CHICAGO BOOTH

The University of Chicago Booth School of Business

Machine Learning

Winter 2020

HW #5

Jiyoon Chung

Arman Bhuiyan

Deepak Putchakayala

Hikaru Sugimori

```
> theme_set(theme_tufte(base_size = 14))
> set.seed(1)
>
> data(Auto)
> Auto <- as.tibble(Auto)
Warning message:
`as.tibble()` is deprecated, use `as_tibble()` (but mind the new semantics).
This warning is displayed once per session.
>
> Auto <- Auto %>%
+    mutate(highmpg = as.integer(mpg > median(mpg))) %>%
+    mutate(highmpg = factor(highmpg),
+            cylinders = factor(cylinders))
>
> Auto %>%
+    sample_n(5) %>%
+    select(mpg, highmpg)
# A tibble: 5 x 2
    mpg highmpg
  <dbl> <fct>
1  44.3 1
2  23   1
3  26   1
4  23.9 1
5  23.2 1
>
> Auto <- Auto %>%
+    select(-mpg, -name)
>
> dummy_trans <- dummyVars(highmpg ~ ., data = Auto)
> Auto_dummy <- predict(dummy_trans, Auto)
Warning message:
In model.frame.default(Terms, newdata, na.action = na.action, xlev =
object$lvls) :
   variable 'highmpg' is not a factor
>
> svm_linear <- train(x = Auto_dummy, y = Auto$highmpg,
+                     method = 'svmLinear2',
+                     trControl = trainControl(method = 'cv', number = 10,
allowParallel = TRUE),
+                     preProcess = c('center', 'scale'),
+                     tuneGrid = expand.grid(cost = seq(1, 20, by = 1)))
>
> svm_linear$finalModel

Call:
svm.default(x = as.matrix(x), y = y, kernel = "linear", cost = param$cost,
    probability = classProbs)


Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  linear
       cost:  2

Number of Support Vectors:  75


>
> svm_poly <- train(x = Auto_dummy, y = Auto$highmpg,
+                     method = 'svmPoly',
+                     trControl = trainControl(method = 'cv', number = 10,
allowParallel = TRUE),
+                     preProcess = c('center', 'scale'),
```

```
+                       tuneGrid = expand.grid(degree = seq(1, 8, by = 1),
+                                              C = seq(1, 5, by = 1),
+                                              scale = TRUE))
>
> svm_poly$finalModel
Support Vector Machine object of class "ksvm"

SV type: C-svc  (classification)
 parameter : cost C = 1

Polynomial kernel function.
 Hyperparameters : degree =  2  scale =  TRUE  offset =  1

Number of Support Vectors : 71

Objective Function Value : -45.587
Training error : 0.045918
>
> svm_radial <- train(x = Auto_dummy, y = Auto$highmpg,
+                     method = 'svmRadial',
+                     trControl = trainControl(method = 'cv', number = 10,
allowParallel = TRUE),
+                     preProcess = c('center', 'scale'),
+                     tuneGrid = expand.grid(C = seq(0.001, 3, length.out =
10),
+                                            sigma = seq(0.2, 2, length.out =
5)))
>
> svm_radial$finalModel
Support Vector Machine object of class "ksvm"

SV type: C-svc  (classification)
 parameter : cost C = 1.00066666666667

Gaussian Radial Basis kernel function.
 Hyperparameter : sigma =  1.55

Number of Support Vectors : 230

Objective Function Value : -73.7206
Training error : 0.02551
>
> plot(svm_linear)
>
> plot(svm_poly)
>
> plot(svm_radial)
>
> postResample(predict(svm_linear), Auto$highmpg)
 Accuracy     Kappa
0.9285714 0.8571429
>
> postResample(predict(svm_poly), Auto$highmpg)
 Accuracy     Kappa
0.9540816 0.9081633
>
> postResample(predict(svm_radial), Auto$highmpg)
 Accuracy     Kappa
0.9744898 0.9489796
```