확률과 통계

섹션 - 6

강사 : James 쌤

로지스틱회귀 원리

키포인트

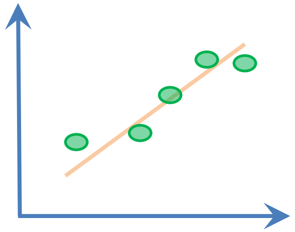
- 로지스틱 회귀의 원리.
- 로지스틱 회귀를 적용한 학습과 예측.

지도학습 머신러닝

지도학습의 세분화

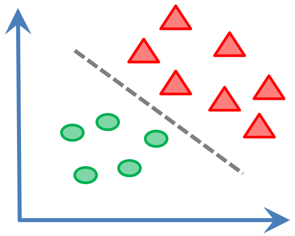
수치형 Y

$Y = 13.45,\ \ 73,\ \ 9.5,\ \ .....$



명목형 Y

$Y =$ *red*, *green*, ...



지도학습 머신러닝

지도학습의 평가

수치형 Y

MSE, MAE, RMSE,

Correlation, 등.

명목형 Y

Confusion Matrix, Accuracy,

Precision, Recall, Specificity, 등.

로지스틱회귀 개요

- 로지스틱 회귀는 이진 분류 방법이다.
- 한 개 이상의 독립변수 (설명변수)가 있다: $X_1, X_2, \dots, X_K$
- 한 개의 종속변수 (반응변수)가 있다: Y
- 종속변수의 값은 0 또는 1이다: 이분법적인 상황을 모델링한다.
- 조건부 확률 예측에 기반하여 종속변수의 값 (0 or 1)을 예측한다.

로지스틱회귀 기초

- K 개의 독립변수 (설명변수)가 있다고 가정한다.

  $\Rightarrow$ 임의의 실수 값을 가질 수 있다.

$$X_1, X_2, \ldots, X_K$$

로지스틱회귀 기초

- 그리고 한개의 종속변수가 있다고 가정한다.

$\Rightarrow$ 그런데 가능한 값은 0과 1로 국한되어 있다.

$$Y = \begin{cases} 1 \\ 0 \end{cases}$$

로지스틱회귀 기초

- 즉, 이분법적인 상황이다.

$$Y = \begin{cases} \text{참 } (True) \\ \text{거짓 } (False) \end{cases}$$

로지스틱회귀 기초

- 즉, 이분법적인 상황이다.

$$Y = \begin{cases} \text{유형 } a \\ \text{유형 } b \end{cases}$$

로지스틱회귀 원리

- 이제는 독립변수 $\{X_i\}$를 선형조합하여 S 변수 (logit)을 만든다.

$$S = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_K X_K$$

독립변수

(데이터로 값이 주어짐)

로지스틱회귀 원리

- 종속변수Y의 값이 1이 될 조건부 확률 $P(Y = 1|\{x_i\})$은 "로지스틱 함수"또는 "Sigmoid 함수"를 사용해서 계산된다.

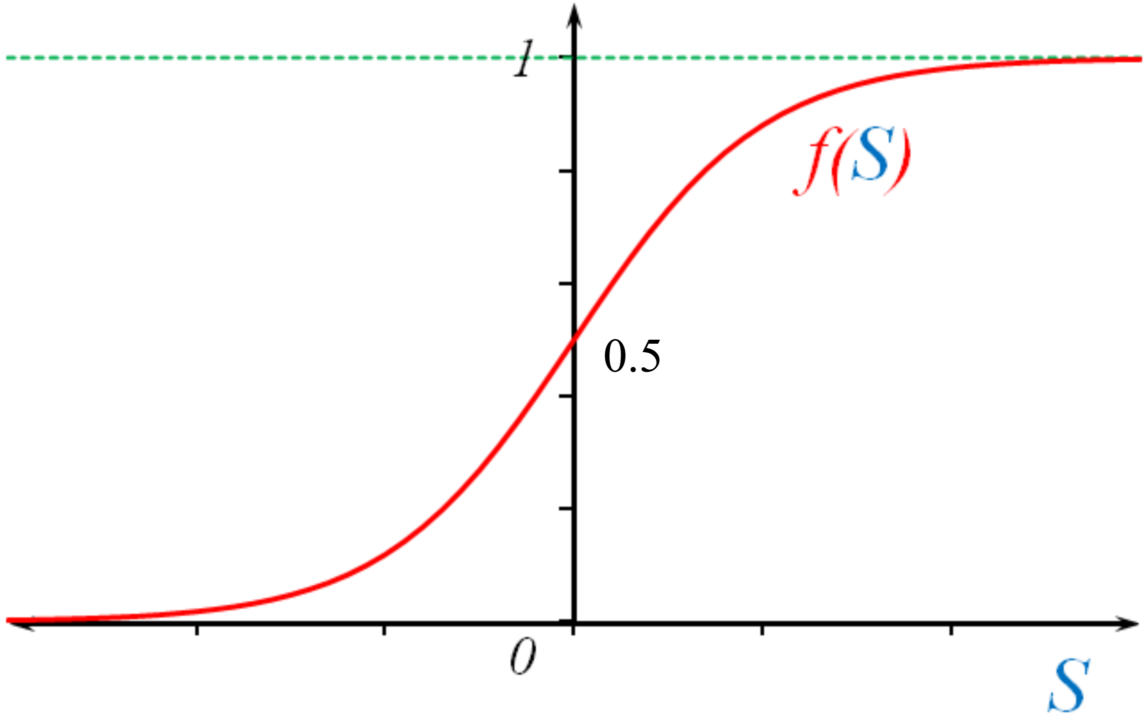
$$f(S) = \frac{e^S}{1 + e^S}$$

$\Rightarrow$ 인공신경망에서 "활성화 함수" (activation function)의 역할을 한다.

$\Rightarrow$ Logit S와 확률 P사이의 관계는 다음과 같다.

$$S = Log\left(\frac{P}{1 - P}\right)$$

로지스틱회귀 원리



로지스틱회귀 학습

- 학습이란 모형의 파라미터, β_i 계수들의 값을 구하는 것을 의미한다.

$$S = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_K X_K$$

$$f(S) = \frac{e^{S}}{1+e^{S}}$$

로지스틱회귀 학습

- 로그우도 L 을 최소화하는 방법으로 β_i 계수들의 값을 구할 수 있다.
- 로그우도 L 은 일종의 “손실함수”이다.

$\Rightarrow$ 예측 오류에 의한 “손실”을 최소화 하고자 한다.

로지스틱회귀 학습

- 로그우도 L의 수식은 다음과 같다:

$$L(\vec{\beta}) = -\sum_{i=1}^{N} Log\left(1 + e^{-y_i \vec{\beta}^t \vec{x}_i}\right)$$

- L의 값은 gradient 방향으로 증가율 최고.

$\Rightarrow$ 그리고 -gradient 방향으로는 감소율 최고.

로지스틱회귀 학습

- L의 gradient는 다음과 같이 계산할 수 있다.

$$\overrightarrow{\nabla L} = -\sum_{i=1}^{N} \frac{y_i \vec{x}_i e^{-\vec{\beta}^t \vec{x}_i}}{1 + e^{-\vec{\beta}^t \vec{x}_i}}$$

$\Rightarrow$ L을 미분하여 구할 수 있다.

$\Rightarrow$ $\vec{x}_i$와 y_i는 실제 데이터 값을 의미한다.

로지스틱회귀 학습

- Gradient descent 알고리즘은 손실함수 $L(\vec{\beta})$를 수렴적으로 최소화 시켜준다.

$\Rightarrow$ $\vec{\beta}$를 감소율이 최고인 $-\nabla L(\vec{\beta})$ 방향으로 조금씩 이동시켜 간다.

a). $\vec{\beta}$를 임의의 값으로 초기화 한다.

b). Gradient $\overrightarrow{\nabla L}$를 계산한다.

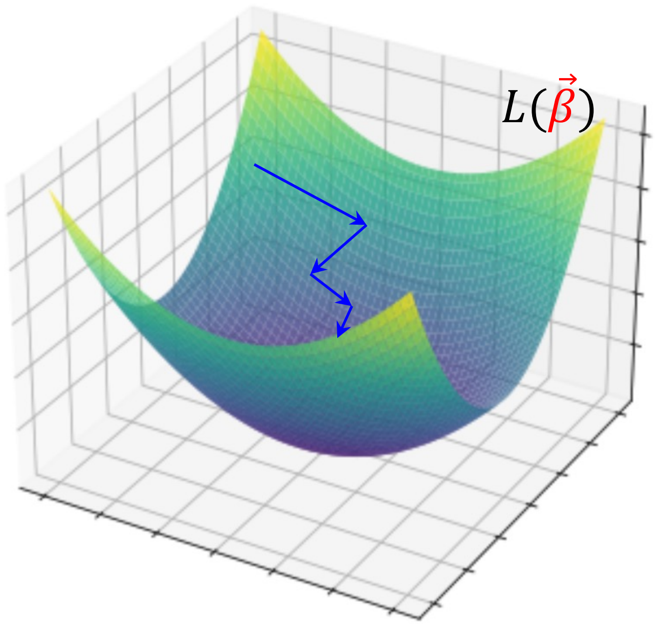
c). $\vec{\beta}$를 $\vec{\beta} - \eta\overrightarrow{\nabla L}$와 같이 갱신한다. "Learning rate" η로 수렴 속도 조절.

d). 스텝 b)로 돌아가서 일정 횟수만큼 반복한다.

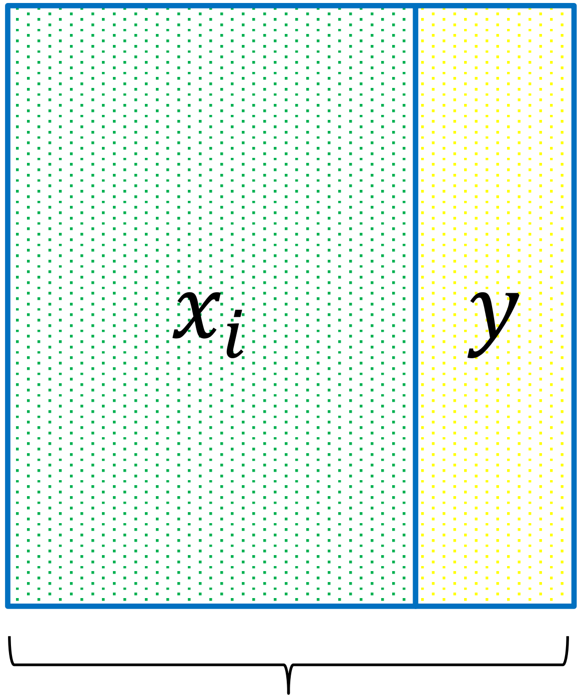
로지스틱회귀 학습

- Gradient descent 알고리즘은 손실함수 $L(\vec{\beta})$를 수렴적으로 최소화 시켜준다.

$\Rightarrow$ $\vec{\beta}$를 감소율이 최고인 $-\nabla L(\vec{\beta})$ 방향으로 조금씩 이동시켜 간다.



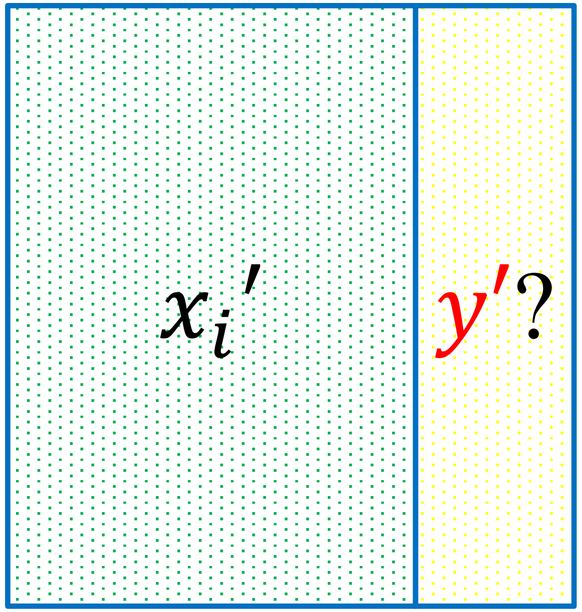
로지스틱회귀 학습



Training set

모형의 파라미터, 즉 $\{\beta_i\}$를 학습용 데이터를 사용하여 계산해 놓는다.

로지스틱회귀 예측



독립변수의 값들 $\{x_i'\}$이 새롭게 주어졌을 때, 모르는 상태인 종속변수의 값 y'을 계산을 통해서 알아낸다.

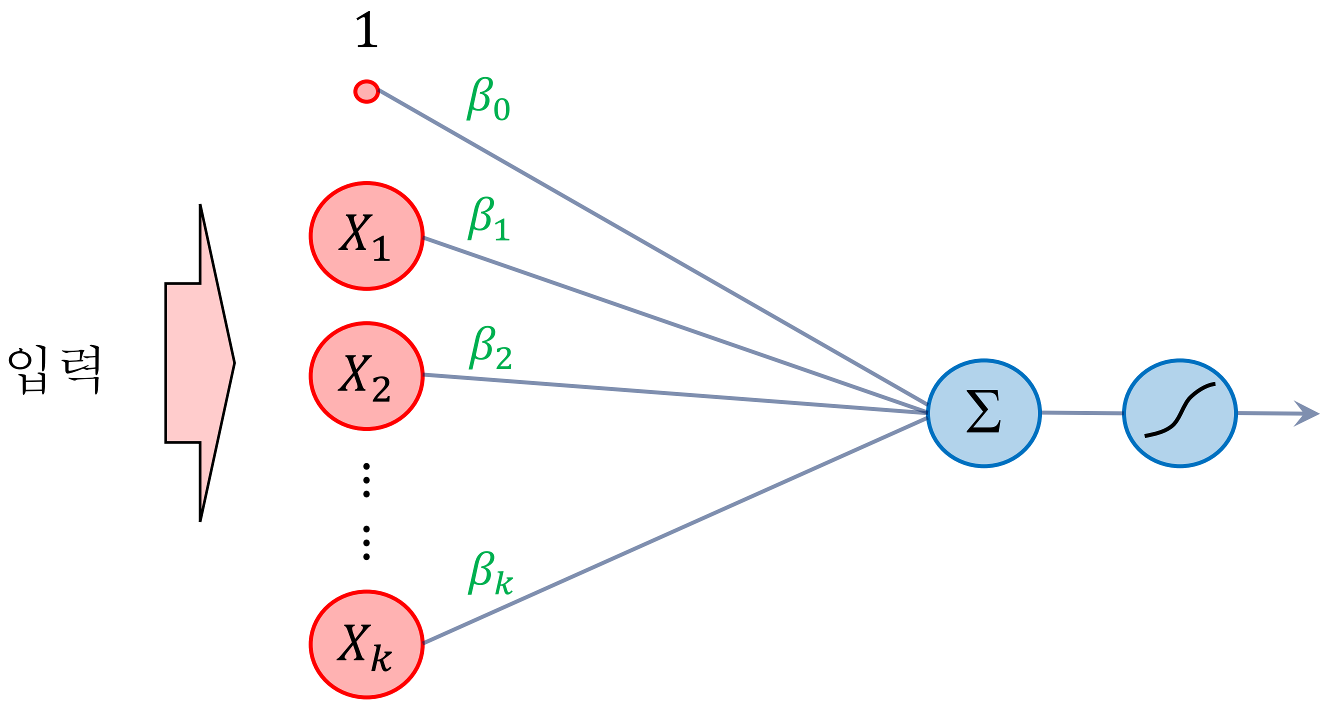
$\Downarrow$

"조건부 확률"

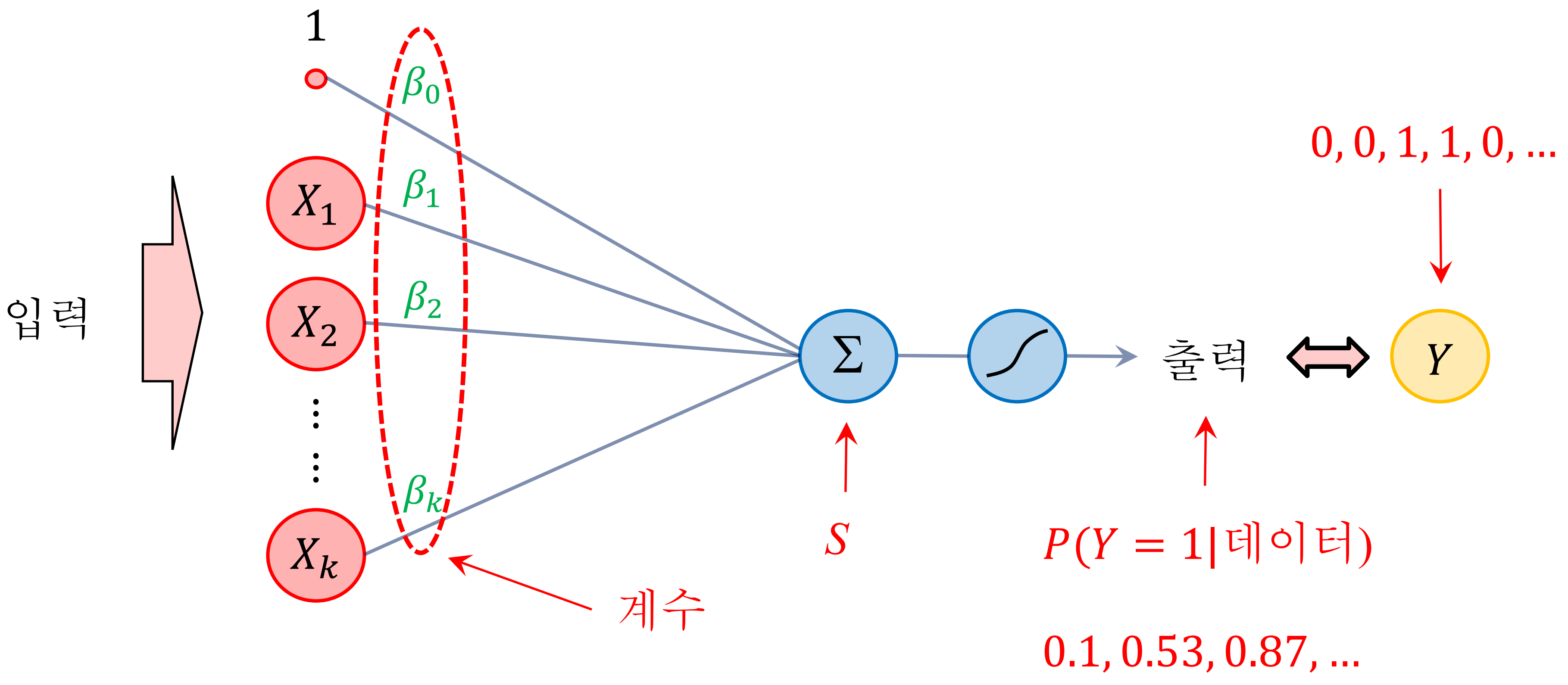
$P(Y' = 1|\{x_i'\})$ > 기준확률?

$\rightarrow \hat{y}' = 1$ or 0

로지스틱회귀 예측



로지스틱회귀 예측



1
β_0
β_1
β_2
β_k
입력
X_1
X_2
X_k
계수
Σ
S
출력
Y
0, 0, 1, 1, 0, …
$P(Y=1|\text{데이터})$
0.1, 0.53, 0.87, …

로지스틱회귀 평가

키포인트

- 혼동행렬 (confusion matrix).
- 정확도, 민감도, 특이도, 정밀도.
- ROC 곡선의 원리. ⇒ 기준확률과의 관계
- ROC 곡선과 AUC.
- 베이즈 정리를 활용한 예측결과 해석.

혼동행렬

	Actual 0	Actual 1
Predicted 0	134	42
Predicted 1	10	14

정확도 (Accuracy)

	Actual 0	*Actual 1*
Predicted 0	134	42
Predicted 1	10	14

⇒ 정확도는 행렬의 대각선의 합과 전체의 합 사이의 비율.

⇒ 예측된 유형이 실제 유형과 일치하는 비율.

정확도의 맹점

- 그런데 정확도 만으로 테스트가 불가능한 상황이 종종 발생한다.

  예). 은행 대출고객 중에서 3% ~ 5%만이 향후 신용불량인 경우.

  ⇒ 목표는 소수인 신용불량 고객을 사전에 검출하는 것.

  ⇒ 만약에 모두를 신용양호로 예측한다면 정확도는 **매우 높다**!

  ⇒ 하지만 신용불량 고객은 한 명도 예측하지 못한다.

성능 지표

- $\text{Accuracy (정확도)} = \dfrac{\text{정확하게 예측된 개수}}{\text{전체 개수}}$

- $\text{Sensitivity (민감도)} = \dfrac{\text{정확하게 예측된 1의 개수}}{\text{실제 1의 총 개수}}$

- $\text{Specificity (특이도)} = \dfrac{\text{정확하게 예측된 0의 개수}}{\text{실제 0의 총 개수}}$

성능 지표

- Precision (정밀도) = $\dfrac{\text{정확하게 예측된 1의 개수}}{\text{1이라고 예측된 개수}}$

- Recall (재현율) = 민감도와 같은 의미

- Cohen의 카파 $\kappa = \frac{Accuracy - p_e}{1 - p_e}$ $\Leftarrow$ p_e는 우연으로 맞을 확률.

민감도 (Sensitivity)

	Actual 0	*Actual 1*
Predicted 0	134	42
Predicted 1	10	14

⇒ 민감도는 실제 1 중에서 정확하게 1이라 예측된 비율.

특이도 (Specificity)

	Actual 0	*Actual 1*
Predicted 0	134	42
Predicted 1	10	14

⇒ 특이도는 실제 0 중에서 정확하게 0이라 예측된 비율.

정밀도 (Precision)

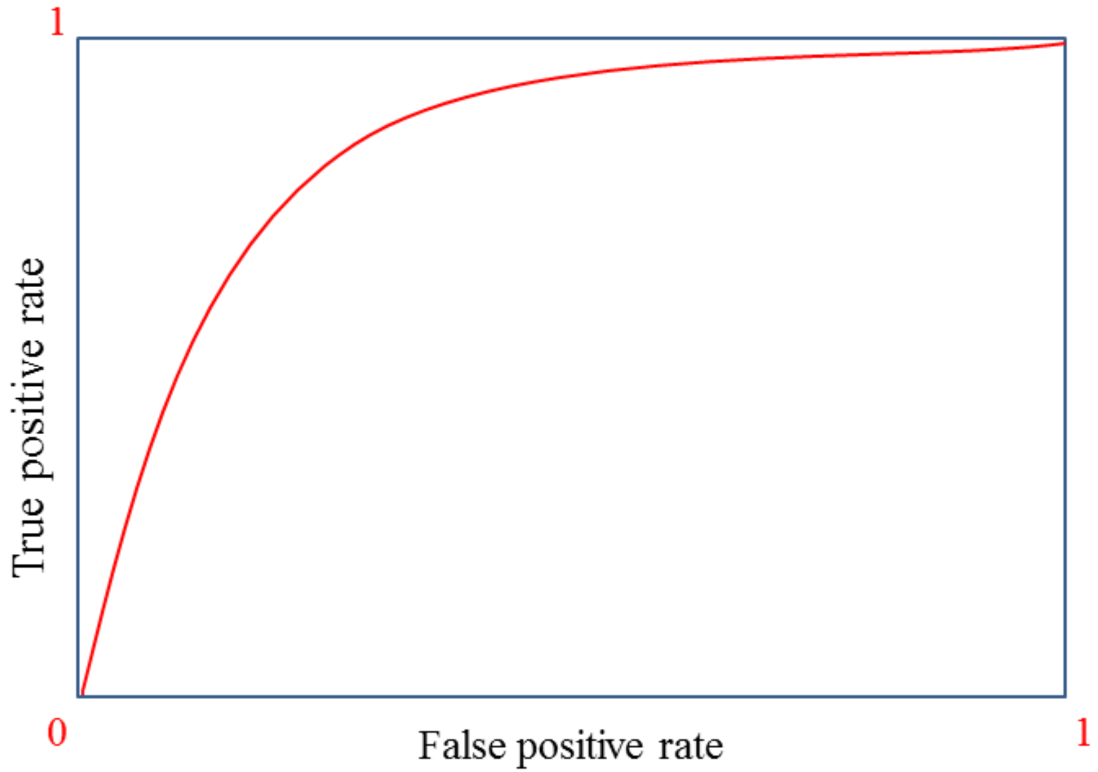
	Actual 0	*Actual 1*
Predicted 0	134	42
Predicted 1	10	14

⇒ 정밀도는 1이라 예측된 경우 중에서 정답의 비율.

성능 지표

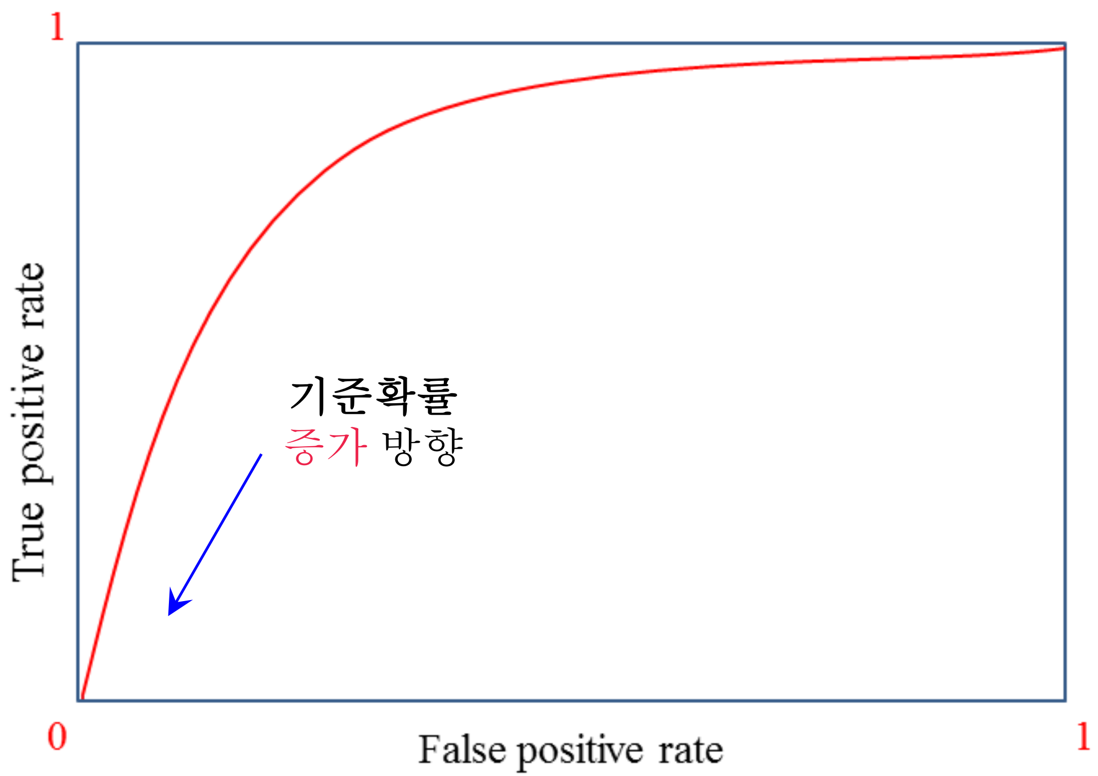
- True Positive Rate = Sensitivity
- True Negative Rate = Specificity
- False Positive Rate = $\dfrac{\text{실제는 0 이지만 1로 오인식된 개수}}{\text{실제 0의 개수}}$ = 1- Specificity
- False Negative Rate = $\dfrac{\text{실제는 1 이지만 0으로 오인식된 개수}}{\text{실제 1의 개수}}$ = 1- Sensitivity
- Positive Predicted Value = Precision

ROC 곡선



⇒ ROC 곡선은 기준확률에 대한 parametric plot이다.

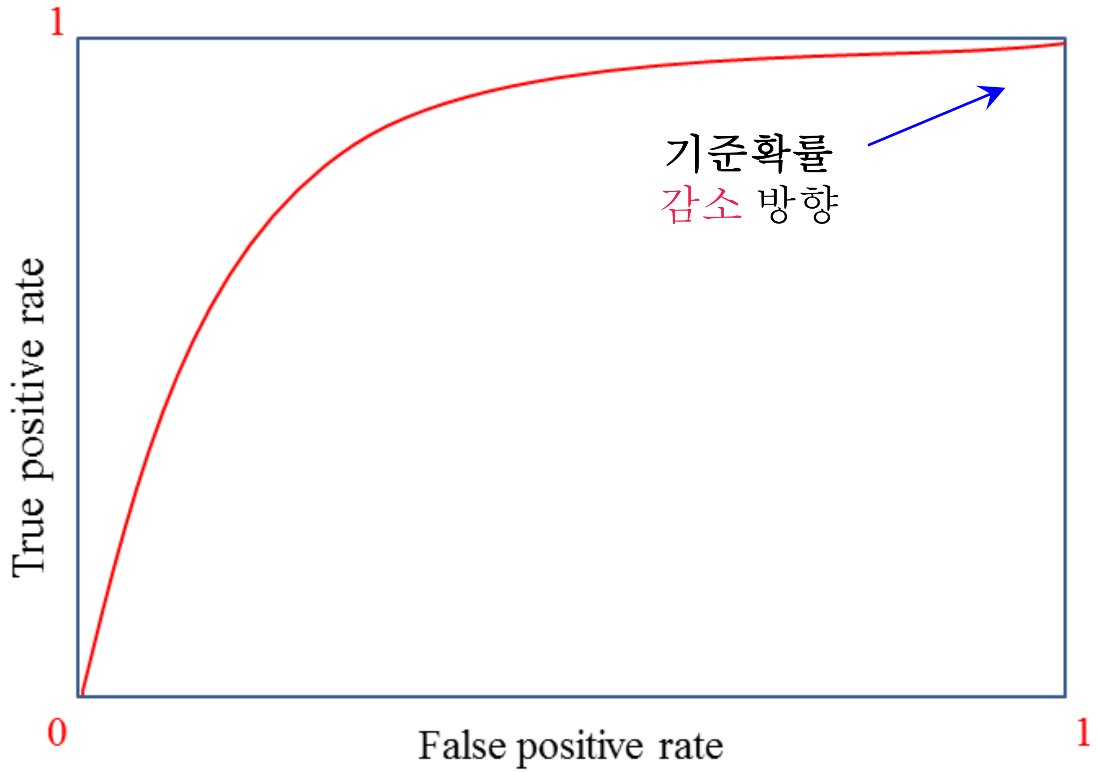
ROC 곡선



기준확률이 증가 (1에 가까워 진다)	
성능척도	방향
민감도 (True Positive)	⇓
특이도	⇑
1-특이도 (False Positive)	⇓
정밀도	⇑

⇒ ROC 곡선은 기준확률에 대한 parametric plot이다.

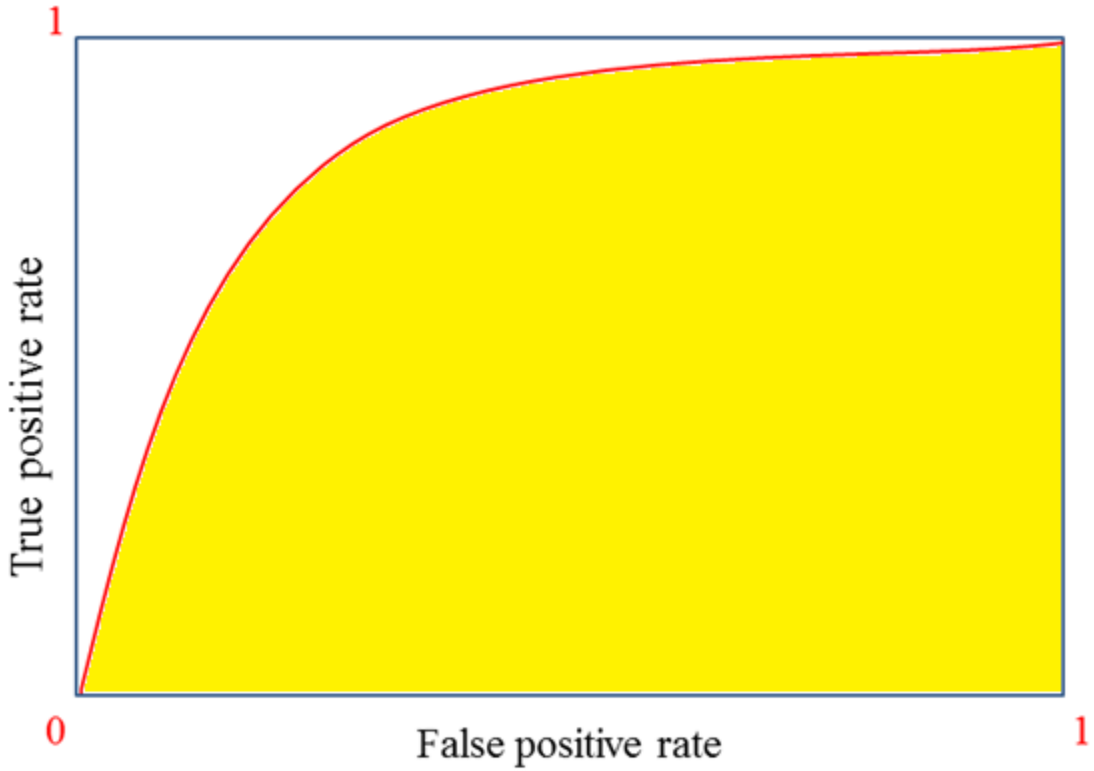
ROC 곡선



기준확률이 감소 (0에 가까워 진다)	
성능척도	방향
민감도 (True Positive)	⇑
특이도	⇓
1-특이도 (False Positive)	⇑
정밀도	⇓

⇒ ROC 곡선은 기준확률에 대한 parametric plot이다.

ROC 곡선과 AUC



⇒ AUC는 곡선 아래의 면적 (**A**rea **U**nder the **C**urve)을 의미한다.

⇒ AUC가 클수록 (1에 가까울 수록) 예측성능이 좋은 것이다.

로지스틱회귀 예측과 베이즈 정리

문제: 500개의 관측치가 있다. 이 중에서 종속변수의 값이 1인 경우는 30회이고 0인 경우는 나머지 470회이다. 그런데 로지스틱회귀 모형의 민감도는 0.92이고 특이도는 0.90이다. 만약에 이 모형을 가지고 한 예측결과가 1이라면 어느정도 믿을 수 있겠는가?

로지스틱회귀 예측과 베이즈 정리

문제: 500개의 관측치가 있다. 이 중에서 종속변수의 값이 1인 경우는 30회이고 0인 경우는 나머지 470회이다. 그런데 로지스틱회귀 모형의 민감도는 0.92이고 특이도는 0.90이다. 만약에 이 모형을 가지고 한 예측결과가 1이라면 어느정도 믿을 수 있겠는가?

이 문제의 조건을 정리해 보면 다음과 같다.

$P(\text{예측 } 1|\text{실제 } 1) = 0.92$ "민감도"

$P(\text{예측 } 0|\text{실제 } 0) = 0.90$ "특이도"

$\Rightarrow P(\text{예측 } 1|\text{실제 } 0) = 1 - P(\text{예측 } 0|\text{실제 } 0) = 0.10$

$P(1) = 30/500 = 0.06$

로지스틱회귀 예측과 베이즈 정리

문제: 500개의 관측치가 있다. 이 중에서 종속변수의 값이 1인 경우는 30회이고 0인 경우는 나머지 470회이다. 그런데 로지스틱회귀 모형의 민감도는 0.92이고 특이도는 0.90이다. 만약에 이 모형을 가지고 한 예측결과가 1이라면 어느정도 믿을 수 있겠는가?

그러므로 이 문제가 요구하는 답은 $P(\text{실제 }1|\text{예측 }1)$이다. 예측 이것을 베이즈 정리를 적용하여 계산해 보면 다음과 같다.

$$P(\text{실제 }1|\text{예측 }1) = \frac{P\left(\text{예측 }1|\text{실제 }1\right)P(1)}{P\left(\text{예측 }1|\text{실제 }1\right)P(1) + P\left(\text{예측 }1|\text{실제 }0\right)P(0)}$$

$$= \frac{0.92 \times 0.06}{0.92 \times 0.06 + 0.1 \times 0.94} \cong \mathbf{0.37}$$

끝