1. Prepare a classification model using SVM for salary data Data Description:

age -- age of a person

workclass -- A work class is a grouping of work

education -- Education of an individuals

maritalstatus -- Marital status of an

individulas occupation -- occupation of an

individuals

relationship –

race -- Race of an Individual

sex -- Gender of an Individual

capitalgain -- profit received from the sale of an investment

capitalloss -- A decrease in the value of a capital asset

hoursperweek -- number of hours work per week

native -- Native of an individual

Salary -- salary of an individual

classify the Size\_Categorie using SVM

month month of the year: 'jan' to 'dec'

day day of the week: 'mon' to 'sun'

FFMC FFMC index from

the FWI system: 18.7 to 96.20

DMC DMC index from the FWI

system: 1.1 to 291.3

DC DC index from the

fWI system: 7.9 to 860.6

ISI ISI index from the

FWI system: 0.0 to 56.10

temp temperature in Celsius

degrees: 2.2 to 33.30

RH relative humidity in %: 15.0 to 100

wind wind speed in km/h: 0.40 to 9.40

rain outside rain in mm/m2 : 0.0 to 6.4

Size\_Categorie the burned area of the forest ( Small , Large)

> library(readr)

> SalaryData\_Train\_1\_ <- read\_csv("E:/data science r studio/Assignment code 1/svm/SalaryData\_Train(1).csv")

> View(SalaryData\_Train\_1\_)

|  |
| --- |
| > head(SalaryData\_Train\_1\_)  # A tibble: 6 x 14  age workclass education educationno maritalstatus occupation relationship  *<dbl>* *<chr>* *<chr>* *<dbl>* *<chr>* *<chr>* *<chr>*  1 39 State-gov Bachelors 13 Never-married Adm-cleri… Not-in-fami…  2 50 Self-emp… Bachelors 13 Married-civ-… Exec-mana… Husband  3 38 Private HS-grad 9 Divorced Handlers-… Not-in-fami…  4 53 Private 11th 7 Married-civ-… Handlers-… Husband  5 28 Private Bachelors 13 Married-civ-… Prof-spec… Wife  6 37 Private Masters 14 Married-civ-… Exec-mana… Wife  # … with 7 more variables: race *<chr>*, sex *<chr>*, capitalgain *<dbl>*,  # capitalloss *<dbl>*, hoursperweek *<dbl>*, native *<chr>*, Salary *<chr>* |
|  |
| |  | | --- | | > | |

> #########3#####################SVM#############

> ###############salary train data###########

> ##svm train and test

> Salary\_train<-SalaryData\_Train\_1\_[1:21112,]

> Salary\_test<-SalaryData\_Train\_1\_[21113:30161,]

> View(SalaryData\_Train\_1\_$Salary)

> install.packages("klaR")

Installing package into ‘C:/Users/DELL/Documents/R/win-library/3.6’

(as ‘lib’ is unspecified)

also installing the dependencies ‘miniUI’, ‘rstudioapi’, ‘highr’, ‘classInt’, ‘labelled’, ‘combinat’, ‘questionr’

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/miniUI\_0.1.1.1.zip'

Content type 'application/zip' length 36567 bytes (35 KB)

downloaded 35 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/rstudioapi\_0.11.zip'

Content type 'application/zip' length 282204 bytes (275 KB)

downloaded 275 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/highr\_0.8.zip'

Content type 'application/zip' length 48776 bytes (47 KB)

downloaded 47 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/classInt\_0.4-2.zip'

Content type 'application/zip' length 90872 bytes (88 KB)

downloaded 88 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/labelled\_2.2.2.zip'

Content type 'application/zip' length 154352 bytes (150 KB)

downloaded 150 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/combinat\_0.0-8.zip'

Content type 'application/zip' length 43075 bytes (42 KB)

downloaded 42 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/questionr\_0.7.0.zip'

Content type 'application/zip' length 1936150 bytes (1.8 MB)

downloaded 1.8 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/klaR\_0.6-14.zip'

Content type 'application/zip' length 568708 bytes (555 KB)

downloaded 555 KB

package ‘miniUI’ successfully unpacked and MD5 sums checked

package ‘rstudioapi’ successfully unpacked and MD5 sums checked

package ‘highr’ successfully unpacked and MD5 sums checked

package ‘classInt’ successfully unpacked and MD5 sums checked

package ‘labelled’ successfully unpacked and MD5 sums checked

package ‘combinat’ successfully unpacked and MD5 sums checked

package ‘questionr’ successfully unpacked and MD5 sums checked

package ‘klaR’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\DELL\AppData\Local\Temp\RtmpugGBGw\downloaded\_packages

> library(kernlab)

Error in library(kernlab) : there is no package called ‘kernlab’

> install.packages("kernlab")

Installing package into ‘C:/Users/DELL/Documents/R/win-library/3.6’

(as ‘lib’ is unspecified)

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/kernlab\_0.9-29.zip'

Content type 'application/zip' length 2641401 bytes (2.5 MB)

downloaded 2.5 MB

package ‘kernlab’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\DELL\AppData\Local\Temp\RtmpugGBGw\downloaded\_packages

> library(kernlab)

> library(caret)

Loading required package: lattice

Loading required package: ggplot2

Attaching package: ‘ggplot2’

The following object is masked from ‘package:kernlab’:

alpha

> model1<-ksvm(Salary ~.,data = Salary\_train,kernel = "vanilladot")

Setting default kernel parameters

> model1

Support Vector Machine object of class "ksvm"

SV type: C-svc (classification)

parameter : cost C = 1

Linear (vanilla) kernel function.

Number of Support Vectors : 7460

Objective Function Value : -7399.792

Training error : 0.153751

> # kernel = rfdot

> model\_rfdot<-ksvm(Salary ~.,data = Salary\_train,kernel = "rbfdot")

> pred\_rfdot<-predict(model\_rfdot,newdata=Salary\_test)

> mean(pred\_rfdot==SalaryData\_Train\_1\_$Salary)

[1] 0.648818

Warning messages:

1: In `==.default`(pred\_rfdot, SalaryData\_Train\_1\_$Salary) :

longer object length is not a multiple of shorter object length

2: In is.na(e1) | is.na(e2) :

longer object length is not a multiple of shorter object length

> # kernal = besseldot

|  |
| --- |
| > # kernel = vanilladot  > model\_vanilla<-ksvm(Salary ~.,data = Salary\_train,kernel = "vanilladot")  Setting default kernel parameters  > pred\_vanilla<-predict(model\_vanilla,newdata=Salary\_test)  > mean(pred\_vanilla==SalaryData\_Train\_1\_$Salary)  [1] 0.6512384  Warning messages:  1: In `==.default`(pred\_vanilla, SalaryData\_Train\_1\_$Salary) :  longer object length is not a multiple of shorter object length  2: In is.na(e1) | is.na(e2) :  longer object length is not a multiple of shorter object length  > # kernal = besseldot  > # kernal = besseldot  > model\_besseldot<-ksvm(Salary ~.,data = Salary\_train,kernel = "besseldot")  Setting default kernel parameters |
|  |
| |  |  |  |  |  | | --- | --- | --- | --- | --- | | >   |  | | --- | | > pred\_bessel<-predict(model\_besseldot,newdata=Salary\_test)  > mean(pred\_bessel==SalaryData\_Train\_1\_$Salary)  [1] 0.6316767  Warning messages:  1: In `==.default`(pred\_bessel, SalaryData\_Train\_1\_$Salary) :  longer object length is not a multiple of shorter object length  2: In is.na(e1) | is.na(e2) :  longer object length is not a multiple of shorter object length | |  | | |  | | --- | | > | |   > model\_poly<-ksvm(Salary ~.,data = Salary\_train,kernel = "polydot")  Setting default kernel parameters  > pred\_poly<-predict(model\_poly,newdata = Salary\_test)  > mean(pred\_poly==SalaryData\_Train\_1\_$Salary)  [1] 0.6512384  Warning messages:  1: In `==.default`(pred\_poly, SalaryData\_Train\_1\_$Salary) :  longer object length is not a multiple of shorter object length  2: In is.na(e1) | is.na(e2) :  longer object length is not a multiple of shorter object length  > | |

2)salary test data

> ###################Salary of test data##########

> library(readr)

> SalaryData\_Test\_1\_ <- read\_csv("E:/data science r studio/Assignment code 1/svm/SalaryData\_Test(1).csv")

Parsed with column specification:

cols(

age = col\_double(),

workclass = col\_character(),

education = col\_character(),

educationno = col\_double(),

maritalstatus = col\_character(),

occupation = col\_character(),

relationship = col\_character(),

race = col\_character(),

sex = col\_character(),

capitalgain = col\_double(),

capitalloss = col\_double(),

hoursperweek = col\_double(),

native = col\_character(),

Salary = col\_character()

)

|===============================================================| 100% 1 MB

> head(SalaryData\_Test\_1\_)

# A tibble: 6 x 14

age workclass education educationno maritalstatus occupation relationship

*<dbl>* *<chr>* *<chr>* *<dbl>* *<chr>* *<chr>* *<chr>*

1 25 Private 11th 7 Never-married Machine-o… Own-child

2 38 Private HS-grad 9 Married-civ-… Farming-f… Husband

3 28 Local-gov Assoc-ac… 12 Married-civ-… Protectiv… Husband

4 44 Private Some-col… 10 Married-civ-… Machine-o… Husband

5 34 Private 10th 6 Never-married Other-ser… Not-in-fami…

6 63 Self-emp… Prof-sch… 15 Married-civ-… Prof-spec… Husband

# … with 7 more variables: race *<chr>*, sex *<chr>*, capitalgain *<dbl>*,

# capitalloss *<dbl>*, hoursperweek *<dbl>*, native *<chr>*, Salary *<chr>*

|  |
| --- |
| > ##svm train and test  > Salary\_train<-SalaryData\_Test\_1\_[1:1092,]  > Salary\_test<-SalaryData\_Test\_1\_[1093:15060,]  > View(SalaryData\_Test\_1\_$Salary) |
|  |
| |  | | --- | | > | |
| > library(kernlab)  Attaching package: ‘kernlab’  The following object is masked from ‘package:ggplot2’:  alpha  > model1<-ksvm(Salary ~.,data = Salary\_train,kernel = "vanilladot")  Setting default kernel parameters  > model1  Support Vector Machine object of class "ksvm"  SV type: C-svc (classification)  parameter : cost C = 1  Linear (vanilla) kernel function.  Number of Support Vectors : 3725  Objective Function Value : -3666.302  Training error : 0.15065 |
|  |
| |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | > > # kernel = rfdot  > model\_rfdot<-ksvm(Salary ~.,data = Salary\_train,kernel = "rbfdot")  > pred\_rfdot<-predict(model\_rfdot,newdata=Salary\_test)  > mean(pred\_rfdot==SalaryData\_Test\_1\_$Salary)  [1] 0.6603586  Warning messages:  1: In `==.default`(pred\_rfdot, SalaryData\_Test\_1\_$Salary) :  longer object length is not a multiple of shorter object length  2: In is.na(e1) | is.na(e2) :  longer object length is not a multiple of shorter object length  > # kernel = vanilladot  > model\_vanilla<-ksvm(Salary ~.,data = Salary\_train,kernel = "vanilladot")  Setting default kernel parameters  > pred\_vanilla<-predict(model\_vanilla,newdata=Salary\_test)  > mean(pred\_vanilla==SalaryData\_Test\_1\_$Salary)  [1] 0.6539841  Warning messages:  1: In `==.default`(pred\_vanilla, SalaryData\_Test\_1\_$Salary) :  longer object length is not a multiple of shorter object length  2: In is.na(e1) | is.na(e2) :  longer object length is not a multiple of shorter object length  > # kernal = besseldot  > model\_besseldot<-ksvm(Salary ~.,data = Salary\_train,kernel = "besseldot")  Setting defau  > pred\_bessel<-predict(model\_besseldot,newdata=Salary\_test)  > mean(pred\_bessel==SalaryData\_Test\_1\_$Salary)#64.70  [1] 0.647012  Warning messages:  1: In `==.default`(pred\_bessel, SalaryData\_Test\_1\_$Salary) :  longer object length is not a multiple of shorter object length  2: In is.na(e1) | is.na(e2) :  longer object length is not a multiple of shorter object length  > model\_poly<-ksvm(Salary ~.,data = Salary\_train,kernel = "polydot")  Setting default kernel parameters   |  | | --- | | > model\_poly<-ksvm(Salary ~.,data = Salary\_train,kernel = "polydot")  Setting default kernel parameters  > pred\_poly<-predict(model\_poly,newdata = Salary\_test)  > mean(pred\_poly==SalaryData\_Test\_1\_$Salary) #  [1] 0.6539841  Warning messages:  1: In `==.default`(pred\_poly, SalaryData\_Test\_1\_$Salary) :  longer object length is not a multiple of shorter object length  2: In is.na(e1) | is.na(e2) :  longer object length is not a multiple of shorter object length | |  | | |  | | --- | | > | |   3)forestfires   |  | | --- | | > library(readr)  > forestfires <- read\_csv("E:/data science r studio/Assignment code 1/svm/forestfires.csv")  Parsed with column specification:  cols(  .default = col\_double(),  month = col\_character(),  day = col\_character(),  size\_category = col\_character()  )  See spec(...) for full column specifications.  > View(forestfires) | |  | | |  | | --- | | > | | | > #svn train and test  > forestfires\_train<-forestfires[1:361,]  > forestfires\_test<-forestfires[362:517,] | |  | | |  | | --- | | > | | | |