



부산대학교

PUSAN NATIONAL UNIVERSITY

Linear Regression Revisited

Linear Models

Regression, Classification, and Loss

BX3600087-140 / DS3600087-140 / AB3600087-140

기계학습개론

2026학년도 2학기

오용경

데이터사이언스전공

yongkyung.oh@pusan.ac.kr

Section 1 • Linear Regression Revisited

Linear Model

선형식의 형태와 매개변수를 구분한다

선형식은 점수 하나를 만든다

선형 모델(linear model)은 설명변수(predictor)의 가중합을 먼저 계산한다

$$z = \beta_0 + \sum_{j=1}^p \beta_j x_j$$

기호	이름	역할
β_0	절편(intercept)	모든 설명변수가 0일 때의 기준값
β_j	계수(coefficient)	설명변수 x_j 에 곱하는 가중치
x_j	설명변수	관측에서 모델에 넣는 값

β_0, β_j 를 θ_0, θ_j 로 쓰기도 한다. 설명변수 x_j 를 특성이라고도 부른다.

Source: James et al. (2023), Section 3.2.

수치형 반응변수의 값을 예측한다

선형 회귀(linear regression)는 선형식의 값을 예측값(prediction)으로 쓴다

$$\hat{y} = \beta_0 + \beta_1 x$$

구분	뜻	예
반응변수(response) y	관측한 수치	판매량
예측값 $\hat{y}$	모델이 낸 수치	예상 판매량
회귀	수치를 예측하는 문제	광고비로 판매량 예측

Source: James et al. (2023), Section 3.1.

설명변수가 하나면 직선이다

단순 선형 회귀는 설명변수가 하나인 선형 회귀다

$$\hat{y} = \beta_0 + \beta_1 x$$

DEFINITION

절편 β_0

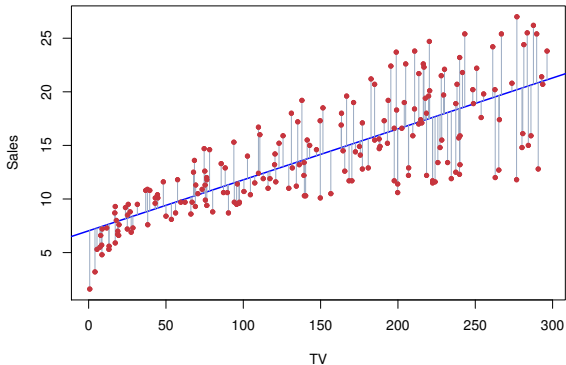
$x = 0$ 일 때의 예측값, 세로축과 만나는 높이

DEFINITION

계수 β_1

x 가 1 증가할 때 예측값의 변화량인 기울기

직선과 관측 사이에 차이가 남는다



파란 직선은 예측값이고 세로 선분은 관측값과 예측값의 차이이다

Source: James et al. (2023), Figure 3.1.

절편과 통계적 편향은 다르다

용어	정의	다루는 맥락
절편 β_0	모델이 학습하는 매개변수	선형식의 상수항
통계적 편향(bias)	반복 학습한 평균 예측의 오차	모델의 일반화 성질

신경망 코드에서는 절편을 bias라 부르기도 하지만, 표의 통계적 편향은 편향-분산 분해의 편향이다

절편은 식의 상수항이고, 통계적 편향은 평균 예측의 체계적인 오차다

Source: James et al. (2023), Sections 2.2.2 and 3.1.

Checkpoint

Linear Model

- β_0 와 β_1 은 직선의 무엇을 정하는가
- 관측값 y 와 예측값 $\hat{y}$ 는 무엇이 다른가
- 절편과 통계적 편향을 구분하는 기준은 무엇인가

Section 2 • Linear Regression Revisited

Loss and Least Squares

잔차로 계수를 고르는 기준을 정한다

잔차는 관측값과 예측값의 차이이다

잔차(residual)는 관측값에서 예측값을 뺀 값이다

$$e_i = y_i - \hat{y}_i$$

상황	잔차	뜻
$y_i > \hat{y}_i$	양수	과소 예측
$y_i < \hat{y}_i$	음수	과대 예측
$y_i = \hat{y}_i$	0	관측과 예측이 일치

아래 첨자 i 는 i 번째 관측을 뜻하며, $y^{(i)}$ 로도 쓴다

한 관측의 예측 오차를 손실로 나타낸다

손실함수(loss function)는 예측값과 관측값의 차이에 벌점을 부여한다

$$\mathcal{L}(\hat{y}_i, y_i) = (y_i - \hat{y}_i)^2 = e_i^2$$

e_i	e_i^2	벌점
1 또는 -1	1	기준(1배)
2 또는 -2	4	네 배
3 또는 -3	9	아홉 배

잔차를 제공하는 이유

성질	결과
부호를 없앤다	직선 위와 아래의 오차가 서로 상쇄되지 않는다
큰 오차를 더 무겁게 센다	멀리 빗나간 관측이 더 큰 벌점을 받는다
손실이 매끄러워진다	계수를 바꿀 때 손실의 변화 방향을 계산할 수 있다

Source: James et al. (2023), Section 3.1.1.

자료 전체에서는 RSS를 계산한다

잔차 제곱합(residual sum of squares, RSS)은 모든 잔차를 제곱해 더한 값이다

$$\text{RSS}(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$$

바뀌는 것	결과
β_0, β_1	직선과 모든 잔차가 바뀐다
각 관측의 잔차	제곱해 더하면 RSS가 된다

Source: James et al. (2023), Section 3.1.1.

최소제곱법은 RSS가 가장 작은 계수를 고른다

DEFINITION

최소제곱법(least squares)

학습 자료의 RSS를 가장 작게 만드는 계수를 고르는 추정 방법

후보

계수로 직선을 만든다

비교

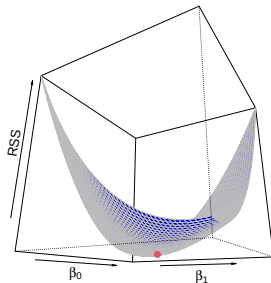
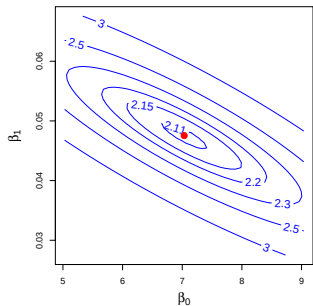
직선마다 RSS를 계산

선택

RSS가 최소인 계수

Source: James et al. (2023), Section 3.11.

RSS는 계수에 따라 달라지는 함수다



같은 RSS 함수를 등고선과 3차원 표면으로 나타냈다. 빨간 점은 RSS가 최소인 계수다

Source: James et al. (2023), Figures 3.2a and 3.2b.

최소제곱해를 행렬로 나타낼 수 있다

행렬 X 의 첫 열을 1로 두면 절편까지 같은 식으로 계산할 수 있다

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

조건	의미
$X^T X$ 가 가역	X 의 열이 선형독립이면 최소제곱해가 하나로 정해진다
가역이 아님	학습 자료에서 같은 적합값을 만드는 계수가 여럿이다. 새 입력의 예측값은 달라질 수 있다

X 는 관측을 행으로 쌓고 첫 열을 1로 둔 행렬, y 는 관측값 벡터다. $(X^T X)^{-1}$ 은 $X^T X$ 의 역행렬이다

Checkpoint

Loss and Least Squares

- 잔차의 부호로 과소 예측과 과대 예측을 어떻게 구분하는가
- 잔차의 크기가 두 배가 되면 제곱 손실은 몇 배가 되는가
- 최소제곱법은 어떤 기준으로 계수를 정하는가

Estimation and Prediction

계수를 정하는 일과 적용하는 일을 구분한다

같은 식을 두 단계에서 쓴다

추정(estimation)과 예측은 입력과 결과가 다르다

단계	입력	결과
추정	학습 자료 (X, y)	계수 $\hat{\beta}$
예측	새 입력 x_0 와 $\hat{\beta}$	예측값 $\hat{y}_0$

학습 자료로 계수를 추정한 뒤, 새 입력에 그 계수를 적용해 예측한다

Source: James et al. (2023), Section 3.1.

행렬 곱으로 관측 전체를 예측한다

다중 선형 회귀(multiple linear regression)는 여러 설명변수의 가중합을 사용한다

$$\hat{y} = X\hat{\beta}$$

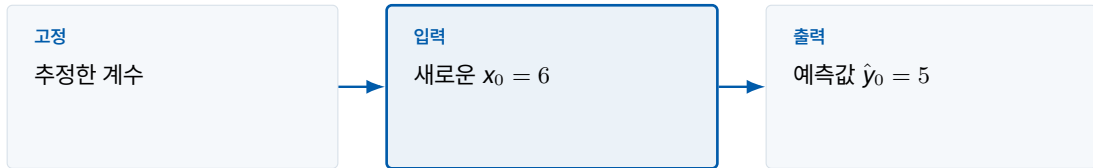
기호	크기	담는 것
X	$n \times (p + 1)$	관측을 행으로 쌓은 설명변수 행렬
$\hat{\beta}$	$(p + 1) \times 1$	추정한 계수 벡터
$\hat{y}$	$n \times 1$	관측마다 하나씩 나온 예측값

p 는 설명변수 수다

새 입력에는 추정한 계수를 적용한다

$\hat{\beta}_0 = 2, \hat{\beta}_1 = 0.5$ 인 모델에 $x_0 = 6$ 을 넣는다

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0 = 2 + 0.5 \times 6 = 5$$



계수는 다른 설명변수가 같을 때의 변화를 나타낸다

다른 설명변수를 고정하고 x_j 가 1 증가할 때의 예측값 변화를 읽는다

$\hat{\beta}_j$	예측값 변화	주의
양수	$\hat{y}$ 가 증가	인과를 뜻하지 않음
음수	$\hat{y}$ 가 감소	인과를 뜻하지 않음
0에 가까움	선택한 단위에서 변화가 작음	다른 설명변수와 단위에 따라 달라짐

Advertising 자료의 계수: TV 0.046, radio 0.189, newspaper -0.001. 광고비는 1,000달러, 판매량은 1,000개 단위다. 다른 두 광고비를 고정하고 TV 광고비를 1,000달러 늘리면 예측 판매량이 약 46개 늘어난다

Source: James et al. (2023), Section 3.2.1.

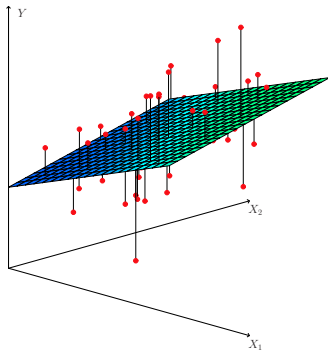
단위가 바뀌면 계수도 바뀐다

같은 관계라도 설명변수의 단위에 따라 계수의 크기가 달라진다

키의 단위	계수	10 cm 증가에 따른 예측값 변화
cm	0.4	$0.4 \times 10 = 4$
m	40	$40 \times 0.1 = 4$

계수의 절댓값만으로 설명변수의 중요도를 비교할 수 없다

설명변수가 둘이면 평면을 적합한다



$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ 는 두 설명변수에 대한 예측 평면이다

Source: James et al. (2023), Figure 3.4.

설명변수의 종류에 따라 식의 항이 달라진다

설명변수	식에 들어가는 항	계수의 뜻
수치형 설명변수	$\beta_1 x_1$	1 단위 증가에 따른 변화
범주의 0/1 표현	$\beta_2 d$	기준 범주와의 차이
상호작용(interaction)	$\beta_3 x_1 x_2$	한 설명변수의 영향이 다른 설명변수의 값에 따라 달라진다

d 는 범주가 A이면 1, 기준 범주 B이면 0인 dummy variable이다. 범주가 k 개이면 0/1 열은 $k - 1$ 개를 쓴다

Checkpoint

Estimation and Prediction

- 추정과 예측의 입력은 무엇이 다른가
- 설명변수의 단위가 바뀌면 계수는 왜 달라지는가
- 범주를 0/1로 나타낸 변수의 계수는 무엇과의 차이인가

From Values to Probabilities

회귀의 선형식에서 무엇을 유지하고 바꿀지 구분한다

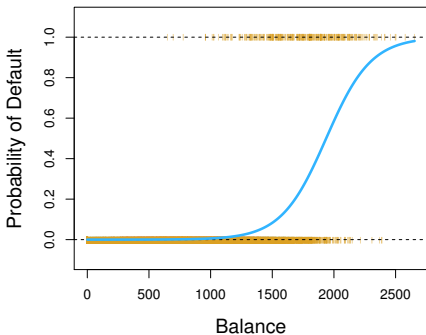
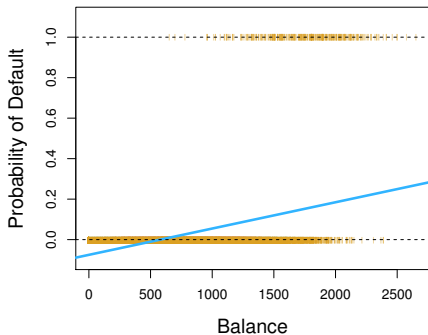
범주형 반응변수에서는 출력의 의미가 달라진다

분류(classification)는 반응변수가 범주인 문제다

문제	반응변수	모델의 출력
회귀	판매량 같은 수치	예측값 $\hat{y}$
분류	연체 여부 같은 범주	범주 1일 추정 확률

$$p(x) = \Pr(Y = 1 \mid X = x)$$

직선을 확률에 그대로 쓰면 범위를 벗어난다



왼쪽 직선은 확률 범위를 벗어나지만 오른쪽 곡선은 0과 1 사이에 머문다

Source: James et al. (2023), Figure 4.2.

sigmoid 함수는 실수를 확률 범위로 바꾼다

sigmoid 함수는 S자 모양의 곡선으로, 실수 z 를 $(0, 1)$ 의 값으로 바꾼다

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

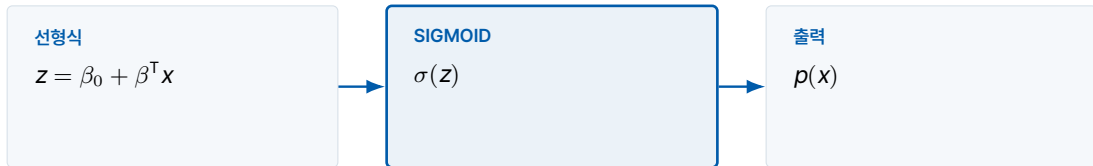
z	$\sigma(z)$	확률의 위치
크게 음수	0에 가까움	낮음
0	0.5	가운데
크게 양수	1에 가까움	높음

Source: MIT 6.390 (2026), Classification.

선형 점수 뒤에 sigmoid 함수를 붙인다

로지스틱 회귀(logistic regression)는 선형 점수를 sigmoid 함수에 넣는다

$$p(x) = \sigma(\beta_0 + \beta^T x) = \frac{1}{1 + \exp[-(\beta_0 + \beta^T x)]}$$



$\beta = (\beta_1, \dots, \beta_p)^T$ 는 절편 β_0 을 뺀 계수 벡터이고, x 에는 1을 넣지 않는다.

로그 승산을 선형식으로 나타낸다

승산(odds)은 발생 확률과 비발생 확률의 비이며, 로그 승산(log odds)은 그 로그다

$$\text{odds} = \frac{p}{1 - p},$$

$$\log \text{odds} = \log \frac{p(x)}{1 - p(x)} = \beta_0 + \beta^T x$$

p	승산	로그 승산의 부호
0.2	1/4	음수
0.5	1	0
0.8	4	양수

Source: James et al. (2023), Section 4.3.1.

같은 계수라도 출력에 따라 해석이 달라진다

다른 설명변수를 고정하고 x_j 가 1 증가할 때의 변화를 비교한다

모델	β_j 가 나타내는 변화	확률 변화
선형 회귀	예측값이 β_j 만큼 변함	확률을 출력하지 않음
로지스틱 회귀	로그 승산이 β_j 만큼 변함	현재 $p(x)$ 에 따라 달라짐

로지스틱 회귀의 계수를 확률의 고정 증가량으로 읽지 않는다

Source: James et al. (2023), Section 4.3.2.

확률 예측과 범주 결정은 서로 다른 단계다

임계값(threshold) t 를 기준으로 예측 범주를 정한다

$$\hat{y} = \begin{cases} 1, & p(x) \geq t, \\ 0, & p(x) < t \end{cases}$$

값	모델 출력	뜻
$p(x)$	0과 1 사이의 확률	범주 1일 추정 확률
$\hat{y}$	0 또는 1	임계값을 적용한 예측 범주

출력에 맞춰 추정 기준도 달라진다

로지스틱 회귀는 관측된 정답 범주에 높은 확률을 주도록 계수를 추정한다

모델	출력	추정 기준
선형 회귀	수치 $\hat{y}$	제곱 손실 최소화
로지스틱 회귀	확률 $p(x)$	우도(likelihood) 최대화

두 회귀 모델은 선형 점수를 쓰지만 출력과 추정 기준이 다르다

Source: James et al. (2023), Section 4.3.2.

Checkpoint

From Values to Probabilities

- 두 회귀 모델이 공통으로 사용하는 선형식은 무엇인가
- sigmoid 함수와 임계값은 서로 어떤 역할을 하는가
- 출력이 달라지면 계수 해석과 추정 기준은 어떻게 달라지는가

Section 5 • Linear Regression Revisited

Why Start Linear?

선형 모델을 비교 기준으로 쓰는 이유

두 모델은 선형 점수를 공유한다

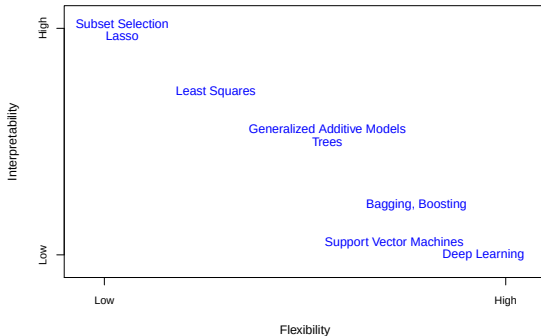
모델	선형 점수 뒤의 출력	손실
선형 회귀	z 를 수치 예측값으로 사용	제곱 손실
로지스틱 회귀	$\sigma(z)$ 를 확률로 사용	음의 로그 우도

우도를 최대화하는 계수는 음의 로그 우도를 최소화하는 계수와 같다

두 모델 모두 설명변수의 가중합인 선형 점수에서 시작한다

Source: MIT 6.390 (2026), Regression and Classification.

최소제곱 선형 회귀의 유연성과 해석 가능성



그림의 유연성(flexibility)은 모델 복잡도를 뜻한다. 최소제곱 선형 회귀는 복잡도가 낮은 대신 해석하기 쉬운 방법에 속한다

Source: James et al. (2023), Figure 2.7.

단순한 모델을 비교 기준으로 쓰는 이유

성질	쓰임
계산이 빠르다	복잡한 모델과 비교할 기준을 빨리 만든다
계수를 해석할 수 있다	설명변수와 예측값 사이의 관계를 확인한다
잔차를 살펴볼 수 있다	잔차의 패턴으로 모델이 놓친 관계를 찾는다

Source: James et al. (2023), Section 2.1.3.

잔차와 계수로 선형식의 한계를 진단한다

자료의 성질	진단 결과	대응
굽은 관계	잔차에 곡선 패턴이 남음	특성 변환
설명변수 사이의 강한 상관	계수 추정이 불안정해짐	해석에 주의
설명변수 사이의 상호작용	잔차에 결합 패턴이 남음	상호작용 항 추가

잔차 패턴과 계수의 불안정성으로 선형식이 놓친 관계를 진단한다

Source: James et al. (2023), Sections 3.3.2 and 3.3.3.

선형식으로 복잡한 관계를 표현하면

실제 관계가 선형식보다 복잡하면 단순한 모델은 과소적합할 수 있다

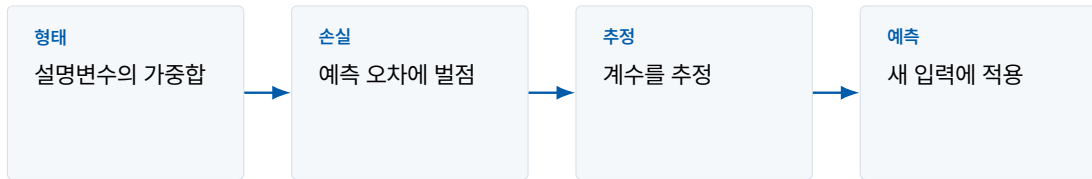
성질	선형 모델	복잡한 모델
편향	커질 수 있다	줄어들 수 있다
분산	작은 경향이 있다	커질 수 있다
학습 자료 적합	표현할 수 있는 관계가 제한된다	학습 오차가 더 작다

모델 선택에는 검증 자료를 쓰고 테스트 자료는 최종 성능 평가에 남긴다

Source: James et al. (2023), Section 2.2.2.

선형 모델을 고르는 네 단계

형태, 손실, 추정, 예측을 서로 다른 결정으로 본다



모델의 형태와 손실을 정한 뒤 학습 자료로 계수를 계산한다

Checkpoint

Why Start Linear?

- 선형 회귀와 로지스틱 회귀가 공유하는 계산은 무엇인가
- 단순한 기준 모델이 비교에 필요한 이유는 무엇인가
- 실제 관계가 선형식보다 복잡하면 어떤 오차가 남을 수 있는가

How Do We Fit a Logistic Model?

Next Lecture: Logistic Regression

- 선형 점수와 확률의 차이
- sigmoid를 이용한 확률 변환
- 확률에 임계값을 적용한 범주 결정