Chapter 3 : Single-Variate Linear Regression without Bias Term.

1-4. Loss and Cost Function 2.

Loss : $\mathcal{L}^{(i)}(\underline{\theta}) = (y^{(i)} - (\theta x^{(i)}))^2$

Weight of Feature values in vector

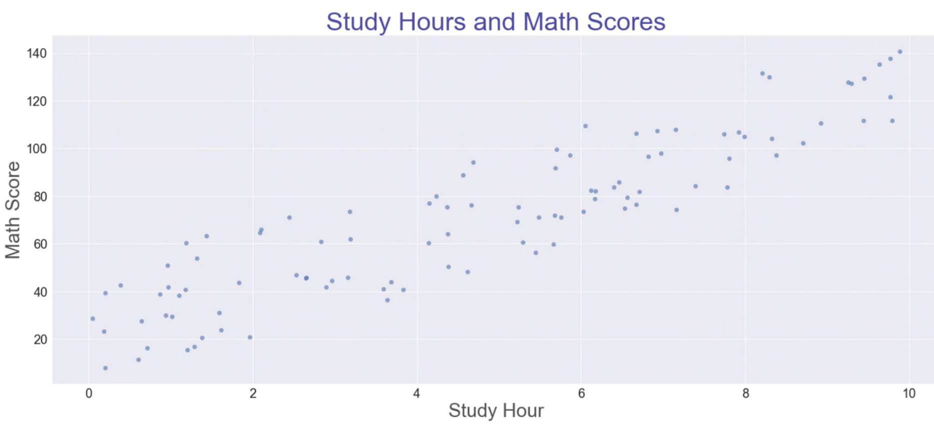
Dot product

Degree of $\theta = 2$

→ 변수 θ에 대한 2차 함수

Cost : $J(\theta) = \frac{1}{n}\sum_{i=1}^{n}(y^{(i)} - (\theta x^{(i)}))^2$ 역시 Degree of $\theta = 2$

→ 실제값과 예측값 사이 차이들의 제곱의 평균.



$$\hat{y} = \theta_1 x + \theta_0$$

Bias Term

선형관계에서 Bias Term이 추가됨.

위에서

Model : $\hat{y} = \theta_1 x + \theta_0$

Loss : $\mathcal{L}^{(i)} = (y^{(i)} - \hat{y}^{(i)})^2$

Cost : $J = \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}^{(i)}$ 여전히 같다.

<Model $\hat{y} = \theta_1 x + \theta_0$ 에 대한 Loss 는>

$\hat{y} = \theta_1 x + \theta_0$

$$\mathcal{L}^{(i)} = (y^{(i)} - \hat{y}^{(i)})^2$$

변수

$$\mathcal{L}^{(i)}(\theta_1, \theta_0) = (\underbrace{y^{(i)}}_{\text{Constant}} - (\theta_1 \underbrace{x^{(i)}}_{\text{Constant}} + \theta_0))^2$$

이때, $\vec{\theta} = \begin{pmatrix} \theta_1 \\ \theta_0 \end{pmatrix}$ $\vec{x}^{(i)} = \begin{pmatrix} x^{(i)} \\ 1 \end{pmatrix}$ 이라고 하면

$$\mathcal{L}^{(i)}(\theta_1, \theta_0) = (y^{(i)} - (\vec{\theta}^T \cdot \vec{x}^{(i)}))^2$$ 가 된다

$$[\theta_1 \ \ \theta_0] \cdot \begin{bmatrix} x^{(i)} \\ 1 \end{bmatrix} = \theta_1 x^{(i)} + \theta_0$$

< Model $\hat{y} = \theta_1 x + \theta_0$에 대한 Cost는 >

$$J = \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}^{(i)}$$

변수

$$J(\theta_1, \theta_0) = \frac{1}{n}\sum_{i=1}^{n}\left(y^{(i)} - (\theta_1 x^{(i)} + \theta_0)\right)^2$$

위와 같은 이유로,

$$J(\theta_1, \theta_0) = \frac{1}{n}\sum_{i=1}^{n}\left(y^{(i)} - (\vec{\theta}^{T} \vec{x}^{(i)})\right)^2$$

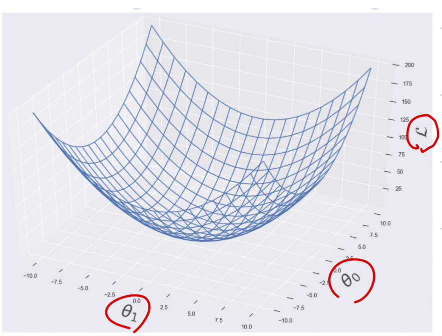
위에서 구한 식 $\mathcal{L}^{(i)}(\theta_1, \theta_0) = \left(y^{(i)} - (\theta_1 x^{(i)} + \theta_0)\right)^2$

$$J(\theta_1, \theta_0) = \frac{1}{n}\sum_{i=1}^{n}\left(y^{(i)} - (\theta_1 x^{(i)} + \theta_0)\right)^2$$

둘 다 각각 θ_1 와 θ_0에 대해서 2차 함수 임

그리고 θ_1^2 와 θ_0^2의 계수 모두 양수 이므로 아래 볼록 $\cup$.

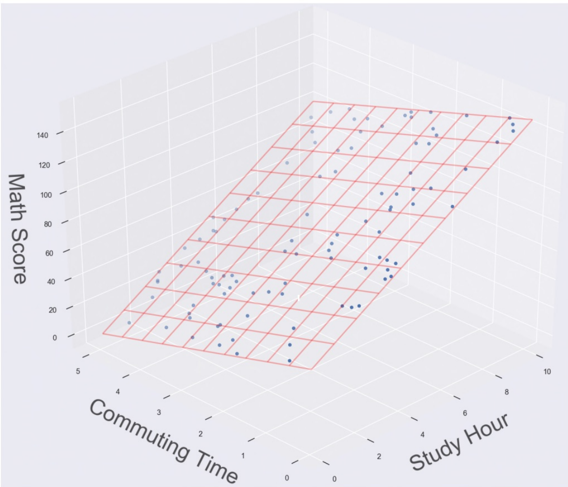
Cost function을 3D로 보면...



gradient로 가장 아래 점으로 가야 함.

위는 하나의 factor, <u>두개인 경우</u> ex) 수학점수와 Commuting time & Study hour 간의 관계를 선형적으로 보는 것.

< feature 두개 일때 그래프 >



선형 Model : $y = \theta_2 x_2 + \theta_1 x_1 + \theta_0$

$$\mathcal{L}^{(i)} = \left(y^{(i)} - \hat{y}^{(i)}\right)^2$$

$$J = \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}^{(i)}$$

한개 feature 일때와 같다.

< feature가 두개 일 때 Loss와 Cost는 ... >

Loss : $$\mathcal{L}^{(i)}(\theta_2, \theta_1, \theta_0) = \left(y^{(i)} - \left(\theta_2 x_2^{(i)} + \theta_1 x_1^{(i)} + \theta_0\right)\right)^2$$

$$\mathcal{L}^{(i)}(\vec{\theta}) = \left(y^{(i)} - \left(\vec{\theta}^{T} \cdot \vec{X}^{(i)}\right)\right)^2$$

Cost : $$J(\vec{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}^{(i)}(\vec{\theta})$$

이때도 Loss 와 Cost 모두 각각 Weight, θ_1, θ_2, θ_0 에 대해서 2차 함수이다. (∪ shape)

* 변수 3개 → 4차원에서 ∪ 모양 (밥그릇)

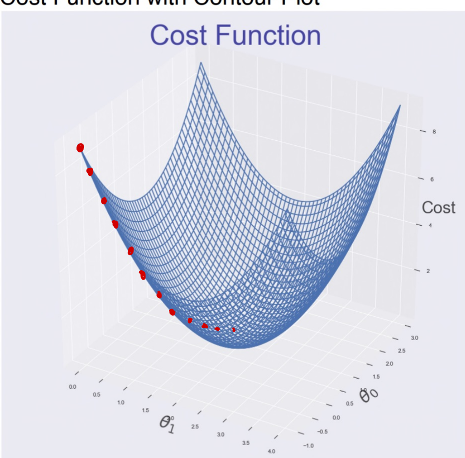
여태 구했던 Loss 와 Cost는 feature의 개수와 상관없이,

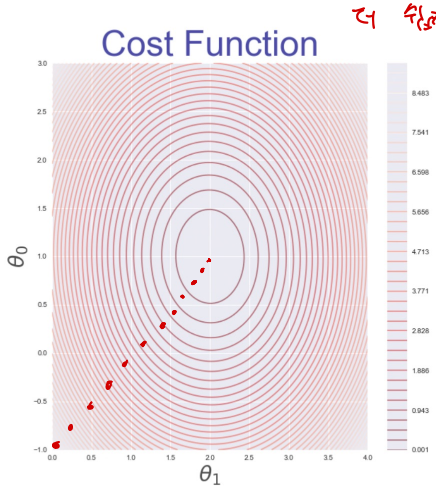
$$\vec{\theta} = \begin{pmatrix} \theta_n \\ \theta_{n-1} \\ \vdots \\ \theta_1 \\ \theta_0 \end{pmatrix} \qquad \vec{x} = \begin{pmatrix} x_n \\ x_{n-1} \\ \vdots \\ x_1 \\ x_0 \end{pmatrix}$$ 일 때,

$$\mathcal{L}^{(i)}(\vec{\theta}) = \left(y^{(i)} - \vec{\theta}^T \cdot \vec{X}^{(i)}\right)^2 \qquad J(\vec{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}^{(i)}(\vec{\theta})$$ 이다.

Contour Plot으로 최적화 진행정도를 더 쉽게 관찰.

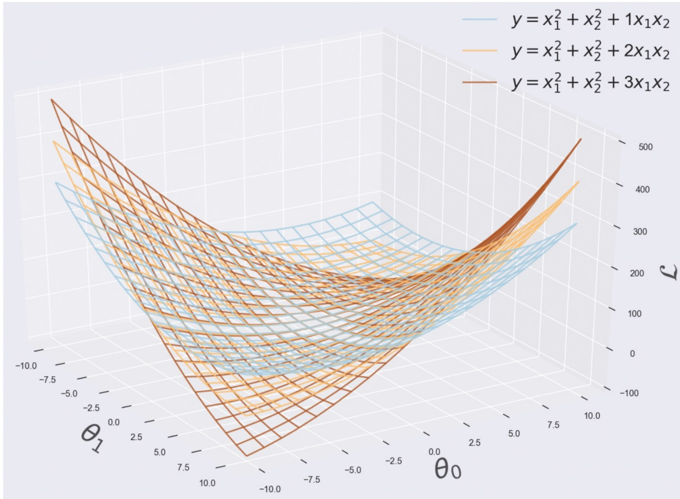
Loss/Cost Function with Contour Plot





각 점은 Cost 값을 의미함.

Real Loss/Cost Functions



실제 Cost function은 x_1x_2와 같은 term도 가질 수 있고

그 계수와 x_1, x_2가 커질수록 등고면은 점점 flat 하게 됨.

Ex) x_1, x_2가 연봉을 나타냄.