

XGBoost 복습문제

#1-1 cance데이터를 createDataPartition을 이용해
train:test = 7:3으로 분할하세요 (set.seed=113)

```
> head(train)
```

	y	x1	x2	x3	x4	x5	x6	x7	x8	x9
2	0	5	4	4	5	7	10	3	2	1
5	0	4	1	1	3	2	1	3	1	1
6	1	8	10	10	8	7	10	9	7	1
7	0	1	1	1	1	2	10	3	1	1
10	0	4	2	1	1	2	1	2	1	1
11	0	1	1	1	1	1	1	3	1	1

```
> head(test)
```

	y	x1	x2	x3	x4	x5	x6	x7	x8	x9
1	0	5	1	1	1	2	1	3	1	1
3	0	3	1	1	1	2	2	3	1	1
4	0	6	8	8	1	3	4	3	7	1
8	0	2	1	2	1	2	1	3	1	1
9	0	2	1	1	1	2	1	1	1	5
12	0	2	1	1	1	2	1	2	1	1

#1-2 RandomForest 모델을 사용하여 train으로 학습한 모델을
test셋에 대한 Accuracy를 구하세요.
(이 때, 옵션으로는 mtry=3, ntrees = 113으로 해주세요)

#1-3 AdaBoost 모델을 사용하여 train으로 학습한 모델을 test
셋에 대한 Accuracy를 구하세요.
(이 때, 옵션은 따로 지정하지 않습니다.)

2-1 수업시간에 배운 내용을 바탕으로 XGBoost를 사용하기 전
데이터를 적절한 매트릭스 형태로 바꿔주세요
(train / test 같은 방식으로 바꿔주셔야 됩니다)

```
> xg.train  
xgb.DMatrix  dim: 479 x 10  info: label  colnames: yes  
> xg.test  
xgb.DMatrix  dim: 204 x 10  info: label  colnames: yes
```

#2-2 XGBoost 모델로 Accuracy를 구하는 문제입니다.

params는 주어진 조건대로 쓰시면 됩니다.

max_depth와 nrounds를 grid search를 통해 가장 큰 Accuracy를 갖는 조합을 찾으세요.
(max_depth의 범위는 1부터 6까지, nrounds의 조합은 seq(100,310,30)으로 설정해주세요!)

```
params = list(booster = "gbtree",  
              eta = 0.01,  
              max_depth = ,  
              gamma = 3 ,  
              subsample = 0.8,  
              objective = "binary:logistic",  
              eval_metric = "error")
```

```
xgb.fit <- xgb.train(  
  params = params,  
  data = xg.train,  
  nrounds = ,  
  watchlist = list(val1 = xg.train, val2 = xg.test),  
  verbose = 0  
)
```