$y = \theta_1 x_1 + \theta_0$

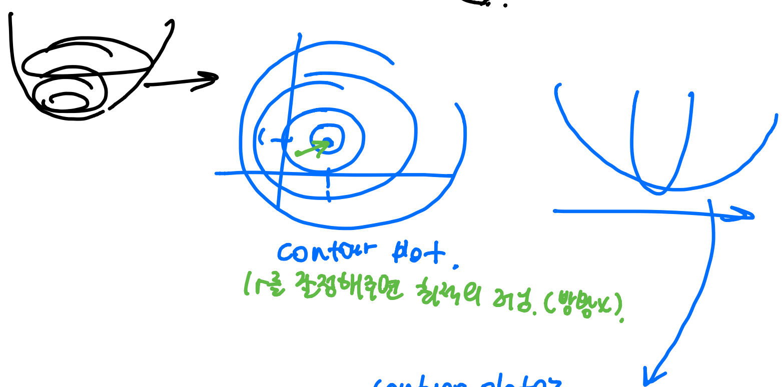
↳ $\theta = \theta - \alpha \cdot x^{(i)}(y^{(i)} - \theta_1 x^{(i)} - b)$

외우.였음.

θ_1, θ_0 의 cost function의 변화하고

3차원으로 그렸을 때 옆이놓은

밥그릇 모양. 



contour plot.

(α를 결정해주면 최적의 경로.(방향x))

contour plot로

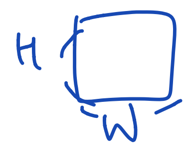
그리면 기울기 파악해염.

feature scaling .

1/20

x_2	x_1	x	y
-1	2	1	2
0	3	2	3
1	4	3	4
2.	5	4	5

+ 여기 다른 변수 넣어주기. ↳ linear 하다는 보장이 없는.



H x W 가 되면 MLP.

$$y = wx + b.$$

$$y = \theta_2 x_2 + \theta_1 x_1 + b$$

$y \propto x_2$ $y \propto x_1$

x_2와 x_1는 서로 상관관계가 없어야. correlation

→ 없어야. 가능.

서로 독립이라는 가정.

$x_2 \sim x_1$

→ covariance
correlation.

np.corrcoef (x_2, x_1)

-1 0 1

$x_1 \uparrow$ $x_2 \downarrow$

0에 가까울수록 상관관계X

$x_1 \uparrow$ $x_2 \uparrow$

	x_1	x_2
x_1	1	
x_2		1

-1 / 1 에 가까우면.

상관관계 0 이여

→ linear regression 에 들어가면 안된다는 뜻.

과제! death rate 추측. linear 하다고 가정하고. b는 eiei.

feature가 여러개 있다고 다 하면 x.

왜? 저번 정규화는 pre-processing. 이것과 같은 개념.

$x_1 \sim x_n$ 사이 상관관계 있는것 빼야. → linear reg. 망침수.

	$x_1 \cdots x_n$
x_1	
$\vdots$	상관관계가 높다? 둘다 넣으면 x.
x_n	

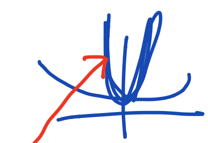
↓

조금 더 target과 유사도가 높은 x_i를 남길것.

$y = \theta_2 \underline{x_2} + \theta_1 \underline{x_1} + b$

각 각의 m, σ 맞춰주어야

($\because$ θ_i에 미치는 영향이 너무 direct?) learning에



~~[illegible]~~

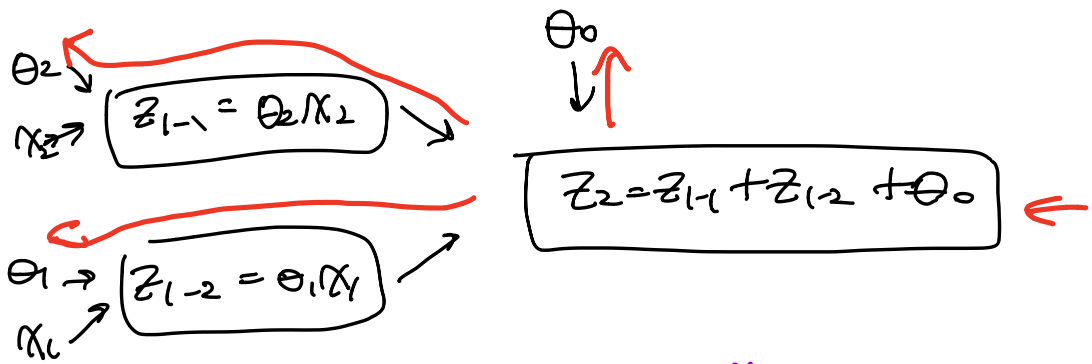
∨ x_i의 μ, σ를 보고. corr도 보고.

$y = \theta_2 x_2 + \theta_1 x_1 + \theta_0$

$\vec{\theta} = \begin{bmatrix} \theta_2 \\ \theta_1 \\ \theta_0 \end{bmatrix}$. $\vec{x} = \begin{bmatrix} x_2 \\ x_1 \\ 1 \end{bmatrix}$.

1 → dummy var. (하나의식표현가능)

$\boxed{\vec{\theta}^T \cdot \vec{x}}$



$x_2 =$

$x_1 =$

$y = x_1 + x_2 + 1 + \text{noise}$

$J(\theta_2, \theta_1, \theta_0)$

$$\text{pred} = \hat{y} = \theta_2 x_2 + \theta_1 x_1 + \theta_0$$

$$\frac{1}{N}\sum (y - \hat{y})^2 \rightarrow \frac{\partial}{\partial \theta_2} = 2(y-\hat{y}) \cdot (-x_2)$$

$$\frac{\partial}{\partial \theta_1} = 2(y-\hat{y}) \cdot (-x_1)$$

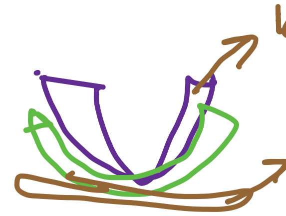
$$\frac{\partial}{\partial \theta_0} = 2(y-\hat{y}) \cdot \boxed{(-1)}$$

Jacobian matrix.

$$\begin{bmatrix} x_1^{(1)} & x_1^{(2)} \\ x_2^{(1)} & x_2^{(2)} \end{bmatrix} \qquad [y^{(1)} \ y^{(2)}].$$

feature scaling 안해주면?

θ가 3개인데



발산가능.

학습 안정.

이렇게 생길수.