

## 4장 모델 훈련 2부

## 감사의 글

자료를 공개한 저자 오렐리앙 제롱과 강의자료를 지원한 한빛아카데미에게 진심어린 감사를 전합니다.

## 4.5 규제 선형 모델

## 자유도와 규제

- 자유도(degree of freedom): 학습 모델 결정에 영향을 주는 요소(특성)들의 수
  - 단순 선형 회귀의 경우: 특성 수
  - 다항 선형 회귀 경우: 차수
- 규제(regularization): 자유도 제한
  - 단순 선형 회귀 모델에 대한 규제: 가중치 역할 제한
  - 다항 선형 회귀 모델에 대한 규제: 차수 줄이기

## 가중치 역할 규제 선형 회귀 모델

- 릿지 회귀
- 라쏘 회귀
- 엘라스틱넷

## 규제 적용 주의사항

규제항은 훈련 과정에만 사용된다. 테스트 과정에는 다른 기준으로 성능을 평가한다.

- 훈련 과정: 비용 최소화 목표
- 테스트 과정: 최종 목표에 따른 성능 평가
  - 예제: 분류기의 경우 재현율/정밀도 기준으로 성능 평가

릿지 회귀

## 릿지 회귀의 비용함수

$$J(\theta) = \text{MSE}(\theta) + \alpha \frac{1}{2} \sum_{i=1}^n \theta_i^2$$

- $\alpha$ (알파): 규제 강도를 지정하는 하이퍼파라미터
- $\alpha = 0$ : 단순 선형 회귀
- $\alpha$ 가 커질 수록 가중치의 역할이 줄어듦.
  - 비용을 줄이기 위해 가중치를 작게 유지하는 방향으로 학습
- 주의사항: 훈련 세트에 대한 특성 스케일링 전처리 실행 후 적용



라쏘 회귀

## 라쏘 회귀의 비용함수

$$J(\theta) = \text{MSE}(\theta) + \alpha \sum_{i=1}^n |\theta_i|$$

- $\alpha$ (알파)
  - 하이퍼파라미터로 지정됨.
  - 규제 강도 지정
  - $\alpha = 0$ 이면 규제가 전혀 없는 기본 선형 회귀
- $\theta_i$ : 덜 중요한 특성을 무시하기 위해  $|\theta_i|$ 가 0에 수렴하도록 학습 유도.
- 주의:  $\theta_0$ 은 규제하지 않음

## 엘라스틱넷

- 릿지 회귀와 라쏘 회귀를 절충한 모델

엘라스틱넷의 비용함수

$$J(\theta) = \text{MSE}(\theta) + r \alpha \sum_{i=1}^n |\theta_i| + \frac{1-r}{2} \alpha \sum_{i=1}^n \theta_i^2$$

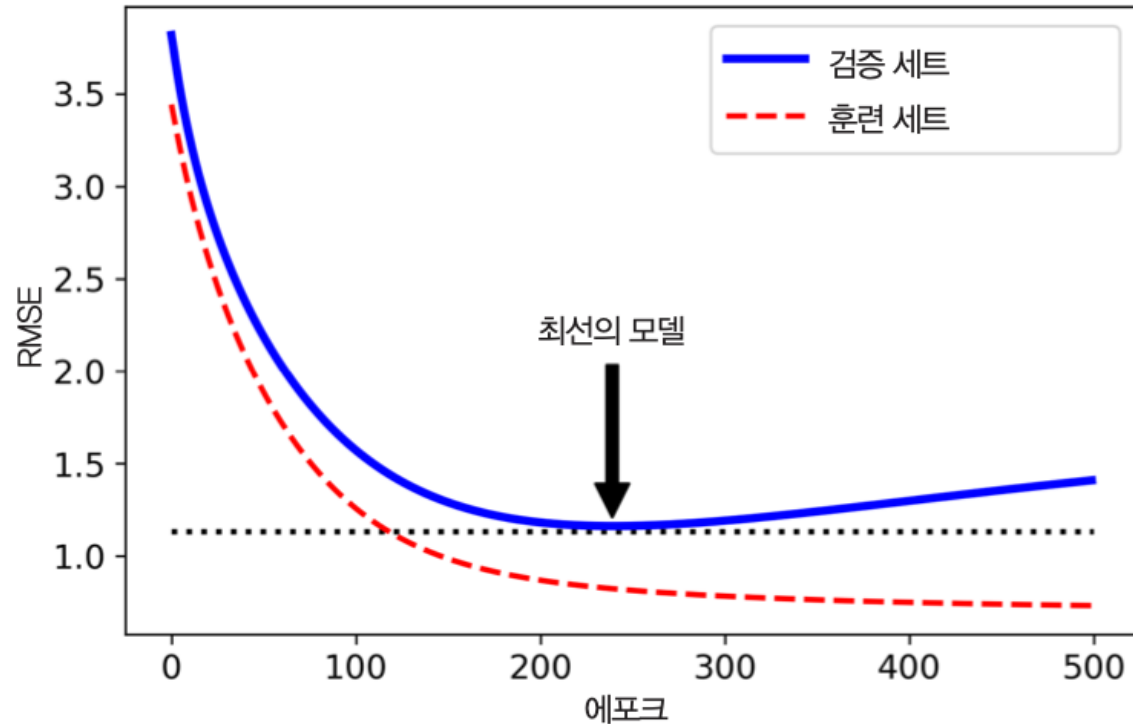
- $r$ 을 이용하여 릿지 규제와 라쏘 규제를 적절하게 조절

## 규제 사용 방법

- 대부분의 경우 약간이라도 규제 사용 추천
- 릿지 규제가 기본
- 유용한 속성이 많지 않다고 판단되는 경우
  - 라쏘 규제나 엘라스틱넷 활용 추천
  - 불필요한 속성의 가중치를 0으로 만들기 때문
- 특성 수가 훈련 샘플 수보다 크거나 특성 몇 개가 강하게 연관되어 있는 경우
  - 라쏘 규제는 적절치 않음.
  - 엘라스틱넷 추천

## 조기 종료 기법

- 반복 훈련 과정 중에 모델이 훈련 데이터에 점점 더 익숙해져서 과대적합 발생 가능
- 따라서 반복 훈련을 적절한 시기에 종료해야 함
- 반복훈련 종료 기준: 검증 데이터에 대한 손실이 줄어 들다가 다시 커지는 순간
- 조기 종료: 검증 오차가 최소에 다다랐을 때 반복 훈련을 멈추게 하는 기법



- 확률적 경사 하강법, 미니배치 경사 하강법의 경우 손실 곡선이 매끄럽지 않고 진동 발생 가능
- 이런 경우에는 검증 오차가 한동안 최솟값보다 높게 유지될 때 반복 훈련을 멈추고 검증 오차가 최소였을 때의 모델 파라미터 확인

## 4.6 로지스틱 회귀와 소프트맥스 회귀



로지스틱 회귀와 소프트맥스 회귀를 이용하여 분류 모델 학습 가능

- 이진 분류: 로지스틱 회귀 활용
- 다중 클래스 분류: 소프트맥스 회귀 활용

## 로지스틱 회귀와 시그모이드 함수

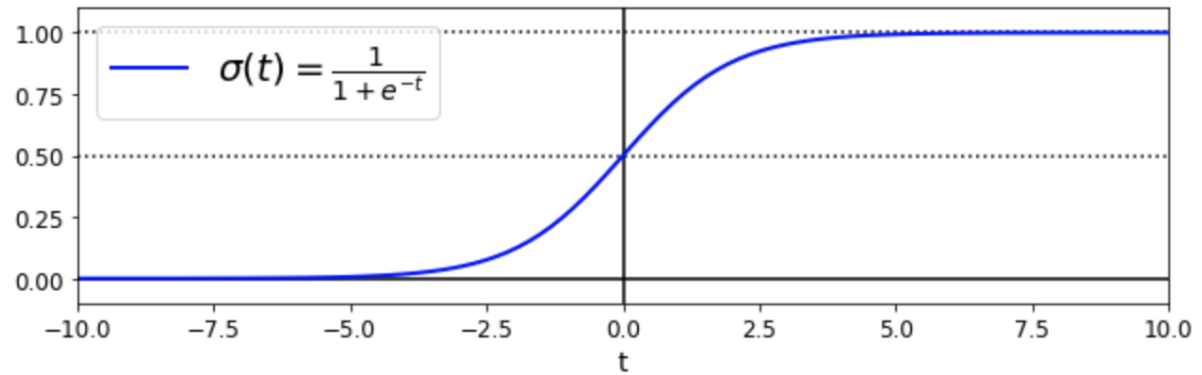
### 로지스틱 회귀

- 특성과 가중치의 곱한 값들을 더한 결과에 **시그모이드 함수**를 적용한 결과 이용
- 로지스틱 회귀 모델에서 샘플  $\mathbf{x}$ 에 대한 예측값

$$\hat{p} = h_{\theta}(\mathbf{x}) = \sigma(\theta^T \mathbf{x}_b^T)$$

## 시그모이드 함수

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



## 로지스틱 회귀 모델의 예측값

$$\hat{y} = \begin{cases} 0 & \text{if } \hat{p} < 0.5 \\ 1 & \text{if } \hat{p} \geq 0.5 \end{cases}$$

- $\theta^T \mathbf{x}_b^T \geq 0$  인 경우: 양성 클래스(1)
- $\theta^T \mathbf{x}_b^T < 0$  인 경우: 음성 클래스(0)

## 로지스틱 회귀 모델의 비용함수

- 로지스틱 회귀 모델을 경사하강법을 이용하여 학습
- 비용함수: 로그 손실(log loss) 함수 사용

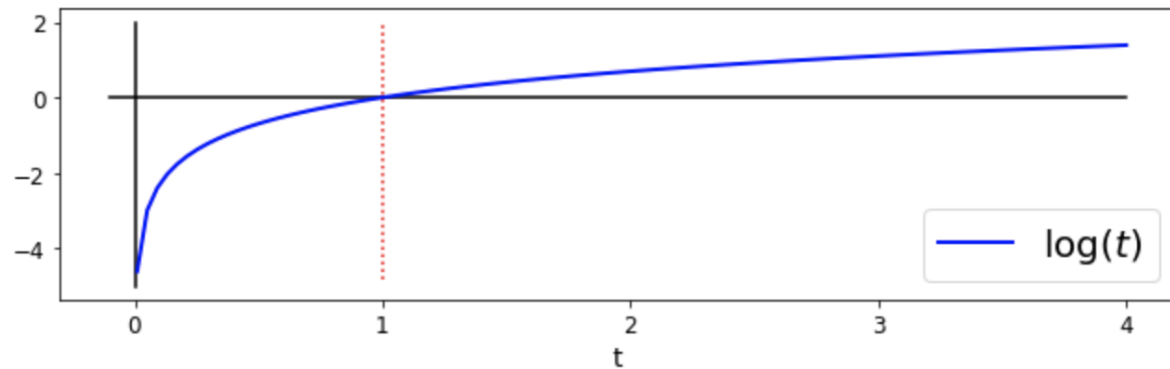
$$J(\theta) = -\frac{1}{m} \sum_{i=0}^{m-1} [y^{(i)} \log(\hat{p}^{(i)}) + (1 - y^{(i)}) \log(1 - \hat{p}^{(i)})]$$

- 이 비용함수에 대해 경사 하강법 적용

- 로그 손실 함수 이해: 하나의 샘플에 대한 아래의 값의 의미 이해 중요

$$-[y^{(i)} \log(\hat{p}^{(i)}) + (1 - y^{(i)}) \log(1 - \hat{p}^{(i)})]$$

- 틀린 예측을 하면 값이 커짐.
- log 함수 성질 참조



결정 경계: 로지스틱 회귀 활용 예제

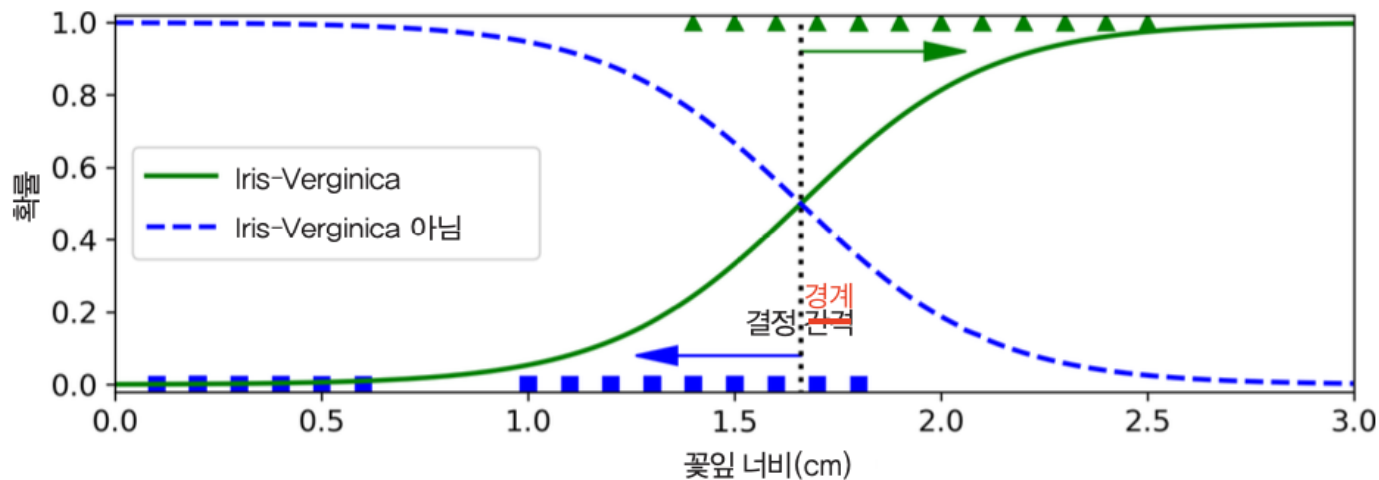
## 사이킷런에서 제공하는 붓꽃 데이터셋 활용

- 4개의 특성 사용
  - 꽃받침 길이
  - 꽃받침 너비
  - 꽃잎 길이
  - 꽃잎 너비
- 샘플 타깃
  - 0: Iris-Setosa
  - 1: Iris-Versicolor
  - 2: Iris-Virginica



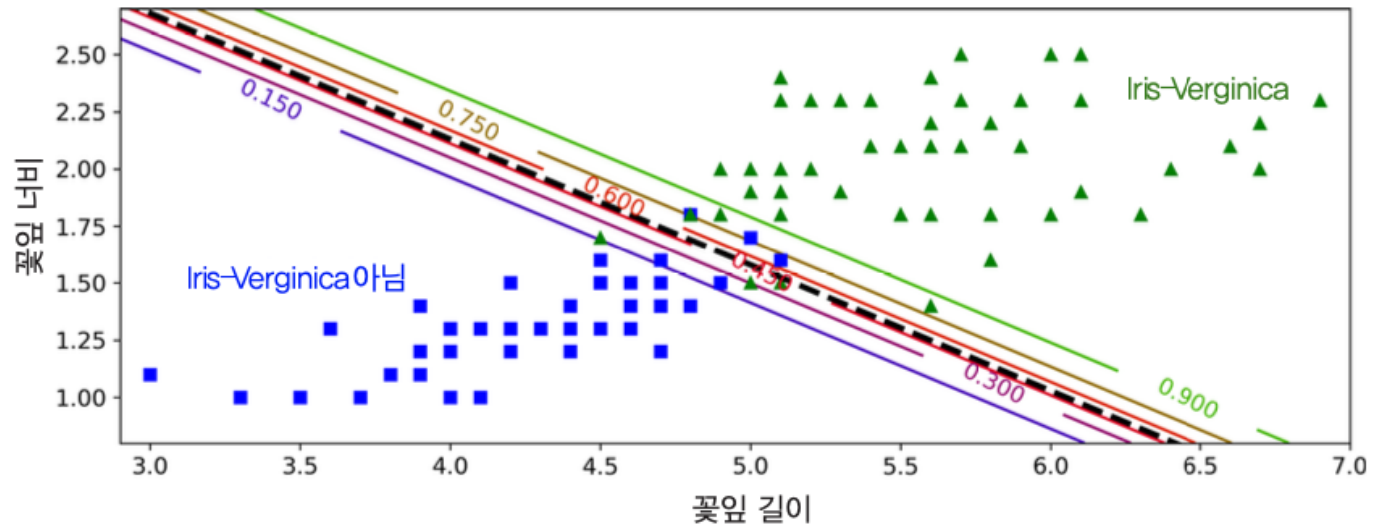
## 꽃잎의 너비를 기준으로 Iris-Virginica 여부 판정하기

- 결정경계: 약 1.6cm



## 꽃잎의 너비와 길이를 기준으로 Iris-Virginica 여부 판정하기

- 결정경계: 검정 점선



## 로지스틱 회귀 규제하기

- 하이퍼파라미터 `penalty` 와 `C` 이용
- `penalty`
  - `l1`, `l2`, `elasticnet` 세 개중에 하나 사용.
  - 기본은 `l2`, 즉,  $\ell_2$  규제를 사용하는 릿지 규제.
  - `elasticnet` 을 선택한 경우 `l1_ratio` 옵션 값을 지정해서 함께 사용.
- `C`
  - 릿지 규제 정도를 지정하는  $\alpha$ 의 역수에 해당.
  - 따라서 0에 가까울 수록 강한 규제 의미.

## 소프트맥스(softmax) 회귀

- 로지스틱 회귀 모델을 일반화하여 다중 클래스 분류를 지원하도록 한 회귀 모델
- 다항 로지스틱 회귀 라고도 불림

## 소프트맥스 회귀 학습 아이디어

- 샘플  $\mathbf{x}$ 이 주어졌을 때 각각의 분류 클래스  $k$ 에 대한 점수  $s_k(\mathbf{x})$  계산

$$s_k(\mathbf{x}) = (\theta^{(k)})^T \mathbf{x}_b^T$$

- 소프트맥스 함수를 이용하여 각 클래스  $k$ 에 속할 확률  $\hat{p}_k$  계산

$$\hat{p}_k = \frac{\exp(s_k(\mathbf{x}))}{\sum_{j=0}^{K-1} \exp(s_j(\mathbf{x}))}$$

- 추정 확률이 가장 높은 클래스 선택

$$\hat{y} = \operatorname{argmax}_k s_k(\mathbf{x})$$

## 주의사항

- 소프트맥스 회귀는 다중 출력 분류 지원 못함.
- 예를 들어, 하나의 사진에서 여러 사람의 얼굴 인식 불가능.

## 소프트맥스 회귀 비용함수

- 각 분류 클래스  $k$ 에 대한 적절한 가중치 벡터  $\theta_k$ 를 학습해 나가야 함.
- 비용함수: 크로스 엔트로피 비용 함수 사용

$$J(\Theta) = -\frac{1}{m} \sum_{i=0}^{m-1} \sum_{k=0}^{K-1} y_k^{(i)} \log(\hat{p}_k^{(i)})$$

- 이 비용함수에 대해 경사 하강법 적용

## 참조

- $K = 2$ 이면 로지스틱 회귀의 로그 손실 함수와 정확하게 일치한다.
- 주어진 샘플의 타깃 클래스를 제대로 예측할 경우 높은 확률값 계산
- 크로스 엔트로피 개념은 정보 이론에서 유래하였다. (자세한 설명은 생략)

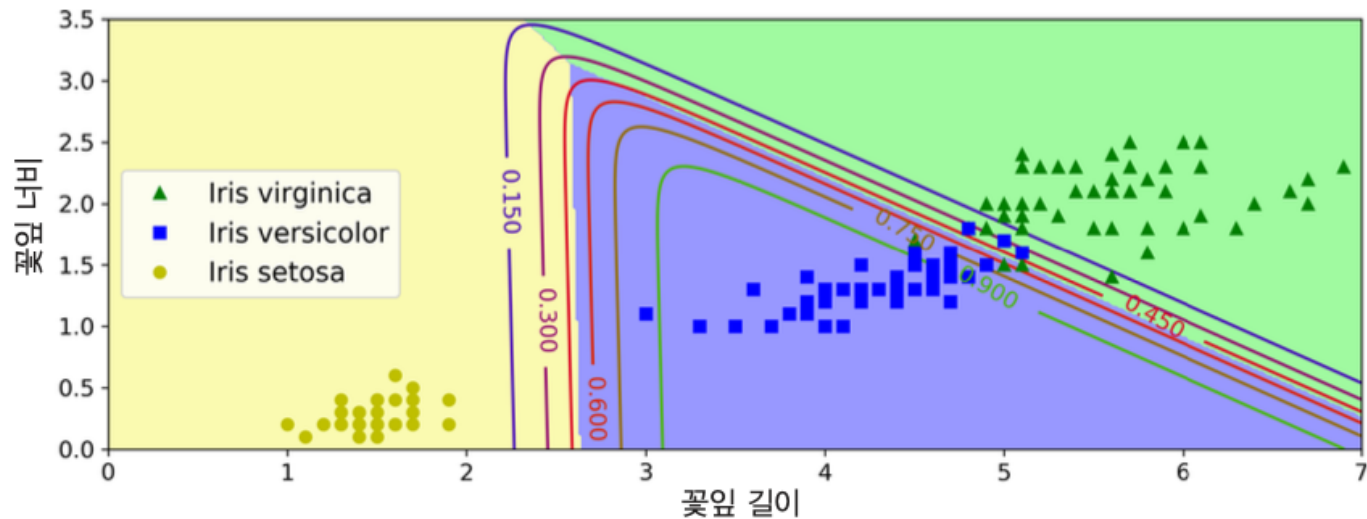


## 멀티 클래스 분류: 소프트맥스 회귀 활용 예제

- 붓꽃을 세 개의 클래스로 분류하기
- 사이킷런의 `LogisticRegression` 예측기 활용
  - `multi_class` 하이퍼파라미터 값을 `multinomial` 로 지정

## 꽃잎의 너비와 길이를 기준으로 붓꽃 클래스 분류

- 결정경계: 배경색으로 구분
- 곡선: Iris-Versicolor 클래스에 속할 확률



- 예제: 꽃잎 길이 5cm, 너비 2cm 인 붓꽃에 대한 품종 클래스 추정
  - 94.2%의 확률로 Iris-Virginica
  - 또는 5.8%의 확률로 Iris-Versicolor