1. SVM, Random Forest and Boosting Models

Code:

```
Train.data=read.csv("WineData.csv",header=T)
Test.data=read.csv("WineHoldoutData.csv",header=T)

summary(Train.data)

##  fixed_acidity    volatile_acidity  citric_acid     residual_sugar
##  Min.   : 3.800   Min.   :0.0800   Min.   :0.0000   Min.   : 0.60
##  1st Qu.: 6.400   1st Qu.:0.2300   1st Qu.:0.2500   1st Qu.: 1.80
##  Median : 7.000   Median :0.2900   Median :0.3100   Median : 3.00
##  Mean   : 7.212   Mean   :0.3418   Mean   :0.3184   Mean   : 5.43
##  3rd Qu.: 7.700   3rd Qu.:0.4100   3rd Qu.:0.3900   3rd Qu.: 8.10
##  Max.   :15.900   Max.   :1.3300   Max.   :1.6600   Max.   :65.80
##    chlorides        free_sulfur_dioxide total_sulfur_dioxide    dens
ity
##  Min.   :0.00900   Min.   :  1.0       Min.   :  6.0         Min.
:0.9871
##  1st Qu.:0.03800   1st Qu.: 17.0       1st Qu.: 78.0         1st Qu.
:0.9924
##  Median :0.04700   Median : 29.0       Median :118.0         Median
:0.9949
##  Mean   :0.05619   Mean   : 30.4       Mean   :115.3         Mean
:0.9947
##  3rd Qu.:0.06600   3rd Qu.: 41.0       3rd Qu.:155.0         3rd Qu.
:0.9969
##  Max.   :0.61000   Max.   :289.0       Max.   :440.0         Max.
:1.0390
##        pH          sulphates         alcohol          quality
##  Min.   :2.74   Min.   :0.220   Min.   : 8.0   Min.   :3.000
##  1st Qu.:3.11   1st Qu.:0.430   1st Qu.: 9.5   1st Qu.:5.000
##  Median :3.21   Median :0.510   Median :10.3   Median :6.000
##  Mean   :3.22   Mean   :0.531   Mean   :10.5   Mean   :5.812
##  3rd Qu.:3.32   3rd Qu.:0.600   3rd Qu.:11.3   3rd Qu.:6.000
##  Max.   :4.01   Max.   :2.000   Max.   :14.9   Max.   :9.000
##     style
##  Length:5198
##  Class :character
##  Mode  :character
##
##
##

summary(Test.data)
```

```
##  fixed_acidity    volatile_acidity  citric_acid     residual_sugar
##  Min.   : 4.200   Min.   :0.100    Min.   :0.0000   Min.   : 0.800
##  1st Qu.: 6.400   1st Qu.:0.230    1st Qu.:0.2500   1st Qu.: 1.800
##  Median : 7.000   Median :0.290    Median :0.3100   Median : 3.000
##  Mean   : 7.229   Mean   :0.331    Mean   :0.3195   Mean   : 5.496
##  3rd Qu.: 7.700   3rd Qu.:0.390    3rd Qu.:0.3900   3rd Qu.: 8.200
##  Max.   :15.600   Max.   :1.580    Max.   :1.2300   Max.   :20.800
##    chlorides        free_sulfur_dioxide total_sulfur_dioxide    dens
ity
##  Min.   :0.01200   Min.   :  3.00      Min.   :  8.0        Min.
:0.9874
##  1st Qu.:0.03800   1st Qu.: 17.00      1st Qu.: 76.0        1st Qu.
:0.9922
##  Median :0.04700   Median : 29.00      Median :119.0        Median
:0.9950
##  Mean   :0.05543   Mean   : 31.01      Mean   :117.5        Mean
:0.9947
##  3rd Qu.:0.06100   3rd Qu.: 43.00      3rd Qu.:159.0        3rd Qu.
:0.9971
##  Max.   :0.61100   Max.   :118.50      Max.   :366.5        Max.
:1.0037
##        pH          sulphates         alcohol         quality
##  Min.   :2.720   Min.   :0.2700   Min.   : 8.40   Min.   :3.000
##  1st Qu.:3.100   1st Qu.:0.4300   1st Qu.: 9.50   1st Qu.:5.000
##  Median :3.200   Median :0.5100   Median :10.20   Median :6.000
##  Mean   :3.214   Mean   :0.5323   Mean   :10.46   Mean   :5.843
##  3rd Qu.:3.320   3rd Qu.:0.6000   3rd Qu.:11.30   3rd Qu.:6.000
##  Max.   :3.800   Max.   :1.9500   Max.   :14.05   Max.   :8.000
##     style
##  Length:1299
##  Class :character
##  Mode  :character
##
##
##
```

```
#Checking correlation between predictors
df=data.frame(Train.data[,c(1:12)])
cor(df)
```

```
##                      fixed_acidity volatile_acidity  citric_acid re
sidual_sugar
## fixed_acidity           1.00000000       0.21599868  0.316851216
-0.11599079
## volatile_acidity        0.21599868       1.00000000 -0.383379676
-0.19143517
```

```
## citric_acid             0.31685122      -0.38337968  1.000000000
0.14382003
## residual_sugar         -0.11599079      -0.19143517  0.143820029
1.00000000
## chlorides               0.29904393       0.38098335  0.040924485
-0.12957070
## free_sulfur_dioxide    -0.28158153      -0.35139672  0.134708918
0.39439450
## total_sulfur_dioxide   -0.33439763      -0.41453324  0.196411471
0.49149983
## density                 0.45499336       0.27477397  0.090671339
0.55323017
## pH                     -0.25613859       0.26688742 -0.327915413
-0.25667780
## sulphates               0.29903244       0.22864088  0.051192324
-0.18573401
## alcohol                -0.09043771      -0.04036961 -0.003289993
-0.35437668
## quality                -0.07735971      -0.26501969  0.091516750
-0.03042104
##                        chlorides free_sulfur_dioxide total_sulfur_d
ioxide
## fixed_acidity         0.29904393         -0.28158153          -0.33
439763
## volatile_acidity      0.38098335         -0.35139672          -0.41
453324
## citric_acid           0.04092449          0.13470892           0.19
641147
## residual_sugar       -0.12957070          0.39439450           0.49
149983
## chlorides             1.00000000         -0.20126300          -0.28
116864
## free_sulfur_dioxide  -0.20126300          1.00000000           0.72
192584
## total_sulfur_dioxide -0.28116864          0.72192584           1.00
000000
## density               0.36307630          0.01281186           0.02
432206
## pH                    0.04450119         -0.14912465          -0.23
854277
## sulphates             0.39161142         -0.19028761          -0.28
157204
## alcohol              -0.25560197         -0.16981465          -0.26
347313
## quality              -0.20197733          0.05365606          -0.04
099676
```

```
##                             density          pH     sulphates       al
cohol
## fixed_acidity          0.45499336 -0.25613859  0.2990324398 -0.09043
77083
## volatile_acidity       0.27477397  0.26688742  0.2286408812 -0.04036
96144
## citric_acid            0.09067134 -0.32791541  0.0511923244 -0.00328
99935
## residual_sugar         0.55323017 -0.25667780 -0.1857340053 -0.35437
66776
## chlorides              0.36307630  0.04450119  0.3916114213 -0.25560
19733
## free_sulfur_dioxide    0.01281186 -0.14912465 -0.1902876070 -0.16981
46470
## total_sulfur_dioxide  0.02432206 -0.23854277 -0.2815720444 -0.26347
31329
## density                1.00000000  0.02023940  0.2590991926 -0.68150
74023
## pH                     0.02023940  1.00000000  0.1940256211  0.11471
18896
## sulphates              0.25909919  0.19402562  1.0000000000 -0.00095
96393
## alcohol               -0.68150740  0.11471189 -0.0009596393  1.00000
00000
## quality               -0.30437481  0.02247475  0.0383116145  0.45030
66020
##                             quality
## fixed_acidity         -0.07735971
## volatile_acidity      -0.26501969
## citric_acid            0.09151675
## residual_sugar        -0.03042104
## chlorides             -0.20197733
## free_sulfur_dioxide    0.05365606
## total_sulfur_dioxide -0.04099676
## density               -0.30437481
## pH                     0.02247475
## sulphates              0.03831161
## alcohol                0.45030660
## quality                1.00000000

#Max correlation is between free sulphur dioxide and total sulphur dio
xide =0.72
```

Explanation:

From the summary of training and test data, it can be seen that there are no missing entries or any absurd values. Correlation matrix shows that all the predictors (except free sulfur dioxide and total sulfur dioxide) are linearly independent. Free sulfur dioxide and total sulfur dioxide have a maximum correlation of 0.72, but they are not removed from the dataset, since Boosting, Random Forest and SVM are not affected by collinearity like linear regression.

Boosting code:

```
#Boosting------------------------------------

library(caret)

boost.caretGrid=expand.grid(interaction.depth=c(3,5),n.trees=seq(from=
100,to=1000,by=50),
                            shrinkage=c(0.001,0.01,0.05,0.1),n.minobsi
nnode=c(100,200))
metric="RMSE"
trainControl=trainControl(method="cv",number=10)
set.seed(1)
gbm.caret=train(quality~., data=Train.data,method="gbm",
                trControl=trainControl, verbose=FALSE,
                tuneGrid=boost.caretGrid, metric=metric)

print(gbm.caret)

## Stochastic Gradient Boosting
##
## 5198 samples
##   12 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 4678, 4679, 4678, 4679, 4678, 4678, ...
## Resampling results across tuning parameters:
##
##   shrinkage  interaction.depth  n.minobsinnode  n.trees  RMSE       Rsquared
##   0.001      3                  100              100     0.8581667  0.2537761
##   0.001      3                  100              150     0.8497450  0.2556744
##   0.001      3                  100              200     0.8419646  0.2575327
##   0.001      3                  100              250     0.8347567  0.2594653
##   0.001      3                  100              300     0.8281291  0.2609338
##   0.001      3                  100              350     0.8219986  0.2622104
##   0.001      3                  100              400     0.8163367  0.2636886
##   0.001      3                  100              450     0.8110510  0.2654652
##   0.001      3                  100              500     0.8061354  0.2672129
##   0.001      3                  100              550     0.8015682  0.2686750
##   0.001      3                  100              600     0.7973308  0.2702253
##   0.001      3                  100              650     0.7933790  0.2716898
##   0.001      3                  100              700     0.7897003  0.2731013
##   0.001      3                  100              750     0.7862677  0.2744495
```

```
##   0.001      3                   100                800     0.7830713  0.2758017
##   0.001      3                   100                850     0.7800848  0.2771151
##   0.001      3                   100                900     0.7772735  0.2784129
##   0.001      3                   100                950     0.7746352  0.2796267
##   0.001      3                   100               1000     0.7722121  0.2806798
##   0.001      3                   200                100     0.8581065  0.2527020
##   0.001      3                   200                150     0.8496727  0.2546334
##   0.001      3                   200                200     0.8418884  0.2564503
##   0.001      3                   200                250     0.8346913  0.2579255
##   0.001      3                   200                300     0.8280128  0.2597249
##   0.001      3                   200                350     0.8218724  0.2612960
##   0.001      3                   200                400     0.8161900  0.2628651
##   0.001      3                   200                450     0.8109393  0.2642958
##   0.001      3                   200                500     0.8060584  0.2658225
##   0.001      3                   200                550     0.8015324  0.2670740
##   0.001      3                   200                600     0.7973493  0.2684179
##   0.001      3                   200                650     0.7934608  0.2697804
##   0.001      3                   200                700     0.7898422  0.2709634
##   0.001      3                   200                750     0.7864460  0.2722831
##   0.001      3                   200                800     0.7832784  0.2736226
##   0.001      3                   200                850     0.7803403  0.2748136
##   0.001      3                   200                900     0.7775956  0.2760282
##   0.001      3                   200                950     0.7750152  0.2772896
##   0.001      3                   200               1000     0.7725780  0.2785035
##   0.001      5                   100                100     0.8563690  0.2762736
##   0.001      5                   100                150     0.8471725  0.2769489
##   0.001      5                   100                200     0.8386787  0.2780040
##   0.001      5                   100                250     0.8308041  0.2794375
##   0.001      5                   100                300     0.8236117  0.2803962
##   0.001      5                   100                350     0.8169135  0.2817434
##   0.001      5                   100                400     0.8107420  0.2830053
##   0.001      5                   100                450     0.8050240  0.2844224
##   0.001      5                   100                500     0.7997346  0.2856655
##   0.001      5                   100                550     0.7948243  0.2870395
##   0.001      5                   100                600     0.7903028  0.2882434
##   0.001      5                   100                650     0.7860821  0.2895431
##   0.001      5                   100                700     0.7821528  0.2908775
##   0.001      5                   100                750     0.7784869  0.2923607
##   0.001      5                   100                800     0.7750240  0.2938434
##   0.001      5                   100                850     0.7717922  0.2952983
##   0.001      5                   100                900     0.7687864  0.2967845
##   0.001      5                   100                950     0.7659472  0.2982348
##   0.001      5                   100               1000     0.7632618  0.2997585
##   0.001      5                   200                100     0.8565232  0.2717616
##   0.001      5                   200                150     0.8473701  0.2723373
##   0.001      5                   200                200     0.8389763  0.2734746
##   0.001      5                   200                250     0.8311861  0.2747451
##   0.001      5                   200                300     0.8240343  0.2758255
##   0.001      5                   200                350     0.8174219  0.2769859
##   0.001      5                   200                400     0.8113172  0.2781582
##   0.001      5                   200                450     0.8056867  0.2793905
##   0.001      5                   200                500     0.8004818  0.2806165
##   0.001      5                   200                550     0.7956782  0.2818344
##   0.001      5                   200                600     0.7912230  0.2830327
##   0.001      5                   200                650     0.7870826  0.2843421
```

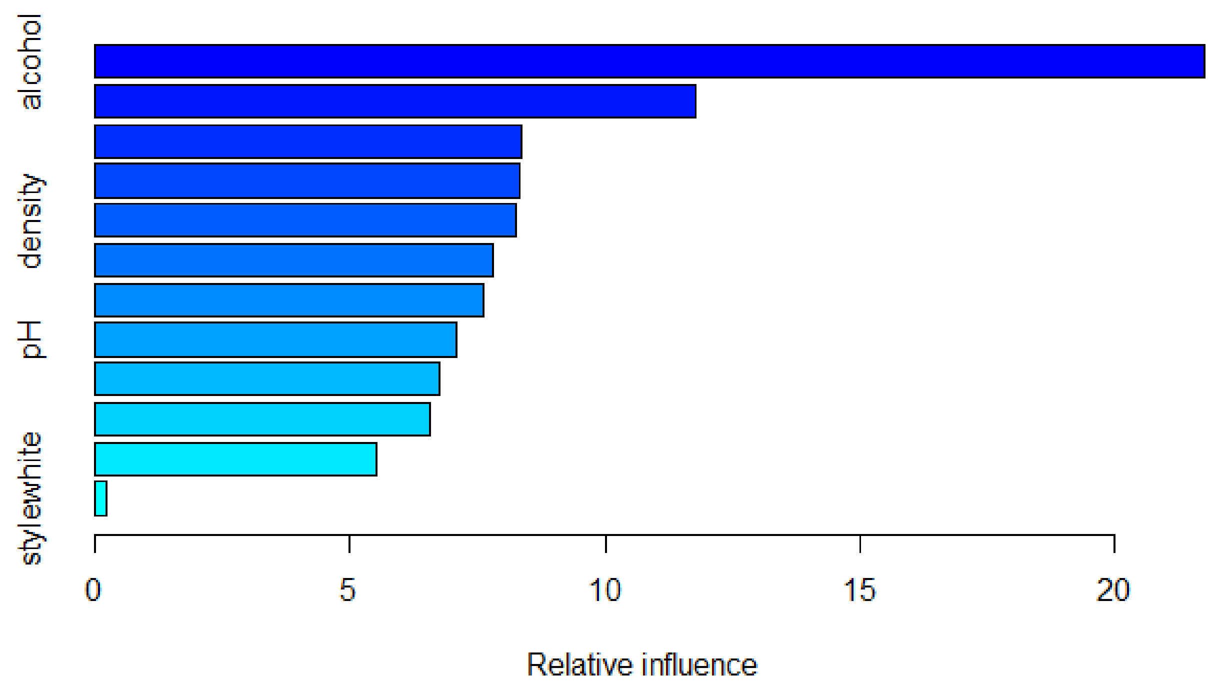
```
##   0.001       5                   200              700      0.7832260  0.2857349
##   0.001       5                   200              750      0.7796175  0.2872939
##   0.001       5                   200              800      0.7762538  0.2886603
##   0.001       5                   200              850      0.7731349  0.2899592
##   0.001       5                   200              900      0.7701682  0.2914777
##   0.001       5                   200              950      0.7673981  0.2928497
##   0.001       5                   200             1000      0.7648069  0.2942007
##   0.010       3                   100              100      0.7722378  0.2802282
##   0.010       3                   100              150      0.7541883  0.2918763
##   0.010       3                   100              200      0.7425068  0.3042441
##   0.010       3                   100              250      0.7339057  0.3154309
##   0.010       3                   100              300      0.7272511  0.3249672
##   0.010       3                   100              350      0.7222227  0.3322135
##   0.010       3                   100              400      0.7180833  0.3381343
##   0.010       3                   100              450      0.7147625  0.3430463
##   0.010       3                   100              500      0.7120017  0.3471731
##   0.010       3                   100              550      0.7095649  0.3509618
##   0.010       3                   100              600      0.7074002  0.3543133
##   0.010       3                   100              650      0.7056832  0.3569313
##   0.010       3                   100              700      0.7042933  0.3590240
##   0.010       3                   100              750      0.7030962  0.3609112
##   0.010       3                   100              800      0.7019227  0.3628360
##   0.010       3                   100              850      0.7008768  0.3645304
##   0.010       3                   100              900      0.6997609  0.3664986
##   0.010       3                   100              950      0.6988078  0.3681127
##   0.010       3                   100             1000      0.6979486  0.3695984
##   0.010       3                   200              100      0.7723865  0.2786929
##   0.010       3                   200              150      0.7545306  0.2898955
##   0.010       3                   200              200      0.7431223  0.3015987
##   0.010       3                   200              250      0.7348233  0.3123707
##   0.010       3                   200              300      0.7284737  0.3213351
##   0.010       3                   200              350      0.7233869  0.3289113
##   0.010       3                   200              400      0.7193678  0.3349121
##   0.010       3                   200              450      0.7162793  0.3393454
##   0.010       3                   200              500      0.7136847  0.3431829
##   0.010       3                   200              550      0.7115183  0.3464700
##   0.010       3                   200              600      0.7096751  0.3492797
##   0.010       3                   200              650      0.7081327  0.3516874
##   0.010       3                   200              700      0.7067797  0.3538644
##   0.010       3                   200              750      0.7056313  0.3557704
##   0.010       3                   200              800      0.7045060  0.3576930
##   0.010       3                   200              850      0.7035810  0.3592744
##   0.010       3                   200              900      0.7027199  0.3607779
##   0.010       3                   200              950      0.7019360  0.3620964
##   0.010       3                   200             1000      0.7011430  0.3634842
##   0.010       5                   100              100      0.7629906  0.3001882
##   0.010       5                   100              150      0.7427265  0.3147969
##   0.010       5                   100              200      0.7301406  0.3269330
##   0.010       5                   100              250      0.7209294  0.3382731
##   0.010       5                   100              300      0.7141570  0.3472246
##   0.010       5                   100              350      0.7093497  0.3537683
##   0.010       5                   100              400      0.7056485  0.3589093
##   0.010       5                   100              450      0.7026952  0.3631873
##   0.010       5                   100              500      0.7002330  0.3668612
##   0.010       5                   100              550      0.6982362  0.3699180
```

```
##   0.010      5                  100                600      0.6965224  0.3727231
##   0.010      5                  100                650      0.6951790  0.3748942
##   0.010      5                  100                700      0.6938305  0.3771220
##   0.010      5                  100                750      0.6926858  0.3790920
##   0.010      5                  100                800      0.6916729  0.3808392
##   0.010      5                  100                850      0.6908287  0.3823110
##   0.010      5                  100                900      0.6899242  0.3838980
##   0.010      5                  100                950      0.6891843  0.3851572
##   0.010      5                  100               1000      0.6885080  0.3863287
##   0.010      5                  200                100      0.7645890  0.2942197
##   0.010      5                  200                150      0.7452152  0.3074191
##   0.010      5                  200                200      0.7329850  0.3197250
##   0.010      5                  200                250      0.7241137  0.3305374
##   0.010      5                  200                300      0.7178349  0.3388740
##   0.010      5                  200                350      0.7129638  0.3457343
##   0.010      5                  200                400      0.7093664  0.3508879
##   0.010      5                  200                450      0.7065778  0.3551221
##   0.010      5                  200                500      0.7042671  0.3587447
##   0.010      5                  200                550      0.7025109  0.3616282
##   0.010      5                  200                600      0.7008780  0.3643663
##   0.010      5                  200                650      0.6995014  0.3667289
##   0.010      5                  200                700      0.6985470  0.3684168
##   0.010      5                  200                750      0.6975834  0.3700856
##   0.010      5                  200                800      0.6966868  0.3716917
##   0.010      5                  200                850      0.6957899  0.3732799
##   0.010      5                  200                900      0.6949960  0.3746784
##   0.010      5                  200                950      0.6942340  0.3760739
##   0.010      5                  200               1000      0.6936515  0.3771210
##   0.050      3                  100                100      0.7122997  0.3463652
##   0.050      3                  100                150      0.7030925  0.3607060
##   0.050      3                  100                200      0.6981429  0.3692634
##   0.050      3                  100                250      0.6946954  0.3753321
##   0.050      3                  100                300      0.6920731  0.3800363
##   0.050      3                  100                350      0.6906194  0.3824908
##   0.050      3                  100                400      0.6894312  0.3845999
##   0.050      3                  100                450      0.6876389  0.3876743
##   0.050      3                  100                500      0.6867382  0.3893886
##   0.050      3                  100                550      0.6856102  0.3913427
##   0.050      3                  100                600      0.6847189  0.3929588
##   0.050      3                  100                650      0.6837825  0.3946673
##   0.050      3                  100                700      0.6833197  0.3954656
##   0.050      3                  100                750      0.6829821  0.3960868
##   0.050      3                  100                800      0.6822524  0.3974018
##   0.050      3                  100                850      0.6814674  0.3987348
##   0.050      3                  100                900      0.6809360  0.3998020
##   0.050      3                  100                950      0.6803608  0.4006761
##   0.050      3                  100               1000      0.6800436  0.4012639
##   0.050      3                  200                100      0.7143470  0.3416080
##   0.050      3                  200                150      0.7067469  0.3535745
##   0.050      3                  200                200      0.7025793  0.3609238
##   0.050      3                  200                250      0.6997932  0.3657855
##   0.050      3                  200                300      0.6976194  0.3696308
##   0.050      3                  200                350      0.6963190  0.3720808
##   0.050      3                  200                400      0.6952079  0.3741005
##   0.050      3                  200                450      0.6941559  0.3759835
```

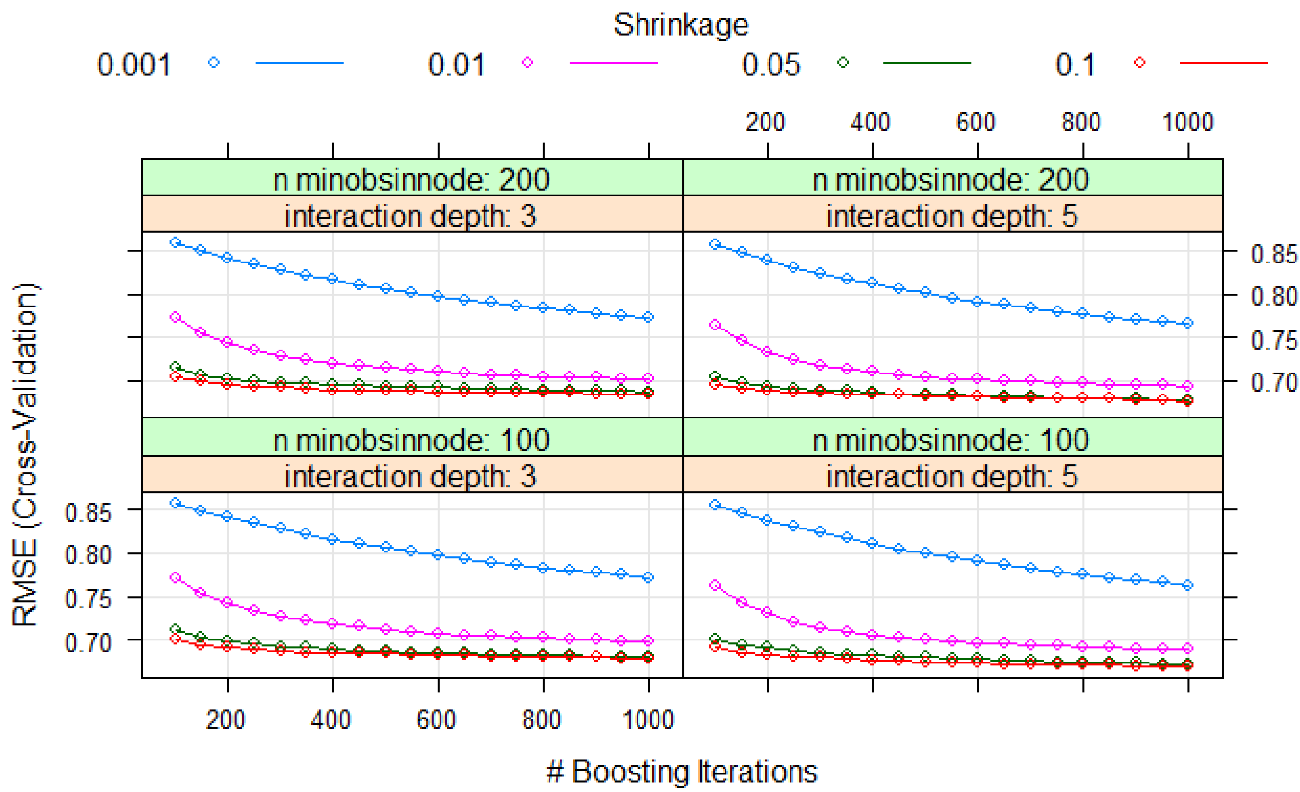
```
##   0.050       3                200                 500      0.6929360  0.3782144
##   0.050       3                200                 550      0.6922306  0.3795495
##   0.050       3                200                 600      0.6915904  0.3806939
##   0.050       3                200                 650      0.6904945  0.3824822
##   0.050       3                200                 700      0.6897224  0.3839647
##   0.050       3                200                 750      0.6896039  0.3842461
##   0.050       3                200                 800      0.6889760  0.3852737
##   0.050       3                200                 850      0.6882345  0.3865871
##   0.050       3                200                 900      0.6881165  0.3868132
##   0.050       3                200                 950      0.6874075  0.3881610
##   0.050       3                200                1000      0.6869976  0.3889328
##   0.050       5                100                 100      0.7012347  0.3646711
##   0.050       5                100                 150      0.6944334  0.3758453
##   0.050       5                100                 200      0.6905381  0.3827920
##   0.050       5                100                 250      0.6873113  0.3884733
##   0.050       5                100                 300      0.6854727  0.3917212
##   0.050       5                100                 350      0.6833814  0.3954141
##   0.050       5                100                 400      0.6819470  0.3978457
##   0.050       5                100                 450      0.6801365  0.4011938
##   0.050       5                100                 500      0.6795695  0.4021190
##   0.050       5                100                 550      0.6788439  0.4035914
##   0.050       5                100                 600      0.6772096  0.4064277
##   0.050       5                100                 650      0.6766611  0.4073065
##   0.050       5                100                 700      0.6754907  0.4093836
##   0.050       5                100                 750      0.6745935  0.4109534
##   0.050       5                100                 800      0.6737978  0.4123566
##   0.050       5                100                 850      0.6729002  0.4138626
##   0.050       5                100                 900      0.6724639  0.4147275
##   0.050       5                100                 950      0.6718121  0.4157630
##   0.050       5                100                1000      0.6712200  0.4168549
##   0.050       5                200                 100      0.7041734  0.3587231
##   0.050       5                200                 150      0.6979078  0.3694969
##   0.050       5                200                 200      0.6932760  0.3777859
##   0.050       5                200                 250      0.6908678  0.3821928
##   0.050       5                200                 300      0.6888490  0.3858865
##   0.050       5                200                 350      0.6870920  0.3889154
##   0.050       5                200                 400      0.6860994  0.3906776
##   0.050       5                200                 450      0.6843188  0.3939007
##   0.050       5                200                 500      0.6830105  0.3961346
##   0.050       5                200                 550      0.6827148  0.3966903
##   0.050       5                200                 600      0.6818150  0.3981789
##   0.050       5                200                 650      0.6814667  0.3988585
##   0.050       5                200                 700      0.6807369  0.4001349
##   0.050       5                200                 750      0.6801798  0.4010136
##   0.050       5                200                 800      0.6794995  0.4021241
##   0.050       5                200                 850      0.6793208  0.4027386
##   0.050       5                200                 900      0.6785405  0.4040740
##   0.050       5                200                 950      0.6781668  0.4047304
##   0.050       5                200                1000      0.6781442  0.4049031
##   0.100       3                100                 100      0.6991857  0.3668671
##   0.100       3                100                 150      0.6937574  0.3767540
##   0.100       3                100                 200      0.6904313  0.3830016
##   0.100       3                100                 250      0.6887795  0.3857501
##   0.100       3                100                 300      0.6868803  0.3892108
##   0.100       3                100                 350      0.6851442  0.3924489
```

```
##   0.100      3                 100                 400     0.6854112  0.3919858
##   0.100      3                 100                 450     0.6842333  0.3939950
##   0.100      3                 100                 500     0.6836957  0.3951078
##   0.100      3                 100                 550     0.6827659  0.3967498
##   0.100      3                 100                 600     0.6822958  0.3976817
##   0.100      3                 100                 650     0.6821849  0.3979155
##   0.100      3                 100                 700     0.6812463  0.3997226
##   0.100      3                 100                 750     0.6809889  0.3999880
##   0.100      3                 100                 800     0.6800760  0.4018395
##   0.100      3                 100                 850     0.6790940  0.4035487
##   0.100      3                 100                 900     0.6792632  0.4035284
##   0.100      3                 100                 950     0.6785498  0.4045025
##   0.100      3                 100                1000     0.6787313  0.4045442
##   0.100      3                 200                 100     0.7033274  0.3595979
##   0.100      3                 200                 150     0.6982301  0.3687036
##   0.100      3                 200                 200     0.6950898  0.3744513
##   0.100      3                 200                 250     0.6934043  0.3773381
##   0.100      3                 200                 300     0.6917798  0.3802137
##   0.100      3                 200                 350     0.6905947  0.3822889
##   0.100      3                 200                 400     0.6891199  0.3849831
##   0.100      3                 200                 450     0.6885200  0.3861144
##   0.100      3                 200                 500     0.6884279  0.3863255
##   0.100      3                 200                 550     0.6873095  0.3884965
##   0.100      3                 200                 600     0.6867638  0.3896127
##   0.100      3                 200                 650     0.6868129  0.3894328
##   0.100      3                 200                 700     0.6860070  0.3909083
##   0.100      3                 200                 750     0.6850943  0.3924902
##   0.100      3                 200                 800     0.6852033  0.3925008
##   0.100      3                 200                 850     0.6853986  0.3921536
##   0.100      3                 200                 900     0.6845150  0.3937575
##   0.100      3                 200                 950     0.6841177  0.3944737
##   0.100      3                 200                1000     0.6832034  0.3959351
##   0.100      5                 100                 100     0.6904254  0.3829091
##   0.100      5                 100                 150     0.6855922  0.3915604
##   0.100      5                 100                 200     0.6826523  0.3969902
##   0.100      5                 100                 250     0.6809679  0.3999928
##   0.100      5                 100                 300     0.6798449  0.4018751
##   0.100      5                 100                 350     0.6784306  0.4046584
##   0.100      5                 100                 400     0.6761041  0.4085639
##   0.100      5                 100                 450     0.6754817  0.4098385
##   0.100      5                 100                 500     0.6744848  0.4116528
##   0.100      5                 100                 550     0.6735755  0.4133974
##   0.100      5                 100                 600     0.6728631  0.4147672
##   0.100      5                 100                 650     0.6720839  0.4162547
##   0.100      5                 100                 700     0.6715979  0.4172170
##   0.100      5                 100                 750     0.6717513  0.4170798
##   0.100      5                 100                 800     0.6715324  0.4174592
##   0.100      5                 100                 850     0.6710122  0.4184630
##   0.100      5                 100                 900     0.6701016  0.4200126
##   0.100      5                 100                 950     0.6698501  0.4207806
##   0.100      5                 100                1000     0.6693939  0.4216555
##   0.100      5                 200                 100     0.6950160  0.3745018
##   0.100      5                 200                 150     0.6901132  0.3832688
##   0.100      5                 200                 200     0.6883418  0.3864236
##   0.100      5                 200                 250     0.6864304  0.3898568
```

```
##   0.100      5                  200               300     0.6853924  0.3919402
##   0.100      5                  200               350     0.6841042  0.3940763
##   0.100      5                  200               400     0.6827993  0.3963075
##   0.100      5                  200               450     0.6826574  0.3968774
##   0.100      5                  200               500     0.6821246  0.3979067
##   0.100      5                  200               550     0.6811965  0.3995464
##   0.100      5                  200               600     0.6804466  0.4010221
##   0.100      5                  200               650     0.6803318  0.4013843
##   0.100      5                  200               700     0.6802246  0.4015018
##   0.100      5                  200               750     0.6789962  0.4037769
##   0.100      5                  200               800     0.6785844  0.4046930
##   0.100      5                  200               850     0.6782839  0.4053268
##   0.100      5                  200               900     0.6773275  0.4071341
##   0.100      5                  200               950     0.6764926  0.4085214
##   0.100      5                  200              1000     0.6754954  0.4104012

## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were n.trees = 1000, interactio
n.depth =
##  5, shrinkage = 0.1 and n.minobsinnode = 100.

> summary(gbm.caret)

##                                       var    rel.inf
## alcohol                           alcohol 21.7660390
## volatile_acidity         volatile_acidity 11.7731859
## free_sulfur_dioxide   free_sulfur_dioxide  8.3500203
## total_sulfur_dioxide total_sulfur_dioxide  8.3149068
## density                           density  8.2468528
## residual_sugar             residual_sugar  7.8071170
## sulphates                       sulphates  7.6158444
## pH                                     pH  7.0877860
## citric_acid                   citric_acid  6.7395801
## chlorides                       chlorides  6.5561664
## fixed_acidity               fixed_acidity  5.5230847
## stylewhite                     stylewhite  0.2194164
```




```
plot(gbm.caret)
```



Explanation:

For Boosting, caret package was used to tune the parameters. Number of variables tried at each node was varied from 3 and 5, number of trees grown was changed from 100 to 1000 in increments of 50, shrinkage/learning rate was tried from 0.001, 0.01, 0.05 and 0.1. Number of minimum observations at node was set at 100 and 200, which is high enough to prevent overfitting of model to training data. 10-fold cross-validation was used with RMSE as the metric. From the plot, it can be observed that for higher shrinkage rates (0.05 and 0.1), there is no significant decrease in RMSE after around 600 trees, whereas the RMSE keeps decreasing for lower shrinkage (0.001 and 0.01), indicating that more number of trees would be required for lower shrinkage rates.

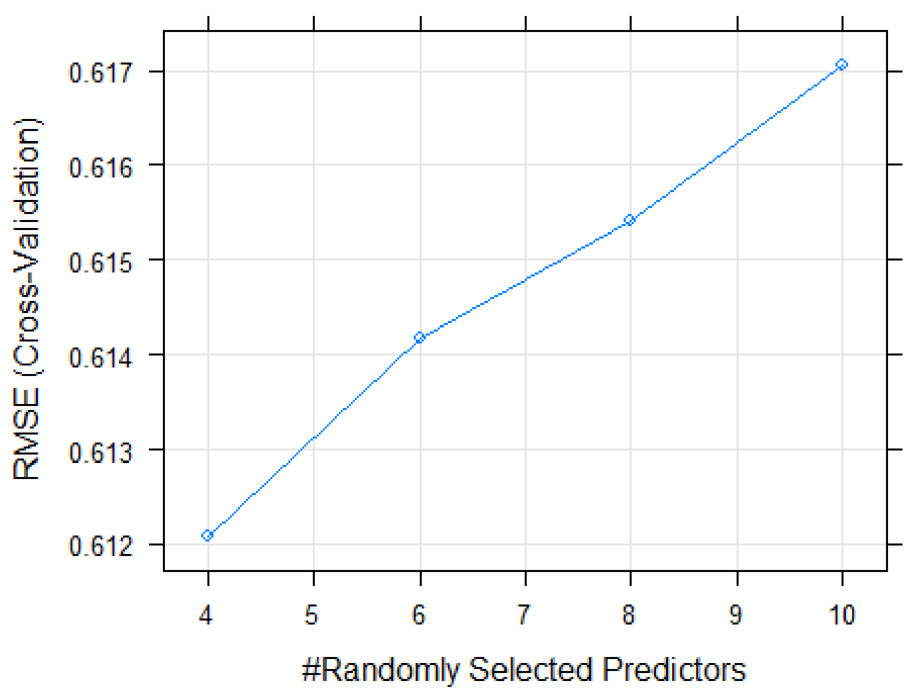
Random Forest code:

```
#Random Forest-----------------------------------------------

rf.caretGrid=expand.grid(mtry=c(4,6,8,10))
metric="RMSE"
trainControl=trainControl(method="cv",number=10)
set.seed(1)
rf.caret=train(quality~., data=Train.data,method="rf",
                trControl=trainControl, verbose=FALSE,
                tuneGrid=rf.caretGrid, metric=metric, ntree=500)
```

```r
print(rf.caret)
```

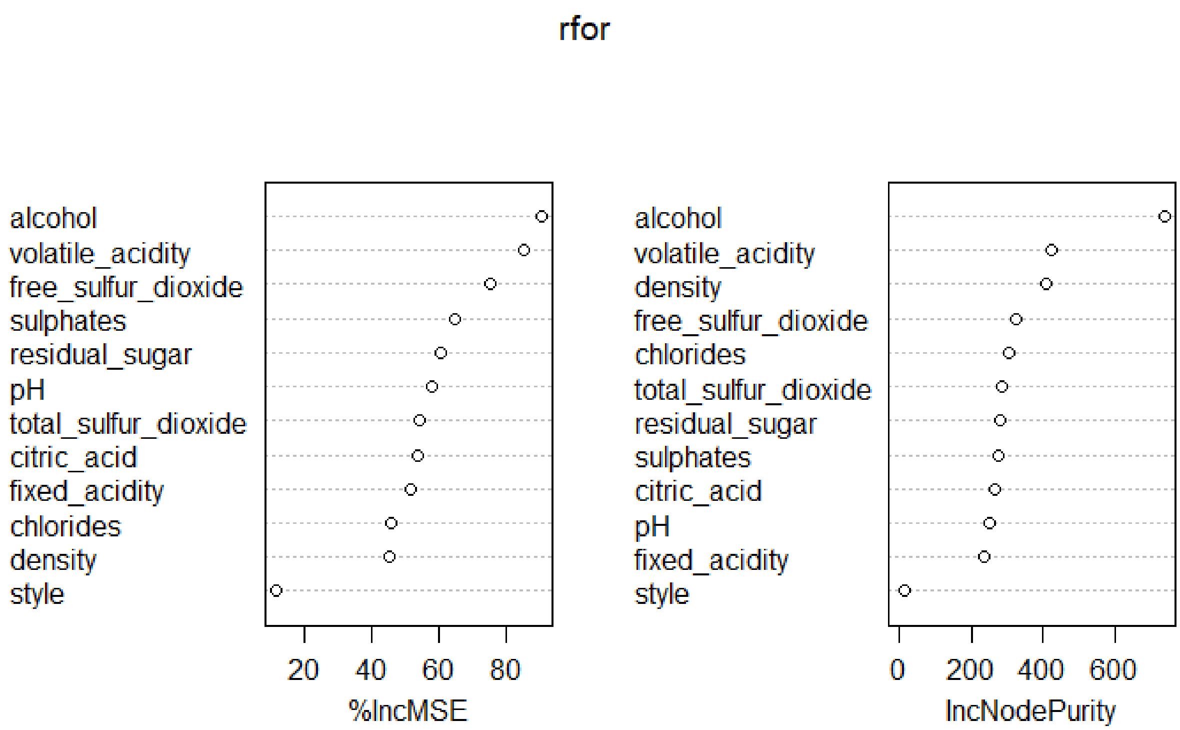
```
## Random Forest
##
## 5198 samples
##   12 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 4678, 4679, 4678, 4679, 4678, 4678, ...
## Resampling results across tuning parameters:
##
##   mtry  RMSE       Rsquared   MAE
##    4    0.6120737  0.5212333  0.4459086
##    6    0.6141681  0.5159709  0.4462314
##    8    0.6154171  0.5133363  0.4468721
##   10    0.6170645  0.5098836  0.4468749
##
## RMSE was used to select the optimal model using the smallest value.
## The final value used for the model was mtry = 4.
```

```r
plot(rf.caret)
```




```r
library(randomForest)
```

```
set.seed(1)
rfor=randomForest(quality~.,data=Train.data,mtry=4,ntree=500,importanc
e=TRUE)
varImp(rfor)

##                        Overall
## fixed_acidity         52.83715
## volatile_acidity      79.70257
## citric_acid           51.67916
## residual_sugar        53.15982
## chlorides             47.43820
## free_sulfur_dioxide   68.80161
## total_sulfur_dioxide  51.33943
## density               41.47568
## pH                    58.65402
## sulphates             62.89265
## alcohol               89.12409
## style                 13.39972

varImpPlot(rfor)
```



Explanation:

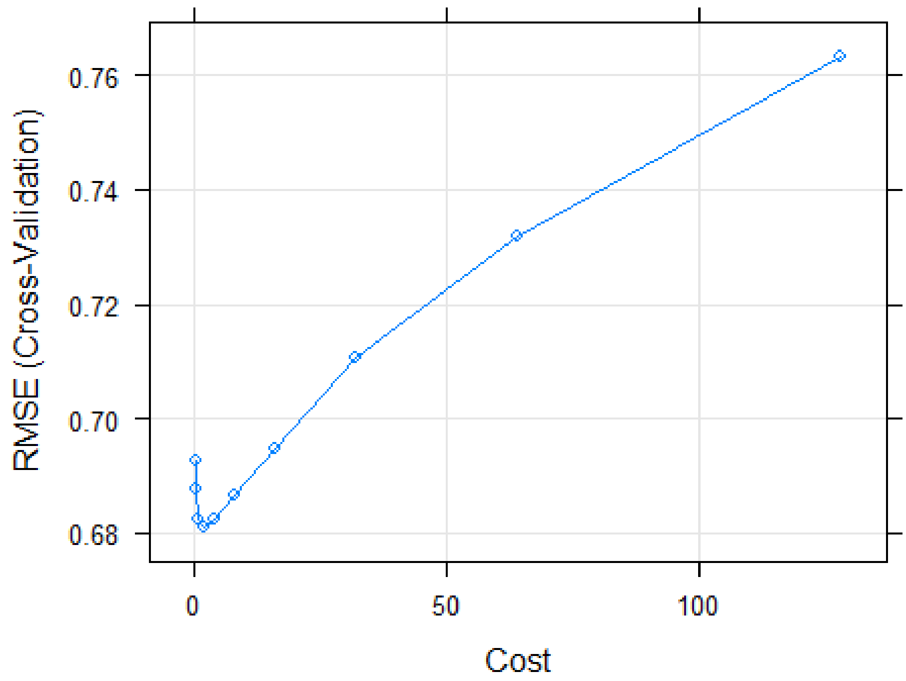
For Random Forest, number of predictors tried at each node was varied between 4, 6, 8 and 10. Since there are a total of 12 predictors, the best value of mtry should be 4 as p/3, where p is the number of predictors, is the thumb rule. This is the case, as seen from the graph and the RMSE values as well. As mtry increases, the RMSE increases, for 500 trees. Variable importance was also derived from the random forest model. It is seen that for predicting the wine quality, %alcohol is the most influential predictor, closely followed by volatile_acidity and free_sulfur_dioxide. Wine style (red or white) does not have any effect on its quality.

SVM code:

```
#SVM---------------------------------------------------

library(e1071)
trainControl=trainControl(method="cv",number=10)

set.seed(1)
svm.caret=train(quality~.,data=Train.data,method="svmRadial",
                preProcess=c("center","scale"),
                tuneLength=10,trControl=trainControl)
svm.caret
```

```
## Support Vector Machines with Radial Basis Function Kernel
##
## 5198 samples
##   12 predictor
##
## Pre-processing: centered (12), scaled (12)
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 4678, 4679, 4678, 4679, 4678, 4678, ...
## Resampling results across tuning parameters:
##
##   C       RMSE       Rsquared   MAE
##     0.25  0.6926980  0.3792132  0.5251839
##     0.50  0.6877696  0.3881147  0.5198233
##     1.00  0.6824698  0.3977056  0.5132648
##     2.00  0.6807917  0.4015503  0.5105128
##     4.00  0.6823434  0.4007518  0.5097913
##     8.00  0.6864692  0.3970218  0.5109669
##    16.00  0.6947185  0.3895968  0.5156426
##    32.00  0.7106061  0.3751361  0.5240890
##    64.00  0.7319058  0.3594585  0.5324724
##   128.00  0.7634800  0.3387071  0.5454074
##
## Tuning parameter 'sigma' was held constant at a value of 0.09179926
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were sigma = 0.09179926 and C =
2.
```

```
plot(svm.caret)
```



Explanation:

For Support Vector Machines, radial kernel was used in the caret package, instead of tune command of e1071 library. This is because using caret, the best model is selected based on RMSE values, whereas for e1071, the best model selection is based on MSE values. Since it is essential to compare the three algorithms, a similar metric to select the best model for each algorithm should be used. In caret, tuneLength argument is used, which varies the cost associated with the model, at a constant sigma value. From the graph, it can be seen that RMSE decreases initially with increase in cost and then starts increasing.

The model RMSE values are compared for all three algorithms in the table below:

Algorithm	**Training RMSE**
Boosting	0.6693
Random Forest	0.6120
SVM (Radial)	0.6807

From the table above, it is seen that Random Forest has the least RMSE on training data

Code for performance on test data:

```
#Performance on holdout data------------------------------------

gbm.predict=predict(gbm.caret,Test.data)
mean((gbm.predict-Test.data$quality)^2)

## [1] 0.4161539
```

```r
rf.predict=predict(rf.caret,Test.data)
mean((rf.predict-Test.data$quality)^2)

## [1] 0.3368632

svmcaret.predict=predict(svm.caret,Test.data)
mean((svmcaret.predict-Test.data$quality)^2)

## [1] 0.4372144

#RMSE gbm:0.416 rf:0.33 svmrad:0.43
```

Explanation:

The performance of all three algorithms were compared for holdout data using RMSE value. The performance is summarized in the table below:

Algorithm	**Test RMSE**
Boosting	0.4161
Random Forest	**0.3368**
SVM (Radial)	0.4372

From the table, it can be seen that Random Forest has the best performance on holdout data.

2. Assumptions made about wine quality data when using a regression model

In a regression model the wine quality data is a numeric value. The underlying assumption is that the data on which wine quality is rated is based on a numeric model. That is, the quality value can be mathematically be derived from the predictor data. It implies that a wine with a score of 8 is twice better than a wine with a quality of 4.

Another assumption about the predicted data is that the wine quality can be a decimal number. This is because the quality data is treated as numeric by the model and the output of the model can be any numeric value, including decimals. This means the wine quality can be 5.5, which is not the case since all the datapoints in training data has an integer value.

This is generally not the case, as the quality cannot be arrived at mathematically. Most of the times, the quality is based on a wine taster’s opinion and experience. Therefore, treating the quality as a numerical value is not correct. It should be treated as a class data, and therefore classification should be performed instead of regression.