) OS X
- Download R for Windows

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2015-08-14, Fire Safety) R-3.2.2.tar.gz, read what's new in the latest version.
- Sources of R alpha and beta releases (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are available here. Please read about new features and bug fixes before filing corresponding feature requests or bug reports.
- Source code of older versions of R is available here.
- Contributed extension packages

Questions About R

- If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

R for Windows

Subdirectories:

base	Binaries for base distribution (managed by Duncan Murdoch). This is what you want to **install R for the first time**.
contrib	Binaries of contributed packages (managed by Uwe Ligges). There is also information on third party software available for CRAN Windows services and corresponding environment and make variables.
Rtools	Tools to build R and R packages (managed by Duncan Murdoch). This is what you want to build your own packages on Windows, or to build R itself.

Please do not submit binaries to CRAN. Package developers might want to contact Duncan Murdoch or Uwe Ligges directly in case of questions / suggestions related to Windows binaries.

You may also want to read the R FAQ and R for Windows FAQ.

Note: CRAN does some checks on these binaries for viruses, but cannot give guarantees. Use the normal precautions with downloaded executables.

R-3.2.2 for Windows (32/64 bit)

Download R 3.2.2 for Windows (62 megabytes, 32/64 bit)

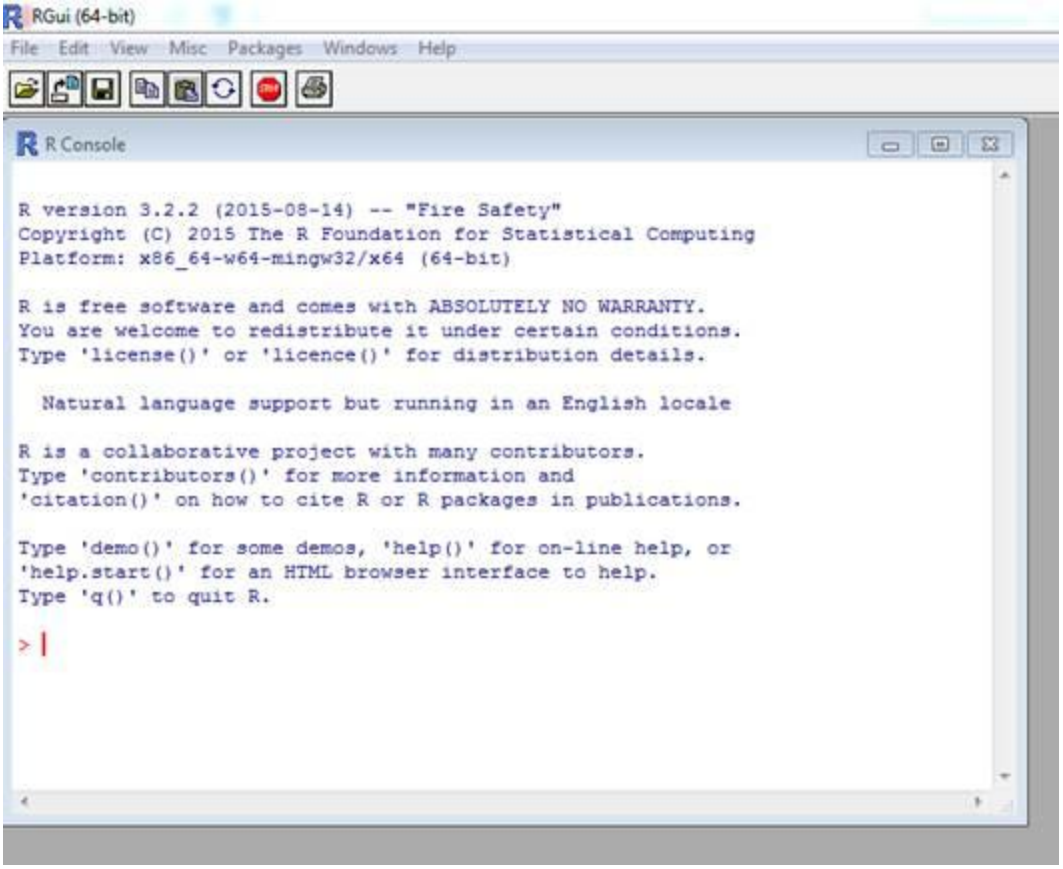
Installation and other instructions

New features in this version

If you want to double-check that the package you have downloaded exactly matches the package distributed by R, you can compare the md5sum of the .exe to the true fingerprint. You will need a version of md5sum for windows: both graphical and command line versions are available.

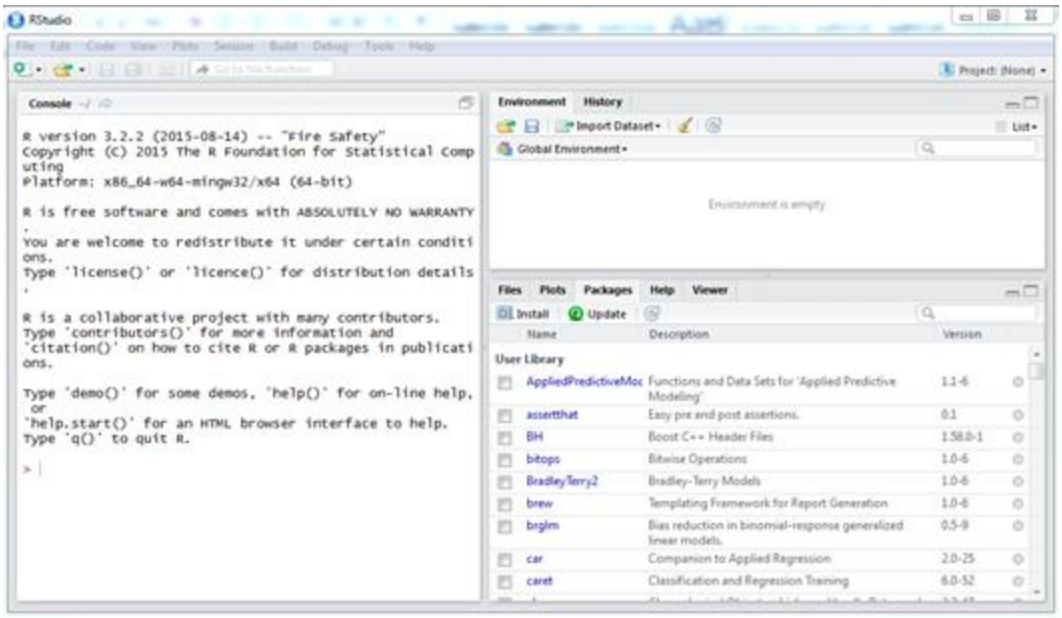
Frequently asked questions

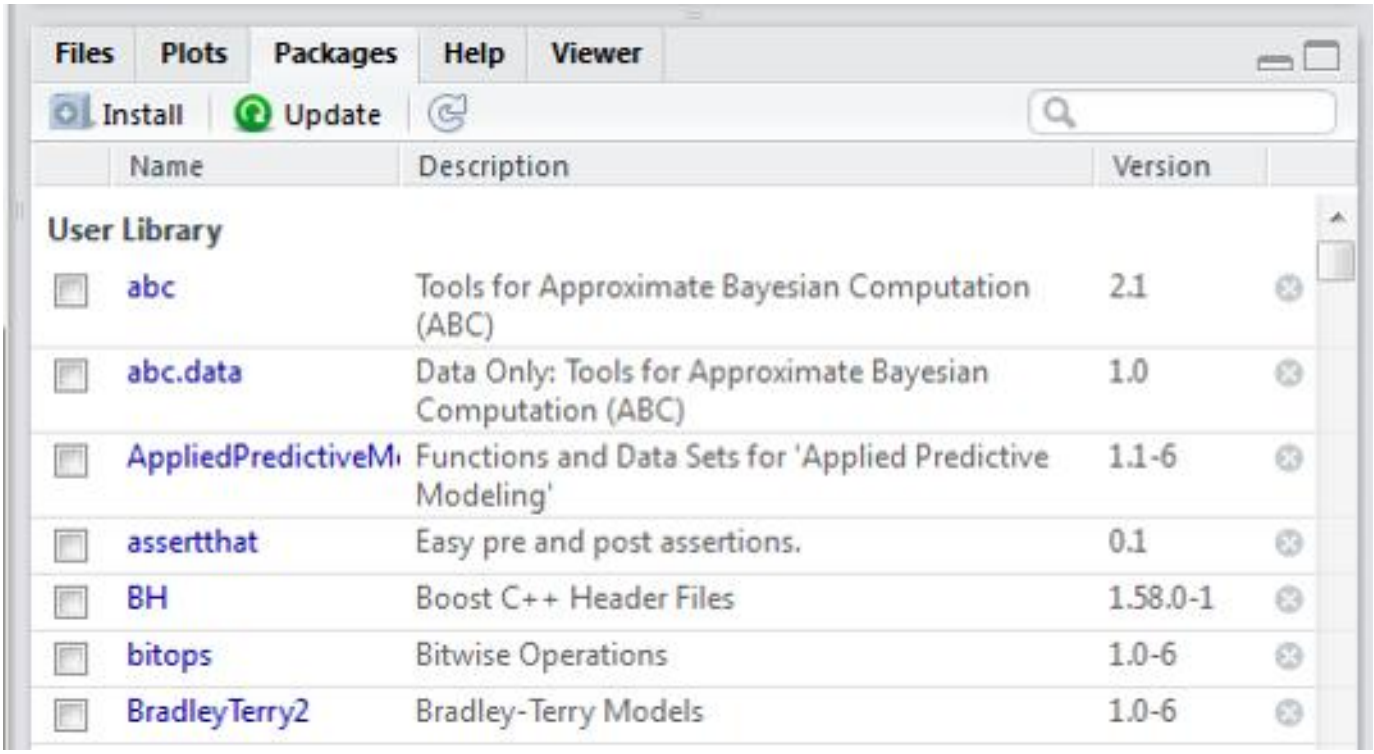
- Does R run under my version of Windows?
- How do I update packages in my previous version of R?
- Should I run 32-bit or 64-bit R?

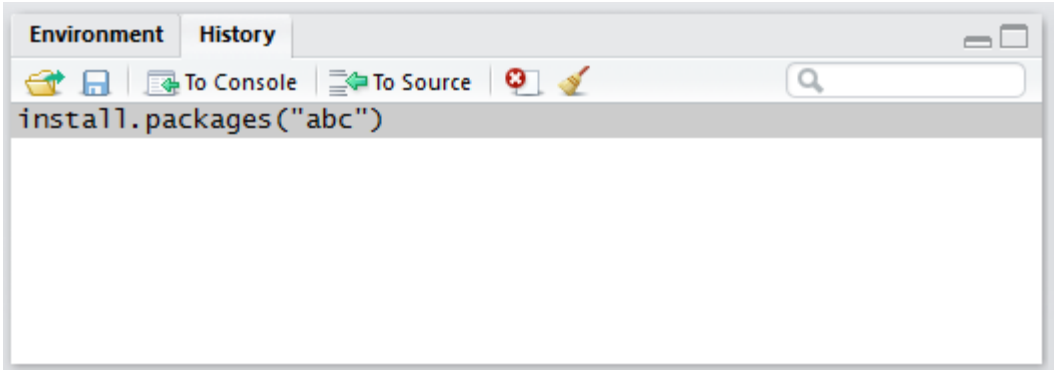


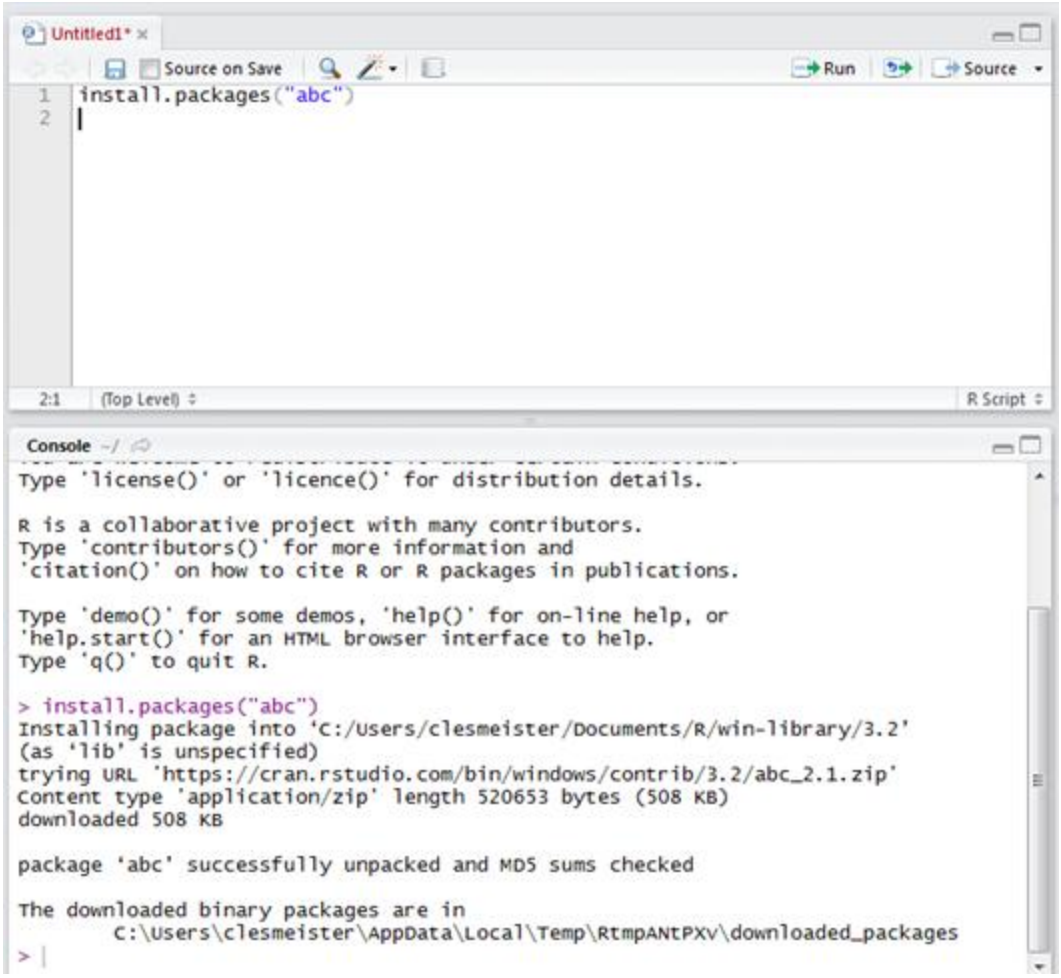
RStudio Desktop

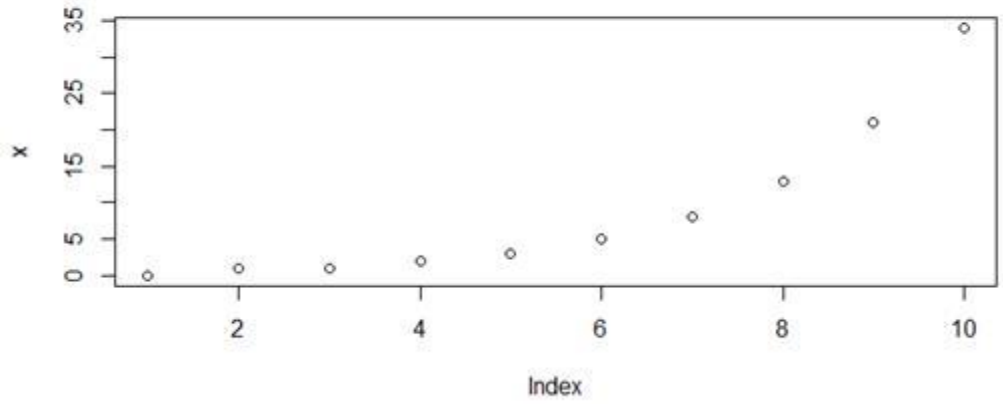
	Open Source Edition	Commercial License
Overview	• Access RStudio locally • Syntax highlighting, code completion, and smart indentation • Execute R code directly from the source editor • Quickly jump to function definitions • Easily manage multiple working directories using projects • Integrated R help and documentation • Interactive debugger to diagnose and fix errors quickly • Extensive package development tools	All of the features of open source, plus: • A commercial license for organizations not able to use AGPL software • Access to priority support
Support	Community forums only	• Priority Email Support • 8 hour response during business hours (ET)
License	AGPL v3	RStudio License Agreement
Pricing	Free	$995/year
	DOWNLOAD RSTUDIO DESKTOP	BUY NOW

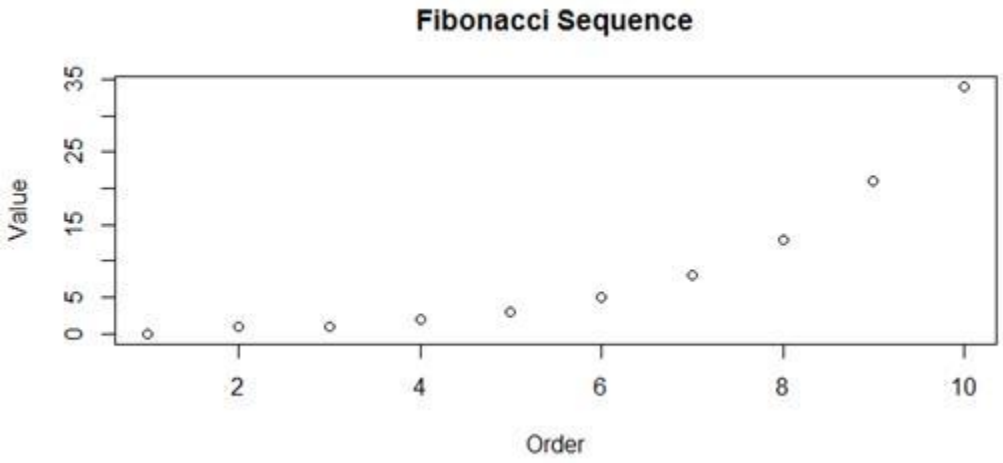

RStudio
Console
R version 3.2.2 (2015-08-14) -- "Fire Safety"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
Environment
History
Import Dataset
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
Install
Update
Name
Description
Version
User Library
AppliedPredictiveMod
Functions and Data Sets for 'Applied Predictive Modeling'
1.1-6
assertthat
Easy pre and post assertions.
0.1
BH
Boost C++ Header Files
1.58.0-1
bitops
Bitwise Operations
1.0-6
BradleyTerry2
Bradley-Terry Models
1.0-6
brew
Templating Framework for Report Generation
1.0-6
brglm
Bias reduction in binomial-response generalized linear models.
0.5-9
car
Companion to Applied Regression
2.0-25
caret
Classification and Regression Training
6.0-52

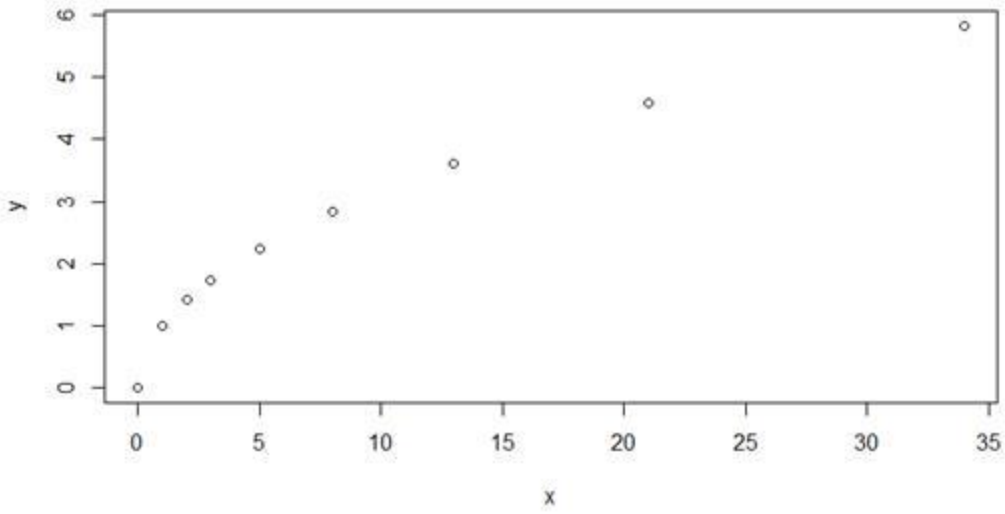

Files
Plots
Packages
Help
Viewer
Install
Update
Name
Description
Version
User Library
abc
Tools for Approximate Bayesian Computation (ABC)
2.1
abc.data
Data Only: Tools for Approximate Bayesian Computation (ABC)
1.0
AppliedPredictiveM
Functions and Data Sets for 'Applied Predictive Modeling'
1.1-6
assertthat
Easy pre and post assertions.
0.1
BH
Boost C++ Header Files
1.58.0-1
bitops
Bitwise Operations
1.0-6
BradleyTerry2
Bradley-Terry Models
1.0-6

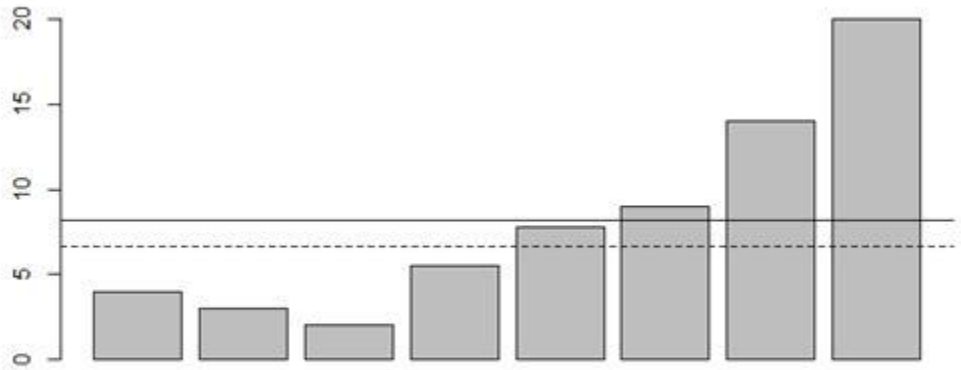

Environment
History
To Console
To Source
install.packages("abc")

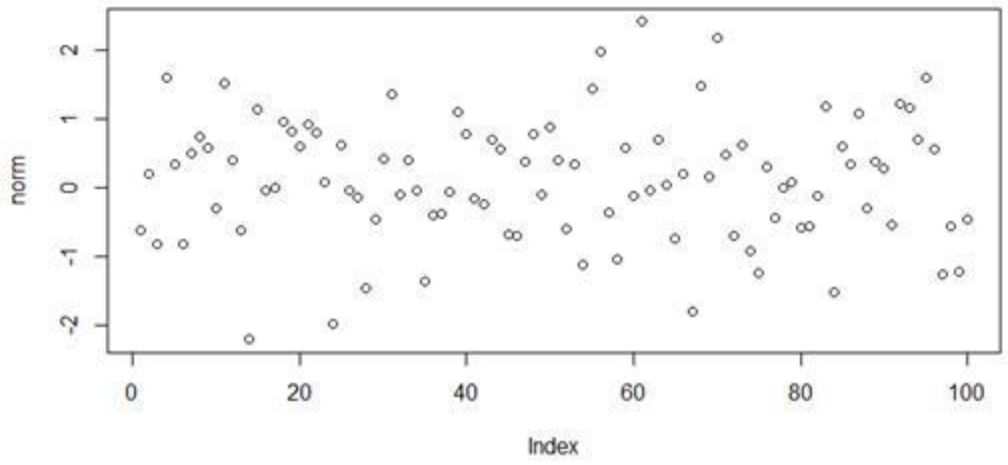

Untitled1*
Source on Save
Run
Source
1 install.packages("abc")
2
2:1
(Top Level)
R Script
Console
Type 'license()' or 'licence()' for distribution details.
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
> install.packages("abc")
Installing package into 'C:/Users/clesmeister/Documents/R/win-library/3.2'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.2/abc_2.1.zip'
Content type 'application/zip' length 520653 bytes (508 KB)
downloaded 508 KB
package 'abc' successfully unpacked and MD5 sums checked
The downloaded binary packages are in
C:\Users\clesmeister\AppData\Local\Temp\RtmpANtPXv\downloaded_packages
>

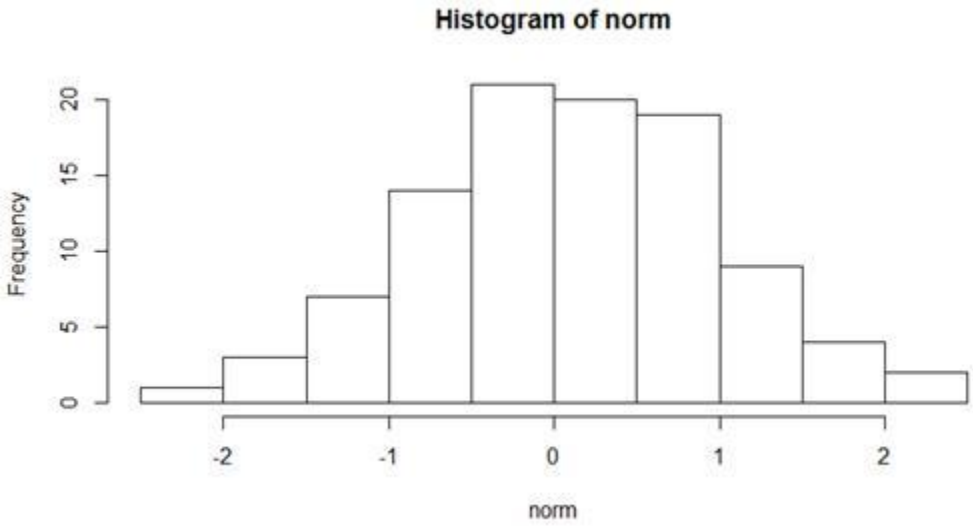

x
0
5
15
25
35
2
4
6
8
10
Index

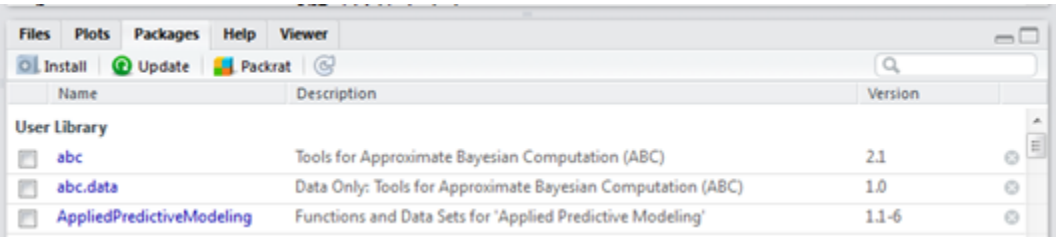

Fibonacci Sequence
Value
0
5
15
25
35
2
4
6
8
10
Order

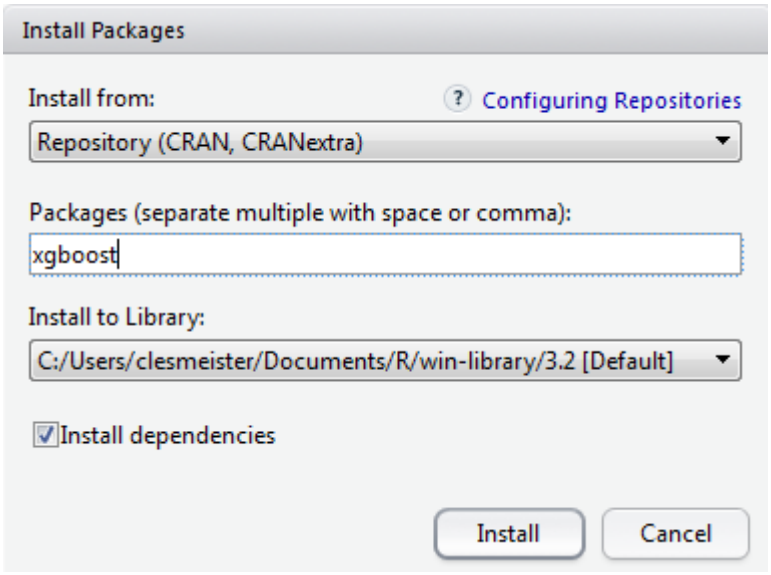

y
x
0
1
2
3
4
5
6
0
5
10
15
20
25
30
35


0
5
10
15
20

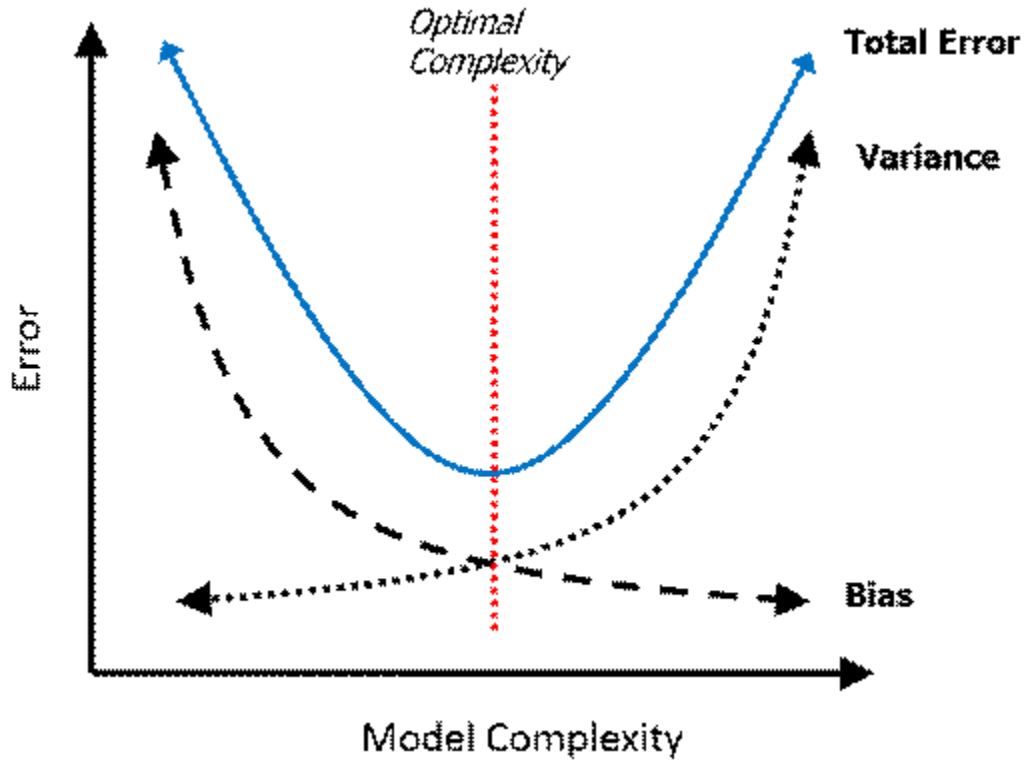

norm
2
1
0
-1
-2
0
20
40
60
80
100
Index


Histogram of norm
Frequency
20
15
10
5
0
-2
-1
0
1
2
norm


Files
Plots
Packages
Help
Viewer
Install
Update
Packrat
Name
Description
Version
User Library
abc
Tools for Approximate Bayesian Computation (ABC)
2.1
abc.data
Data Only: Tools for Approximate Bayesian Computation (ABC)
1.0
AppliedPredictiveModeling
Functions and Data Sets for 'Applied Predictive Modeling'
1.1-6


Install Packages
Install from:
Configuring Repositories
Repository (CRAN, CRANextra)
Packages (separate multiple with space or comma):
xgboost
Install to Library:
C:/Users/clesmeister/Documents/R/win-library/3.2 [Default]
Install dependencies
Install
Cancel

Preface


Optimal Complexity
Total Error
Variance
Error
Bias
Model Complexity