



15주차 과제

과 목	데이터사이언스프로그래밍
담당 교수	이 두 호
학 번	201720970
학 과	소프트웨어·미디어·산업공학부
이 름	권 대 한

1. rpf.csv를 로딩하시오. 일주일 동안 평균 4.9시간 공부하고, 출석점수가 82점일 때 pass, fail을 예측하라.

```
가) rpf <- read.csv("rpf.csv", header = T)[-1, ]
나) rpf$outcomes %<>% as.factor()
다) sample_num <- sample(1:nrow(rpf), nrow(rpf)*0.7)
라) rpf.train <- rpf[sample_num, ]
마) rpf.test <- rpf[-sample_num, ]
바) nnet_object <- nnet(outcomes ~., data=rpf.train, size=20, maxit=10000)
사) score.prediction <- predict(nnet_object, rpf.test[, -3], type = "class")
아) print(score.prediction)
자) submission <- data.frame(4.9, 82)
차) colnames(submission)<- c("hours","absence")
카) submission.prediction <- predict(nnet_object, submission, type = "class")
타) print(submission.prediction)
```

코드 가)에서 1열의 행 번호 데이터를 제외하고 로드한다.

코드 나)에서 Fail, Pass의 데이터를 레벨로 정의해 학습하기 위해서 rpf 객체의 outcomes Field 데이터를 Factor로 변환해주었다.

코드 다)에서는 임의의 행 번호를 추출하였으며, 제대로 학습이 되었는지 검증하기 위해 코드 라), 마)에서는 추출한 행 번호에 대한 데이터를 train, test 객체에 복사하였다.

코드 바)에서는 nnet 함수를 사용해서 outcomes Field를 제외한 데이터를 input에 대한 학습 데이터(독립변수)로 입력하였으며, outcomes Field 데이터를 output(종속변수)에 대한 학습 데이터로 정의하며, 동시에 학습 Layer의 개수는 20개, 최대 학습 횟수는 10000번으로 정의하였다.

코드 사)에서 test 데이터를 통해 모델의 학습 검증을 하였으며, 코드 아)에서 결과를 확인하였다.

이후 코드 자), 차)에서 4.9시간 학습, 출석시간 82점에 대한 데이터 프레임을 정의하였으며, 코드 카)에서 Output 데이터를 예측하며 객체에 저장하였다. 이후 코드 타)에서 예측 결과를 출력한다.

```
> print(score.prediction)
[1] "fail" "fail" "pass" "pass" "pass" "pass" "pass" "pass" "fail" "pass" "fail" "pass" "fail"
> submission <- data.frame(4.9, 82)
> colnames(submission)<- c("hours","absence")
> submission.prediction <- predict(nnet_object, submission, type = "class")
> print(submission.prediction)
[1] "pass"
```