**HW2 (\*이 워드 파일에 R code와 output을 기입하여 제출해주시기 바랍니다.)**

1. iris 데이터를 70 대 30으로 random split 하고 Linear Regression model을 사용해서 training MSE와 test MSE를 구하시오 **(\*데이터 분리할 때 set.seed(1) 사용)**

|  |  |
| --- | --- |
|  | Sepal.Length를 y로 하고, 나머지 변수를 x1, x2, ..로 하는 선형 회귀 분석을 하려 한다. 여기서 x에 해당하는 4번째 변수인 Species가 categorical variable이기 때문에 이 변수를 dummy variable로 변환하는 작업을 먼저 진행하였다.  Species는 총 3가지 종류가 있으며 1~50행=setosa, 51~100행=versicolor, 101~150행=virginica임을 참고하여 dummy variable d1, d2를 만들어 새로운 data frame iris\_aug를 만들었다.  Setosa의 경우 (d1,d2)=(1,0)  Versicolor의 경우 (d1,d2)=(0,1)  Virginica의 경우 (d1, d2)=(0,0)  이다. |
|  | 또한 Species를 dummy variable이 아닌 allocated code를 이용하여 setosa=1, versicolor=2, virginica=3으로 대응시켜 새로운 dataframe iris aug\_2를 만들었다.  한편, 선형 회귀분석을 하기 전 Sepal.Length와 qualitative variable들 사이의 관계를 ggplot을 통해 확인해보았다. |
|  | |
| Training data를 이용해 Sepal.Length를 Sepal.Width, Petal.Length, Petal.Width,Species(using d1,d2 dummy variable)로 선형회귀분석한 뒤 잔차를 제곱하여 training mse를 얻었고, test data에서 Sepal.Length(true)와 Trained Model로 도출된 Sepal.Length(Fitted value)를 이용해 test mse를 얻었다.    마찬가지로, allocated code를 이용해 표현된 iris dataframe에도 동일한 과정을 거쳐 training mse와 test mse를 구하였다. | |

1. iris 데이터를 60 대 40으로 random split 하고 Linear Regression model을 사용해서 training MSE와 test MSE를 구하시오 **(\*데이터 분리할 때 set.seed(1) 사용)**

|  |
| --- |
|  |
| 위는 Dummy variable을 이용해 만든 dataframe에서의 training mse, test mse를 구하는 과정과 결과이다.    위는 allocated code를 이용해 만든 dataframe에서의 training mse, test mse를 구하는 과정과 결과이다. |