

문제 2 mpg 자료를 이용하여 다음을 코딩하라.(결측값은 제거할 것). (조건: 반응 변수: mpg01, 예측 변수: cylinders, weight, displacement, horsepower로 지정)

1. mpg의 값이 중앙값보다 크면 1로, 아니면 0으로 할당하는 이항 변수 자료를 만들어라. 그리고 Auto 자료에 mpg01 이라는 변수명으로 추가하라.

	mpg	cylinders	displacement	horsepower	weight	acceleration	year	origin	name	mpg01
1	18	8	307.0	130	3504	12.0	70	1	chevrolet chevelle malibu	0
2	15	8	350.0	165	3693	11.5	70	1	buick skylark 320	0
3	18	8	318.0	150	3436	11.0	70	1	plymouth satellite	0
4	16	8	304.0	150	3433	12.0	70	1	amc rebel sst	0
5	17	8	302.0	140	3449	10.5	70	1	ford torino	0
6	15	8	429.0	198	4341	10.0	70	1	ford galaxie 500	0
7	14	8	454.0	220	4354	9.0	70	1	chevrolet impala	0
8	14	8	440.0	215	4312	8.5	70	1	plymouth fury iii	0

2. 선형 판별 분석(LDA)을 수행하라. 이때, 자료를 짝수년을 기준으로 훈련 자료와 시험 자료로 분리하고 전체 오차율을 구하여라.

```
> table(lda.class,mpg01.test) # 혼동 행렬을 생성하여 예측값과 test set 비교
      mpg01.test
lda.class 0  1
      0 86  9
      1 14 73
> mean(lda.class==mpg01.test)
[1] 0.8736264
> mean(lda.class != mpg01.test)
[1] 0.1263736
> (14+9)/(182) # 전체 오차율은 12.6%
[1] 0.1263736
```

3. 이차 판별 분석(QDA)을 수행하라. 그리고 혼동 행렬 (confusion matrix)을 만들고 전체 오차율을 구하여라.

```
> mean(qda.pred$class != mpg01.test)
[1] 0.1318681
> mean(qda.class==mpg01.test)
[1] 0.8681319
> table(qda.class,mpg01.test)
      mpg01.test
qda.class 0  1
      0 89 13
      1 11 69
> (13+11)/(182) # 전체 오차율은 13.2%
[1] 0.1318681
```

4. 다중 로지스틱 회귀 분석을 수행하라. 그리고 시험자료에서 예측한 값과 실제 값을 비교하고 전체 오차율을 구하여라.

```
> table(glm.pred, mpg01.test)
      mpg01.test
glm.pred  0  1
      0 89 11
      1 11 71
> mean(glm.pred == mpg01.test)
[1] 0.8791209
> mean(glm.pred != mpg01.test)
[1] 0.1208791
> (11+11)/(182) # 전체 오차율은 12.1%
[1] 0.1208791
```

5. 최근접이웃방법(KNN)을 수행하라. 그리고 $K = 1$, $K = 10$, $K = 100$ 일 때, 각 전체 오차율을 구하여라. .

```
-----\
> knn.pred = knn(train.x, test.x, train.mpg01, k = 1) #k가 1인 KNN 수행
> mean(knn.pred != mpg01.test) #k가 1일때의 전체오차율은 15.4%
[1] 0.1538462
> knn.pred = knn(train.x, test.x, train.mpg01, k = 10) #k가 10인 KNN 수행
> mean(knn.pred != mpg01.test) #k가 10일때의 전체오차율은 16.5%
[1] 0.1648352
> knn.pred = knn(train.x, test.x, train.mpg01, k = 100) #k가 100인 KNN 수행
> mean(knn.pred != mpg01.test) #k가 100일때의 전체오차율은 14.3%
[1] 0.1428571
```