HW2 KNN Regression

2025-02-13

```r
library(MASS)
set.seed(123)

# Subset the data to only the variables of interest: lstat and medv
data(Boston)
boston_sub <- Boston[, c("lstat", "medv")]

# Create a train/test split (70% training, 30% testing)
n <- nrow(boston_sub)
train_index <- sample(1:n, size = round(0.7 * n))
train_data <- boston_sub[train_index, ]
test_data  <- boston_sub[-train_index, ]

# --- Least Squares Fit ---
lm_fit <- lm(medv ~ lstat, data = train_data)
lm_pred <- predict(lm_fit, newdata = test_data)
lm_mse <- mean((test_data$medv - lm_pred)^2)
cat("Least Squares MSE:", lm_mse, "\n")
```

```
## Least Squares MSE: 41.08968
```

```r
# --- KNN Regression ---

knn_reg <- function(x_train, y_train, x_test, k) {
  predictions <- sapply(x_test, function(x) {
    # Compute Euclidean distance
    distances <- abs(x_train - x)
    # Find indices of the k nearest neighbors
    neighbor_indices <- order(distances)[1:k]
    # Return the average response of the neighbors
    mean(y_train[neighbor_indices])
  })
  return(predictions)
}

k <- 5
knn_pred <- knn_reg(train_data$lstat, train_data$medv, test_data$lstat, k)
knn_mse <- mean((test_data$medv - knn_pred)^2)
cat("KNN (k =", k, ") MSE:", knn_mse, "\n")
```

```
## KNN (k = 5 ) MSE: 36.91268
```

```r
# --- Compare the Two Methods ---
if (knn_mse < lm_mse) {
  cat("KNN regression performs better on the test set (lower MSE).\n")
} else if (lm_mse < knn_mse) {
```

```
  cat("Least Squares regression performs better on the test set (lower MSE).\n")
} else {
  cat("Both methods perform equally well on the test set.\n")
}
```

```
## KNN regression performs better on the test set (lower MSE).
```

```
plot(test_data$lstat, test_data$medv, pch = 16, col = "gray",
     xlab = "lstat", ylab = "medv", main = "Model Predictions vs Actual")
points(test_data$lstat, lm_pred, col = "blue", pch = 17)
points(test_data$lstat, knn_pred, col = "red", pch = 18)
legend("topright", legend = c("Actual", "Least Squares", "KNN"),
       col = c("gray", "blue", "red"), pch = c(16, 17, 18))
```