



		감독교수확인	이 두 호	
2025 학년도 1학기 기말고사	시험과목 : 머신러닝1 (2반)	전공 : AI소프트웨어학과 학번 : 20212285	출 제 자	이두호
		성명 : 이해준	채점성적	

```
#2번 (x1, x2) = (3, 62.5)
w0_hat*(3^w1_hat)*(62.5^w2_hat)
> #2번 (x1, x2) = (3, 62.5)
> w0_hat*(3^w1_hat)*(62.5^w2_hat)
(Intercept)
5.674608
```

이후 입력값으로 X1,X2 = 3, 62.5 가 들어갔을 시의 결과 값입니다.

[3]

```
load("C:/Users/user/Desktop/p3.RData")

#1번 능형회귀모델, Adam 알고리즘, 학습모델 가중치 파라미터를 구하시오
X <- p3[, -5]
X <- as.matrix(cbind(1, p3[, -5]))
y <- p3[, 5]
y <- as.matrix(y)

alpha <- 0.5
beta1 <- 0.9
beta2 <- 0.999
omega <- 10^-8
lamda <- 0.004
epoch <- 19000

w <- matrix(0, nrow = ncol(X), ncol = 1) |
m <- matrix(0, nrow = ncol(X), ncol = 1) # 1차
v <- matrix(0, nrow = ncol(X), ncol = 1) # 2차

for (t in 1:epoch) {
  g <- -2 * t(X) %>% (y - X %>% w) + 2 * lamda * w
  m <- beta1*m + (1-beta1)*g
  v <- beta2*v + (1-beta2)*(g*g)
  m_hat <- m/(1 - beta1^t)
  v_hat <- v/(1 - beta2^t)
  w <- w - alpha * m_hat/(sqrt(v_hat)+omega)
}

w_hat <- w
w_hat
```

```
> w_hat
      [,1]
1  30.37526239
x1 -6.64963020
x2  0.55719493
x3  0.03167138
x4 -0.14527607
```

(1)문제: 위 결과 값들이 각 변수의 회귀계수(파라미터 추정치)입니다.



```
#2번 Adam 알고리즘을 통해 추정된 능형회귀 학습모델의 결정계수를 구하시오
y_hat <- X %>% w_hat

SST <- sum((y - mean(y))^2)
SSE <- sum((y - y_hat)^2)

R2 <- 1 - SSE / SST
R2
```

```
> R2
[1] 0.9116833
```

(2)문제: 능형회귀 학습모델의 결정계수입니다.

		감독교수확인	이 두 호 	
2025 학년도 1학기 기말고사	시험과목 : 머신러닝1 (2반)	전공 : AI소프트웨어학과 학번 : 202122852 성명 : 이해준	출 제 자	이두호 
			채점성적	y_hat_lassoreg <- predict(lassoreg, x_test) # 검증데이터 예측오차제곱합의 평균 mean((y_test-y_hat_lassoreg)^2)

[4]

```
#1 데이터 불러오기
load("C:/Users/user/Desktop/p4.RData")

# 열 이름을 변수로 저장
col_name <- names(p4_train)[1] # 첫 번째(유일한) 열 이름 추출

# 데이터 분할 함수 정의
split_data <- function(data) {
  return(do.call(rbind, strsplit(data[[col_name]], ",")))
}

# 함수 적용
train_split_matrix <- split_data(p4_train)
test_split_matrix <- split_data(p4_test)

# 3. 데이터프레임으로 변환
train_split <- as.data.frame(train_split_matrix)
test_split <- as.data.frame(test_split_matrix)

# 4. 열 이름 지정
colnames(train_split) <- c(paste0("x", 1:100), "y")
colnames(test_split) <- c(paste0("x", 1:100), "y")

# 5. 숫자로 변환 (필요 시)
train_split <- as.data.frame(lapply(train_split, as.numeric))
test_split <- as.data.frame(lapply(test_split, as.numeric))
```

위와 같이 데이터를 불러온 뒤에 전처리 과정을 거칩니다 이후

```
#install.packages("glmnet")
library(glmnet)

x_train <- data.matrix(train_split[, -101]) #train 입력변수
y_train <- data.matrix(train_split[, 101]) #train 출력변수
x_test <- data.matrix(test_split[, -101]) #valid 입력변수
y_test <- data.matrix(test_split[, 101]) #valid 출력변수

#능형화귀모델 생성
lassoreg <- glmnet(x_train, y_train, alpha = 1, lambda = 0.04467)
round(coef(lassoreg))

coef_vec <- as.vector(coef(lassoreg))

count <- sum(coef_vec[-1] != 0)
count
```

```
> round(coef(lassoreg))
101 x 1 sparse Matrix of class "dgCMatrix"

      s0
(Intercept) 0
x1          1
x2          1
x3          1
x4          1
x5          1
x6          1
x7          1
x8          1
x9          1
x10         1
x11         0
x12         .
x13         .
x14         .
x15         .
```

```
> count <- sum(coef_vec[-1] != 0)
> count
[1] 25
```

(1)문제: matrix를 확인하고 0이 아닌 값들을 확인해보면 25라는 결과가 나오게 됩니다.

```
y_hat_lassoreg <- predict(lassoreg, x_test)

# 검증데이터 예측오차제곱합의 평균
mean((y_test-y_hat_lassoreg)^2)
```

```
> y_hat_lassoreg <- predict(lassoreg, x_test)
> # 검증데이터 예측오차제곱합의 평균
> mean((y_test-y_hat_lassoreg)^2)
[1] 1.058185
```

(2)문제: 검증데이터에 대해 예측오차제곱합의 평균을 구한 결과 값으로 1.058185가 나왔습니다.

		감독교수확인	이 두 호	
2025 학년도 1학기 기말고사	시험과목 : 머신러닝1 (2반)	전공 : AI소프트웨어학과 학번 : 202122852	출 제 자	이두호
		이혜준	채점성적	

[5]

```

load("C:/Users/user/Desktop/p5.RData")

library(caret)
#install.packages("ROCR")
library(ROCR)

x2_spiral_train <- data.frame(spiral_train$x2, spiral_train$y)
colnames(x2_spiral_train) <- c("x2", "y")

x2_logreg <- glm(y ~ x2, data=x2_spiral_train, family = "binomial")
x2_w0_hat <- coef(x2_logreg)[1]
x2_w1_hat <- coef(x2_logreg)[2]
x2_prob <- predict(x2_logreg, x2_spiral_train, type="response")
x2_y_hat <- ifelse(x2_prob > 0.5, 1, 0)

# 정오표
x2_cfm <- confusionMatrix(as.factor(x2_y_hat), as.factor(spiral_train$y))
x2_cfm

# roc 커브
x2_pr <- prediction(x2_prob, spiral_train$y)
x2_prf <- performance(x2_pr, measure="tpr", x.measure="fpr")

plot(x2_prf, col="blue", lwd=3)
grid()
abline(0,1,ty=2)

# 정확도
x2_auc <- performance(x2_pr, measure="auc")
x2_auc <- x2_auc@y.values[[1]]
text(0.3, 0.8, paste("auc=", x2_auc))
x2_auc

```

```
> x2_cfm
```

Confusion Matrix and Statistics

	Reference	0	1
Prediction 0	56	91	
Prediction 1	94	59	

(1)번 입력변수 중 x2만 사용할 경우 검증데이터에 대해 정오표, 정확도, ROC 곡선 및 AUC 결과입니다.

```

logreg <- glm(y ~ ., data=spiral_train, family = "binomial")
w0_hat <- coef(logreg)[1]
w1_hat <- coef(logreg)[2]
w2_hat <- coef(logreg)[3]
prob <- predict(logreg, spiral_train, type="response")
y_hat <- ifelse((prob) > 0.5, 1, 0)

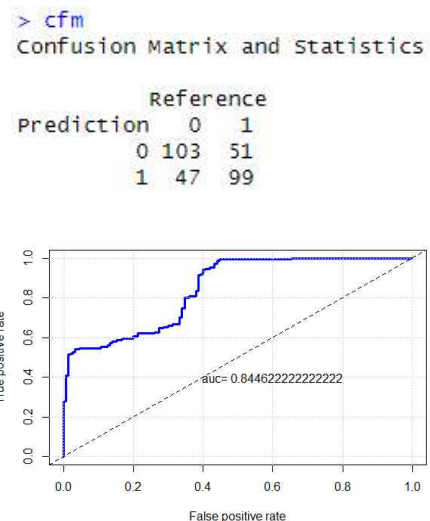
# 정오표
cfm <- confusionMatrix(as.factor(y_hat), as.factor(spiral_train$y))
cfm

pr <- prediction(prob, spiral_train$y)
prf <- performance(pr, measure="tpr", x.measure="fpr")

plot(prf, col="blue", lwd=3)
grid()
abline(0,1,ty=2)

# 정확도
auc <- performance(pr, measure="auc")
auc <- auc@y.values[[1]]
text(0.6, 0.4, paste("auc=", auc))
auc

```



(2)번 입력변수를 모두 사용할 경우 검증데이터에 대해 정오표, 정확도, ROC 곡선 및 AUC 결과입니다.

```

y_pred <- ifelse(predict(logreg, spiral_test, type="response")>0.5, 1, 0)

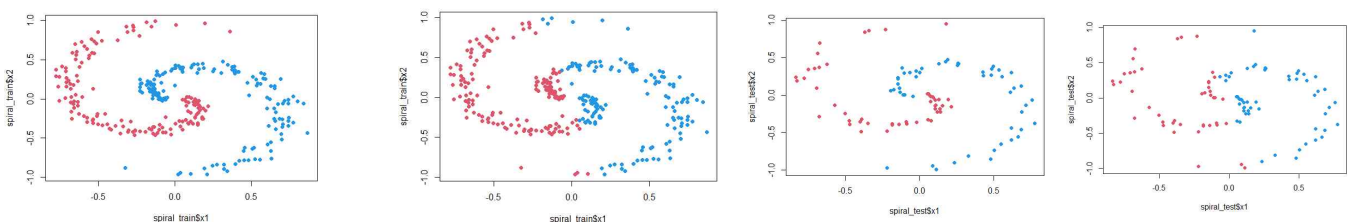
mean((spiral_train$y - y_hat)^2)
mean((spiral_test$y - y_pred)^2)
plot(spiral_train$x1, spiral_train$x2, col = (spiral_train$y+1)*2, pch=19, cex=1)
plot(spiral_train$x1, spiral_train$x2, col = (y_hat+1)*2, pch=19, cex=1)
plot(spiral_test$x1, spiral_test$x2, col = (spiral_test$y+1)*2, pch=19, cex=1)
plot(spiral_test$x1, spiral_test$x2, col = (y_pred+1)*2, pch=19, cex=1)

```

```

> mean((spiral_train$y - y_hat)^2)
[1] 0.3266667
> mean((spiral_test$y - y_pred)^2)
[1] 0.36

```



과대적합을 확인하기 위해서 train과 test 데이터에서 mse를 계산해 보았을 때 크게 차이가 나지 않는 걸 확인가능합니다. 이후 train의 예측값과 실제값, test의 예측값과 실제값을 시각화 해보았을 때 결과를 보았을 때 x1과 x2의 입력변수를 모두 사용한 경우에는 과대적합이 나타나지 않았다고 판단합니다.



		감독교수확인	이 두 호	
2025 학년도 1학기 기말고사	시험과목 : 머신러닝1 (2반)	전공 : AI소프트웨어학과 학번: 202122852    성명:이혜준	출 제 자	이두호
			채점성적	
<p>[6]</p> <p>여름방학에는 캡스톤프로젝트로 제작하였던 가상피팅프로젝트를 팀원들과 함께 논문으로 작성해볼 생각이며, 6월에 잠시 본가에 올라가서 한달 동안만 있다 올 생각입니다. 그리고 2025 하반기부터 기사자격증에 대한 자격이 되어서 정보처리기사를 응시하기 위해 7월에 필기 원서를 작성하고 공부를 할 생각입니다. 또한 토익을 공부하여 방학이 끝나기 전에 2번은 응시하여 최소한으로 750점 이상을 맞춰볼 생각입니다. 이전에 응시했던 SQLD 시험의 결과가 현재 확실하게 예상이 가질 않아 너무 벼락치기로 공부했던 것 같아, 만약 불합격이 나온다면 다시 개념부터 잡고 공부할 것이고, 다시 공부를 하는 김에 Dasp도 찾아보니 배우는 내용이 겹치는 게 상당 부분이 많아서 함께 공부할 생각입니다</p>				