



# 12-2. 로지스틱 회귀분석

## chap12\_2\_LogisticRegression 수업내용

1. Logit 변환
2. Sigmoid Function
3. 이항 로지스틱 회귀
4. 다항 로지스틱 회귀
5. 오분류표[confusion matrix]



# 1. Logit 변환

## ● 오즈비 vs 로짓변환

- ## 1. 오즈비(Odds ratio) : 0(실패)에 대한 1(성공)의 비율(0:no, 1:yes)
  - # no인 상태와 비교하여 yes가 얼마나 높은지 or 낮은지 정량화한 것
  - #  $\text{odds\_ratio} = p(\text{success}) / 1-p(\text{fail})$
  - #  $p : y(\text{반응변수})=1$  이 나올 확률,  $1-P : y(\text{반응변수})=1$ 의 여사건
- ## 2. 로짓변환 : 오즈비에 log 함수 적용
  - #  $\text{logit} = \log(p / 1-p)$



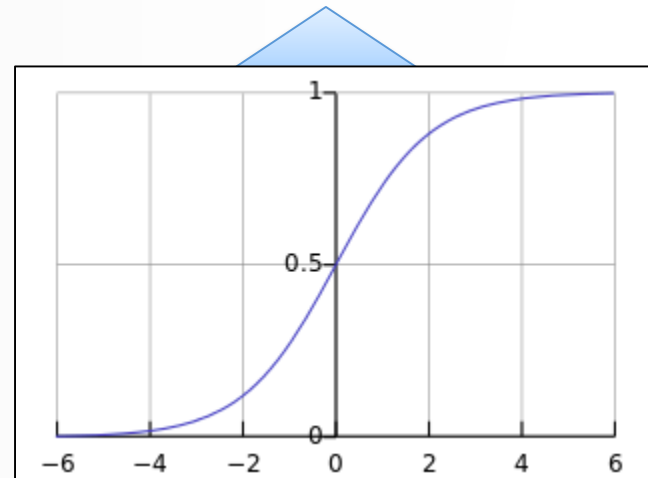
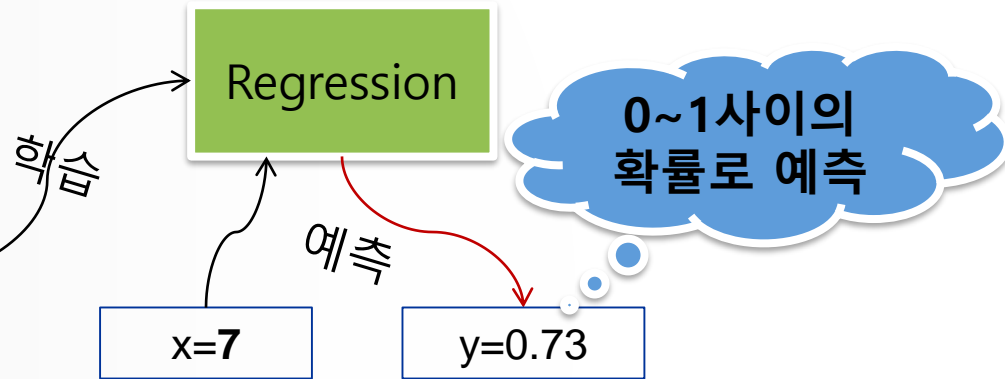
## 2. Sigmoid Function

### ● Sigmoid Function

➤ 합격/불합격 분류

hours	score
10	pass
9	pass
5	fail
3	fail

Training data set



Sigmoid function



### 3. 이항 로지스틱 회귀

- 이항 로지스틱 회귀모형

# 로지스틱 회귀모델 생성 : 학습데이터

```
weater_model <- glm(RainTomorrow ~ ., data = train, family = 'binomial')
```

```
weater_model
```

```
summary(weater_model)
```

# 로지스틱 회귀모델 예측치 생성 : 검정데이터

# newdata=test : 새로운 데이터 셋, type="response" : 0~1 확률값으로 예측

```
pred <- predict(weater_model, newdata=test, type="response")
```

```
Pred
```



## 4. 다항 로지스틱 회귀

### ● 다항 로지스틱 회귀모형

```
model <- multinom(Species ~ ., data = train)
```

```
fit <- model$fitted.values
```

```
# type='response' : 0~1 확률 예측 -> sigmoid 함수(yes/no)
```

```
# type='probs' : 0~1 확률 예측 -> softmax 함수(a, b, c)
```

```
pred_prob <- predict(model, newdata=test, type="probs")
```

```
pred_prob
```



## 5. 오분류표 (confusion matrix)

		예측치	
		Positive	Negative
실제값	POS	TP[참 긍정]	FN[거짓 부정]
	NEG	FP[거짓 긍정]	TN[참 부정]

정분류율(Accuracy) =  $(TP + TN) / \text{전체관측치}$

오분류율(Inaccuracy) =  $(FN + FP) / \text{전체관측치}$

정확률(Precision) =  $TP / (TP + FP)$

재현율(Recall) =  $TP / (TP + FN)$

F 측정치(F measure) =  $2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$

정분류율(Accuracy) : 알고리즘의 성능평가 척도

오분류율(Inaccuracy) : 알고리즘의 오차 비율

정확률(Precision) : 알고리즘이 Yes로 판단한 것 중에서 실제로 Yes인 비율

재현율(Recall) : 실제값이 Yes인 것 중에서 알고리즘이 Yes로 판단한 비율

F 측정치(F measure) : 정확률과 재현율을 동시에 고려하는 측정치



# 오분류표(confusion matrix)와 ROC 그래프

## 예측치

실제값

	Positive	Negative
POS	TP[참 긍정]	FN[거짓 부정]
NEG	FP[거짓 긍정]	TN[참 부정]

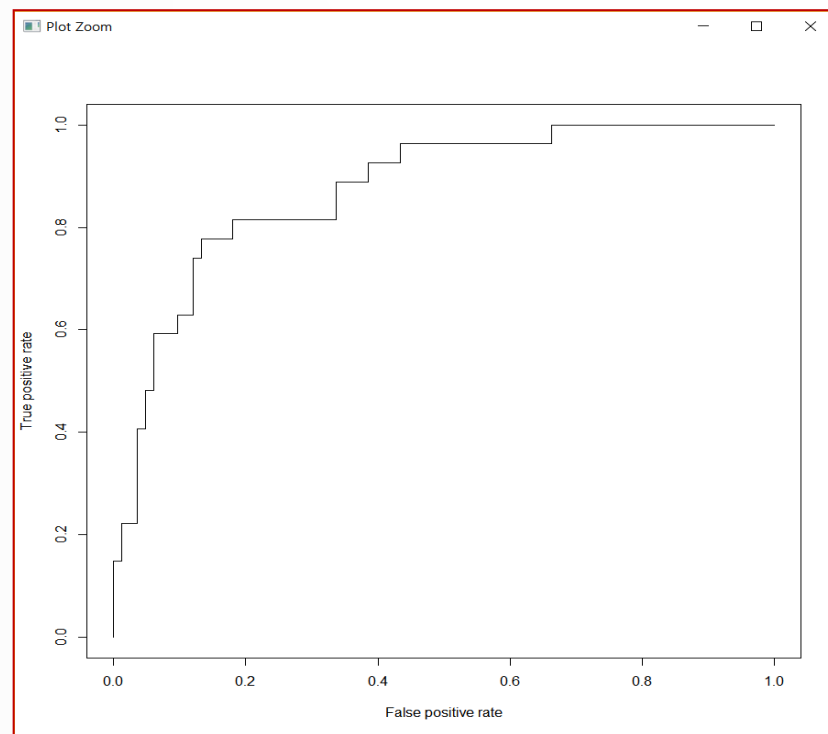
민감도

민감도(Sensitivity) =  $TP / (TP + FN)$

특이도(Specificity) =  $TN / (FP + TN)$

민감도(Sensitivity) : 실제값 Yes인 경우 Yes 예측 비율  
= 재현율(Recall)

특이도(Specificity) : 실제값 No인 경우 No 예측 비율



특이도(Specificity)