선형모델 예제

23.03.2019

Presenter: **KyungTae Lim**

Contents

본 자료는 기계학습 **(오일석)** 의 슬라이드의 일부를 재구성 한 후 이와 연관된 Pytorch 예제 코드를 작성했습니다.

슬라이드 출처 http://cv.jbnu.ac.kr/index.php?mid=ml

선형모델의 예제

선형 회귀 문제

- [그림 1-4]: 식 (1.2)의 직선 모델을 사용하므로 두 개의 매개변수 $\Theta = (w, b)^{\mathrm{T}}$

$$y = wx + b \tag{1.2}$$

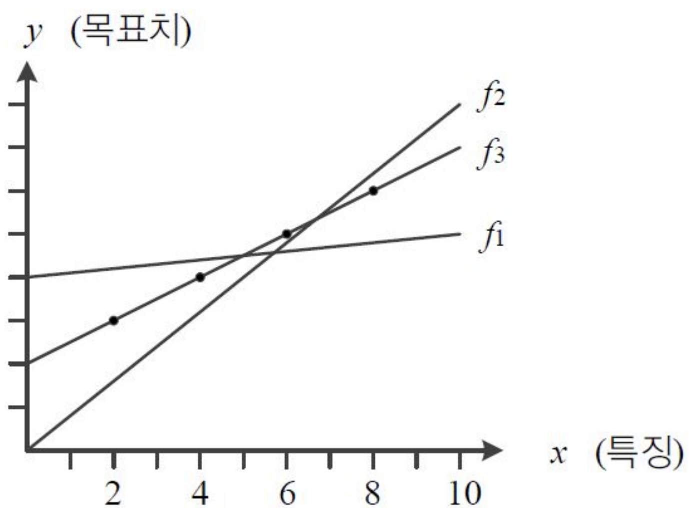


그림 1-4 간단한 기계 학습 예제

선형모델의 목적함수

- 목적 함수objective function (또는 비용 함수cost function)
  - 식 (1.8)은 선형 회귀를 위한 목적 함수
    - $f_\Theta(\mathbf{x}_i)$는 예측함수의 출력, y_i는 예측함수가 맞추어야 하는 목푯값이므로 $f_\Theta(\mathbf{x}_i)-y_i$는 오차
    - 식 (1.8)을 평균제곱오차MSE(mean squared error)라 부름

$$J(\Theta) = \frac{1}{n}\sum_{i=1}^{n}(f_\Theta(\mathbf{x}_i) - y_i)^2 \qquad (1.8)$$

  - 처음에는 최적 매개변수 값을 알 수 없으므로 난수로 $\Theta_1 = (w_1, b_1)^{\mathrm{T}}$ 설정 → $\Theta_2 = (w_2, b_2)^{\mathrm{T}}$로 개선 → $\Theta_3 = (w_3, b_3)^{\mathrm{T}}$로 개선 → Θ_3는 최적해 $\hat{\Theta}$
    - 이때 $J(\Theta_1) > J(\Theta_2) > J(\Theta_3)$

출처: 기계학습 (오원철

선형모델의 예제

[예제 1-1]

- 훈련집합

  $\mathbb{X} = \{x_1 = (2.0), x_2 = (4.0),\ x_3 = (6.0),\ x_4 = (8.0)\}$,

  $\mathbb{Y} = \{y_1 = 3.0,\ y_2 = 4.0,\ y_3 = 5.0,\ y_4 = 6.0\}$

- 초기 직선의 매개변수 $\Theta_1 = (0.1, 4.0)^{\mathrm{T}}$라 가정

$$
\begin{aligned}
\mathbf{x}_1, y_1 &\longrightarrow \left(f_{\Theta_1}(2.0) - 3.0\right)^2 = \left((0.1 * 2.0 + 4.0) - 3.0\right)^2 = 1.44 \\
\mathbf{x}_2, y_2 &\longrightarrow \left(f_{\Theta_1}(4.0) - 4.0\right)^2 = \left((0.1 * 4.0 + 4.0) - 4.0\right)^2 = 0.16 \\
\mathbf{x}_3, y_3 &\longrightarrow \left(f_{\Theta_1}(6.0) - 5.0\right)^2 = \left((0.1 * 6.0 + 4.0) - 5.0\right)^2 = 0.16 \\
\mathbf{x}_4, y_4 &\longrightarrow \left(f_{\Theta_1}(8.0) - 6.0\right)^2 = \left((0.1 * 8.0 + 4.0) - 6.0\right)^2 = 1.44
\end{aligned}
\longrightarrow \quad J(\Theta_1) = 0.8
$$

출처: 기계학습 (오원철

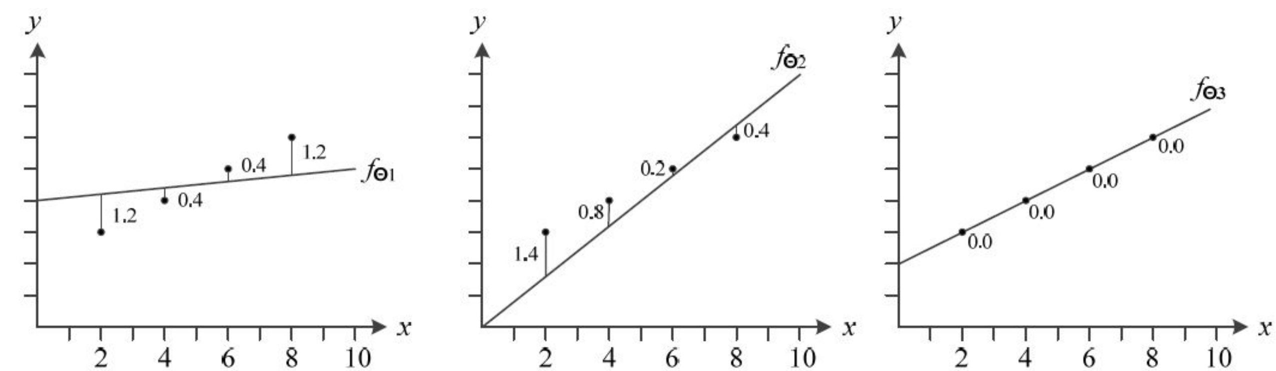
선형모델의 예제

[예제 1-1] 훈련집합

- Θ_1을 개선하여 $\Theta_2 = (0.8, 0.0)^{\mathrm{T}}$가 되었다고 가정

$$
\begin{aligned}
\mathbf{x}_1, \mathrm{y}_1 &\rightarrow \left(f_{\Theta_2}(2.0) - 3.0\right)^2 = \left((0.8 * 2.0 + 0.0) - 3.0\right)^2 = 1.96 \\
\mathbf{x}_2, \mathrm{y}_2 &\rightarrow \left(f_{\Theta_2}(4.0) - 4.0\right)^2 = \left((0.8 * 4.0 + 0.0) - 4.0\right)^2 = 0.64 \\
\mathbf{x}_3, \mathrm{y}_3 &\rightarrow \left(f_{\Theta_2}(6.0) - 5.0\right)^2 = \left((0.8 * 6.0 + 0.0) - 5.0\right)^2 = 0.04 \\
\mathbf{x}_4, \mathrm{y}_4 &\rightarrow \left(f_{\Theta_2}(8.0) - 6.0\right)^2 = \left((0.8 * 8.0 + 0.0) - 6.0\right)^2 = 0.16
\end{aligned}
\quad \longrightarrow \quad J(\Theta_2) = 0.7
$$

- Θ_2를 개선하여 $\Theta_3 = (0.5, 2.0)^{\mathrm{T}}$가 되었다고 가정
- 이때 $J(\Theta_3) = 0.0$이 되어 Θ_3은 최적값 $\hat{\Theta}$ 이 됨



(a) 초기 매개변수 Θ_1 (b) Θ_1을 개선하여 Θ_2가 됨 (c) Θ_2를 개선하여 최적의 Θ_3을 찾음

그림 1-11 기계 학습에서 목적함수의 역할

출처: 기계학습 (오원철

선형모델의 일반화

- 기계 학습이 할 일을 공식화하면,

$$\widehat{\Theta} = \underset{\Theta}{\operatorname{argmin}} J(\Theta) \qquad (1.9)$$

  - 기계 학습은 작은 개선을 반복하여 최적해를 찾아가는 수치적 방법으로 식 (1.9)를 풂

- 알고리즘 형식으로 쓰면,

알고리즘 1-1 기계 학습 알고리즘

입력 : 훈련집합 $\mathbb{X}$와 $\mathbb{Y}$

출력 : 최적의 매개변수 $\widehat{\Theta}$

난수를 생성하여 초기 해 Θ_1을 설정한다.
t=1
while ($J(\Theta_t)$가 0.0에 충분히 가깝지 않음) // 수렴 여부 검사
    $J(\Theta_t)$가 작아지는 방향 $\Delta\Theta_t$를 구한다. // $\Delta\Theta_t$는 주로 미분을 사용하여 구함
    $\Theta_{t+1} = \Theta_t + \Delta\Theta_t$
    t=t+1
$\widehat{\Theta} = \Theta_t$

출처: 기계학습 (오원철

선형모델의 한계

- 좀더 현실적인 상황
  - 지금까지는 데이터가 선형을 이루는 아주 단순한 상황을 고려함
  - 실제 세계는 선형이 아니며 잡음이 섞임 □ 비선형 모델이 필요

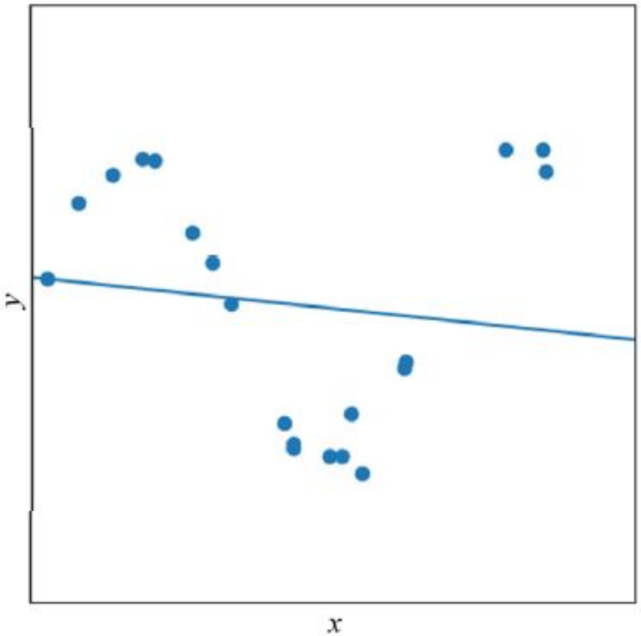


그림 1-12 선형 모델의 한계

출처: 기계학습 (오원철

직접 만들어 봅시다.