Bfox.R

arpan

2023-07-26

```
library(car)
```

```
## Loading required package: carData
```

```
# Load the Bfox dataset
data("Bfox")


bfox <- Bfox
head(bfox)
```

```
##      partic  tfr menwage womwage  debt parttime
## 1946   25.3 3748   25.35   14.05 18.18    10.28
## 1947   24.4 3996   26.14   14.61 28.33     9.28
## 1948   24.2 3725   25.11   14.23 30.55     9.51
## 1949   24.2 3750   25.45   14.61 35.81     8.87
## 1950   23.7 3669   26.79   15.26 38.39     8.54
## 1951   24.2 3682   26.33   14.58 26.52     8.84
```

```
# Split the data into train and test datasets
set.seed(123)
train_index <- sample(1:nrow(bfox), 0.7 * nrow(bfox))
train <- bfox[train_index, ]
test <- bfox[-train_index, ]

# Fit a supervised linear regression model
lm_model <- lm(debt ~ ., data = train)
summary(lm_model)
```

```
##
## Call:
## lm(formula = debt ~ ., data = train)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -7.747 -3.073 -0.066  3.312  6.516
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.005e+02  2.159e+01  -9.283 1.32e-07 ***
```

```
## partic       9.073e+00  1.347e+00   6.737 6.67e-06 ***
## tfr         -2.199e-03  4.101e-03  -0.536  0.59974
## menwage      2.018e+00  1.008e+00   2.001  0.06378 .
## womwage      8.815e-01  1.883e+00   0.468  0.64635
## parttime    -4.357e+00  1.346e+00  -3.236  0.00554 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.495 on 15 degrees of freedom
## Multiple R-squared:  0.9959, Adjusted R-squared:  0.9945
## F-statistic: 721.1 on 5 and 15 DF,  p-value: < 2.2e-16
```

```
# Fit a KNN regression model
library(caret)
```

```
## Loading required package: ggplot2
```

```
## Loading required package: lattice
```

```
knn_model <- knnreg(debt ~ ., data = train, k = 5)
summary(knn_model)
```

```
##         Length Class  Mode
## learn   2      -none- list
## k       1      -none- numeric
## terms   3      terms  call
## xlevels 0      -none- list
## theDots 0      -none- list
```

```
# Predict the debt variable in the test data
lm_pred <- predict(lm_model, test)
knn_pred <- predict(knn_model, test)

# Get the fit indices of the predicted models
lm_r2 <- 1 - sum((test$debt - lm_pred)^2) / sum((test$debt - mean(test$debt))^2)
lm_mse <- mean((test$debt - lm_pred)^2)
lm_rmse <- sqrt(lm_mse)

knn_r2 <- 1 - sum((test$debt - knn_pred)^2) / sum((test$debt - mean(test$debt))^2)
knn_mse <- mean((test$debt - knn_pred)^2)
knn_rmse <- sqrt(knn_mse)

# Print the fit indices
print(paste("Linear regression R^2:", lm_r2))
```

```
## [1] "Linear regression R^2: 0.976473840117479"
```

```
print(paste("Linear regression MSE:", lm_mse))
```

```
## [1] "Linear regression MSE: 94.4393060089475"
```

```r
print(paste("Linear regression RMSE:", lm_rmse))
```

```
## [1] "Linear regression RMSE: 9.71798878415423"
```

```r
print(paste("KNN regression R^2:", knn_r2))
```

```
## [1] "KNN regression R^2: 0.837863813290572"
```

```r
print(paste("KNN regression MSE:", knn_mse))
```

```
## [1] "KNN regression MSE: 650.851181333333"
```

```r
print(paste("KNN regression RMSE:", knn_rmse))
```

```
## [1] "KNN regression RMSE: 25.5117851459543"
```