모델 훈련 1

By Hands-On

경제학과 하지민

선형 회귀

$$y = Wx + b$$

$$y = Wx$$

$$cost = \frac{1}{m}\sum_{i=1}^{m}\left(\boxed{H\left(x^{(i)}\right)} - y^{(i)}\right)^2$$

선형 회귀

정규 방정식

∑ 로 표현되는 제곱의 합은 그 수들을 요소로 하는 행렬과 그 행렬의 전치행렬의 곱과 같음

$$1 \quad 2 \quad 3 \times 1 \quad 2 \quad 3^{T} = 1 \quad 2 \quad 3 \times \begin{matrix} 1 \\ 2 \\ 3 \end{matrix}$$

$$= 1 \times 1 + 2 \times 2 + 3 \times 3$$

$$= \sum_{i=1}^{3} i^{2}$$

선형 회귀

정규 방정식

정치행렬은 다음과 같은 성질 존재

$$(A^T)^T = A$$

$$(A+B)^T = A^T + B^T$$

$$(AB)^T = B^T A^T$$

$$(kA)^T = kA^T \text{(}k\text{는 임의의 상수)}$$

$$A^T B = B^T A$$

선형 회귀

정규 방정식

$$\mathrm{MSE}(W) = \frac{1}{m}\sum_{i=1}^{m}(Wx^{(i)} - y^{(i)})^2$$

$$= \frac{1}{m}((WX - y)^T(WX - y))$$

$$= \frac{1}{m}((WX)^T - y^T)\,(WX - y))$$

$$= \frac{1}{m}((WX)^T WX - (WX)^T y - y^T WX + y^T y)$$

$$= \frac{1}{m}(X^T W^T WX - 2(WX)^T y + y^T y)$$

선형 회귀

정규 방정식

$$MSE(W) = \frac{1}{m}(X^T \boxed{W^T W} X - 2\boxed{(WX)^T} y + y^T y)$$

$$MSE(W) = \frac{1}{m}(X^T X\boxed{W^2} - 2\boxed{X^T} y\boxed{W^T} + y^T y)$$

$$\frac{dMSE(W)}{dW} = \frac{1}{m}(2X^T XW - 2X^T y) = 0$$

$$\frac{dMSE(W)}{dW} = 2X^T XW - 2X^T y = 0$$

$$2X^T XW - 2X^T y = 0$$

$$2X^T XW = 2X^T y$$

$$X^T XW = X^T y$$

$$W = (X^T X)^{-1} X^T y$$

선형 회귀

정규 방정식 특징

행렬식으로 경사 하강법에 비해 많은 연산량이 필요하지도 않고 학습률 설정 등 골치 아픈 하이퍼파라미터 신경을 쓰지 않아도 됨

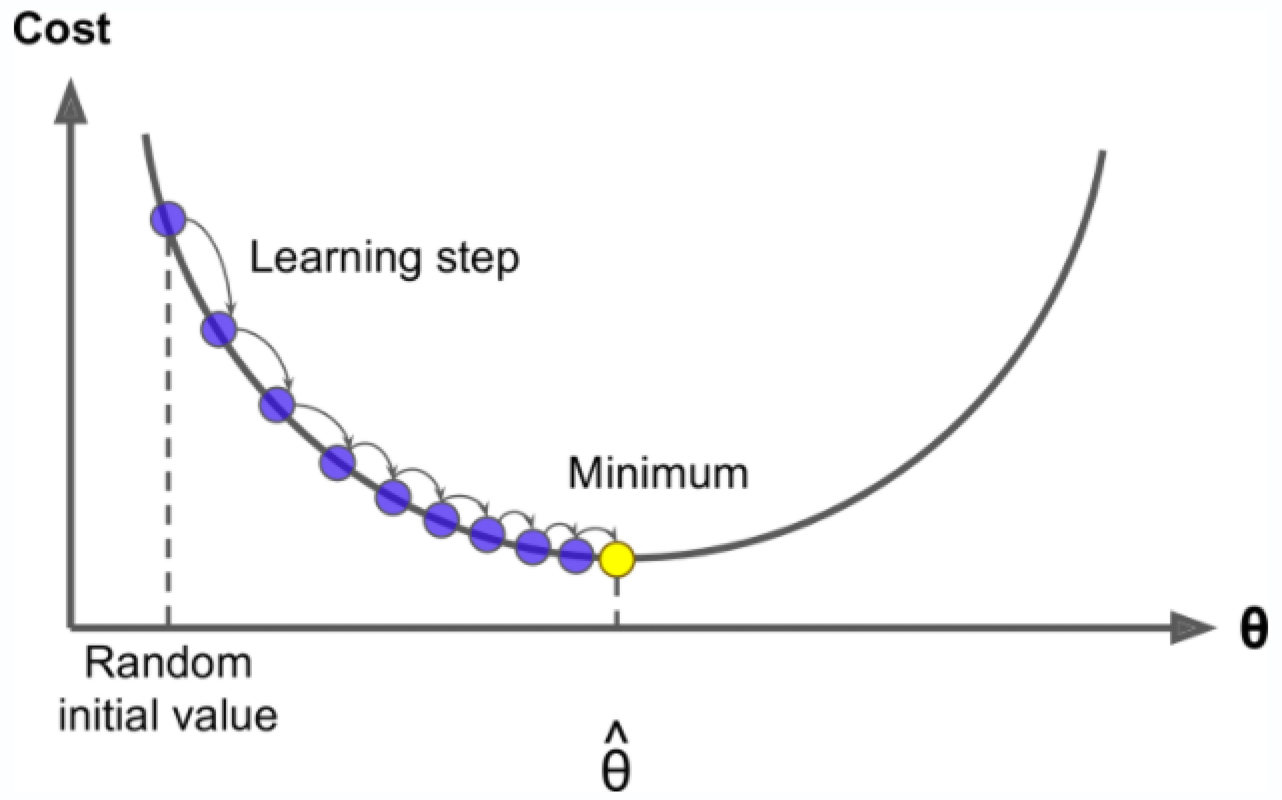
행렬 연상이라서 특성의 수가 늘어나면 계산속도가 많이 느려지게 됨
샘플 수에 대해서는 선형적으로 비례

정규방정식으로 학습된 선형 모델은 예측이 매우 빠름

특성이 매우 많고 훈련 샘플이 너무 많아 메모리에 모두 담을 수 없을 때 적합

경사 하강법

경사 하강법

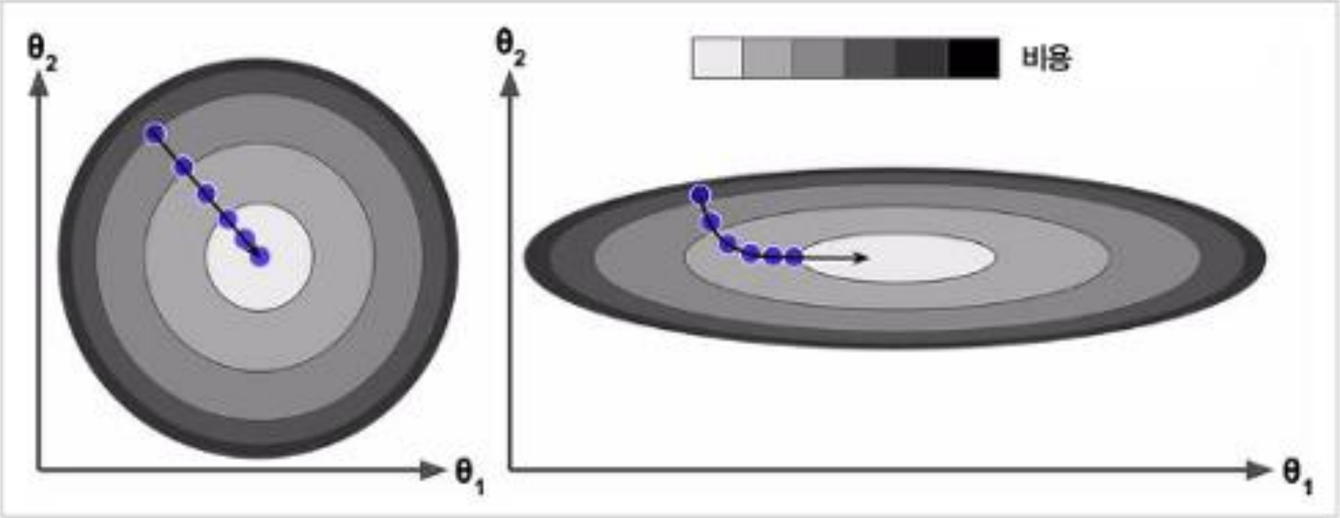


Cost
Learning step
Minimum
θ
Random
initial value
$\hat{\theta}$

경사 하강법

경사 하강법

그림 4-7 특성 스케일에 따른 경사 하강법



경사 하강법

경사 하강법 – step 공식

1. $h_\theta(x) = \theta \cdot x$
2. $h_\theta(\mathrm{X}) = \theta^T \cdot \mathrm{X}$
3. $h_\theta(\mathrm{X}) = \mathrm{X} \cdot \theta$

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 \ldots + \theta_n x_n$$

$$\theta = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ . \\ . \\ . \\ \theta_n \end{pmatrix} \qquad X = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ . \\ . \\ . \\ x_n \end{pmatrix}$$

경사 하강법

경사 하강법 – step 공식

$\boldsymbol{\theta}$와 x 가 모두 n X 1벡터라고 한다면 (n X 1) · (n X 1)이 되어 벡터(행렬)의 연산 법칙으로 인해 계산을 할 수 없게 됨

따라서 앞에 있는 $\boldsymbol{\theta}$를 전치행렬로 만들어 (1 X n) · (n X 1)이 되게 함으로써 연산이 가능하게 만드는 것

이 것이 바로 두 번째 식

물론 전치행렬의 성질에 따라 다음과 같이 표현할 수도 있음

$$h_\theta(\mathrm{X}) = X^T \cdot \theta$$

경사 하강법

경사 하강법 – step 공식

3번째 식은 2번째 식을 조금 더 확장한 것

2번이 식에서 X는 n개의 요소를 갖는 벡터

이러한 식이 m개, 즉 n개의 특성을 갖는 샘플이 m개가 있다고 보는 것

따라서 이 때는 식의 결과 역시 벡터가 되는 것

즉, 3번의 식을 구성하는 각 요소는 다음의 의미가 존재

$$h_\theta(\mathrm{X}) = \begin{pmatrix} h_\theta(x^1) \\ h_\theta(x^2) \\ h_\theta(x^3) \\ . \\ . \\ . \\ h_\theta(x^m) \end{pmatrix} \qquad X = \begin{pmatrix} 1\ x_1^1\ x_2^1\ \ldots x_n^1 \\ 1\ x_1^2\ x_2^2\ \ldots x_n^2 \\ 1\ x_1^3\ x_2^3\ \ldots x_n^3 \\ .\ \ .\ \ .\ \ . \\ .\ \ .\ \ .\ \ . \\ .\ \ .\ \ .\ \ . \\ 1\ x_1^m x_2^m \ldots x_n^m \end{pmatrix} \qquad \theta = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ . \\ . \\ . \\ \theta_n \end{pmatrix} \qquad MSE(\theta) = \frac{1}{m}\sum_{i=1}^{m}(\theta^T \cdot X^{(i)} - y^{(i)})^2$$

경사 하강법

배치 경사 하강법

$$\frac{\partial}{\partial\theta}MSE(\theta) = \frac{2}{m}\sum_{i=1}^{m}(\theta^T \cdot X^{(i)} - y^{(i)}) \cdot X^{(i)}$$

$$\frac{2}{m}((\theta^T \cdot X^{(1)} - y^{(1)}) \cdot X^{(1)} + (\theta^T \cdot X^{(2)} - y^{(2)}) \cdot X^{(2)} + \cdots + (\theta^T \cdot X^{(m)} - y^{(m)}) \cdot X^{(m)})$$

경사 하강법

배치 경사 하강법

$$\begin{pmatrix} (\theta^T \cdot X^{(1)} - y^{(1)}) \\ (\theta^T \cdot X^{(2)} - y^{(2)}) \\ (\theta^T \cdot X^{(3)} - y^{(3)}) \\ \cdot \\ \cdot \\ \cdot \\ (\theta^T \cdot X^{(m)} - y^{(m)}) \end{pmatrix}$$

$$\begin{pmatrix} X^{(1)} \\ X^{(2)} \\ X^{(3)} \\ \cdot \\ \cdot \\ \cdot \\ X^{(m)} \end{pmatrix}$$

$$\begin{pmatrix} X^{(1)} \\ X^{(2)} \\ X^{(3)} \\ \cdot \\ \cdot \\ \cdot \\ X^{(m)} \end{pmatrix}^T = (X^{(1)}\ X^{(2)}\ X^{(3)} \ldots X^{(m)})$$

$$\begin{pmatrix} X^{(1)} \\ X^{(2)} \\ X^{(3)} \\ \cdot \\ \cdot \\ \cdot \\ X^{(m)} \end{pmatrix}^T \cdot \begin{pmatrix} (\theta^T \cdot X^{(1)} - y^{(1)}) \\ (\theta^T \cdot X^{(2)} - y^{(2)}) \\ (\theta^T \cdot X^{(3)} - y^{(3)}) \\ \cdot \\ \cdot \\ \cdot \\ (\theta^T \cdot X^{(m)} - y^{(m)}) \end{pmatrix}$$

경사 하강법

배치 경사 하강법

$$\begin{pmatrix} (\theta^T \cdot X^{(1)} - y^{(1)}) \\ (\theta^T \cdot X^{(2)} - y^{(2)}) \\ (\theta^T \cdot X^{(3)} - y^{(3)}) \\ \vdots \\ (\theta^T \cdot X^{(m)} - y^{(m)}) \end{pmatrix} = \begin{pmatrix} \theta^T \cdot X^{(1)} \\ \theta^T \cdot X^{(2)} \\ \theta^T \cdot X^{(3)} \\ \vdots \\ \theta^T \cdot X^{(m)} \end{pmatrix} - \begin{pmatrix} y^{(1)} \\ y^{(2)} \\ y^{(3)} \\ \vdots \\ y^{(m)} \end{pmatrix}$$

$$\begin{pmatrix} \theta^T \cdot X^{(1)} \\ \theta^T \cdot X^{(2)} \\ \theta^T \cdot X^{(3)} \\ \vdots \\ \theta^T \cdot X^{(m)} \end{pmatrix} = \theta^T \cdot \begin{pmatrix} X_0^{(1)} X_1^{(1)} X_2^{(1)} \dots X_n^{(1)} \\ X_0^{(2)} X_1^{(2)} X_2^{(2)} \dots X_n^{(2)} \\ X_0^{(3)} X_1^{(3)} X_2^{(3)} \dots X_n^{(3)} \\ \vdots \\ X_0^{(m)} X_1^{(m)} X_2^{(m)} \dots X_n^{(m)} \end{pmatrix}$$

$$\frac{2}{m} X^T \cdot (X \cdot \theta - y)$$

$$\theta_{i+1} = \theta_i - \eta \frac{2}{m} X^T \cdot (X \cdot \theta - y)$$

경사 하강법 종류

	경사하강법	확률적 경사하강법
1회의 학습에 사용되는 데이터	모든 데이터 사용	랜덤으로 추출된 1개의 데이터 사용(중복 선택 가능)
반복에 따른 정확도	학습이 반복 될 수록 최적해에 근접	학습이 반복 될 수록 최적해에 근접
노이즈	거의 없음	비교적 노이즈가 심함
해를 찾는 과정의 이미지 비교