확률적 경사 하강법

데이터의 양 증가.

추가되는 데이터.

기존의 모델을 업데이트 하는 방법?

=) 확률적 경사 하강법.

(머신러닝 모델 X).

=) 업데이트 할 연속함수 모델이 필요함.

(손실함수)

회귀의 경우 target data 와의 거리를 이용해 손실함수를 정의할 수 있지만

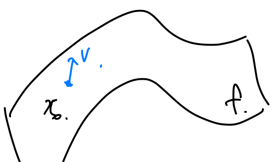
분류 모델인 경우 예측 확률을 도입해 손실을 정의 한다.

(4-1 로지스틱 모델).

Gradient

$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right).$$

Th)



$v = (v_1, v_2, v_3)$

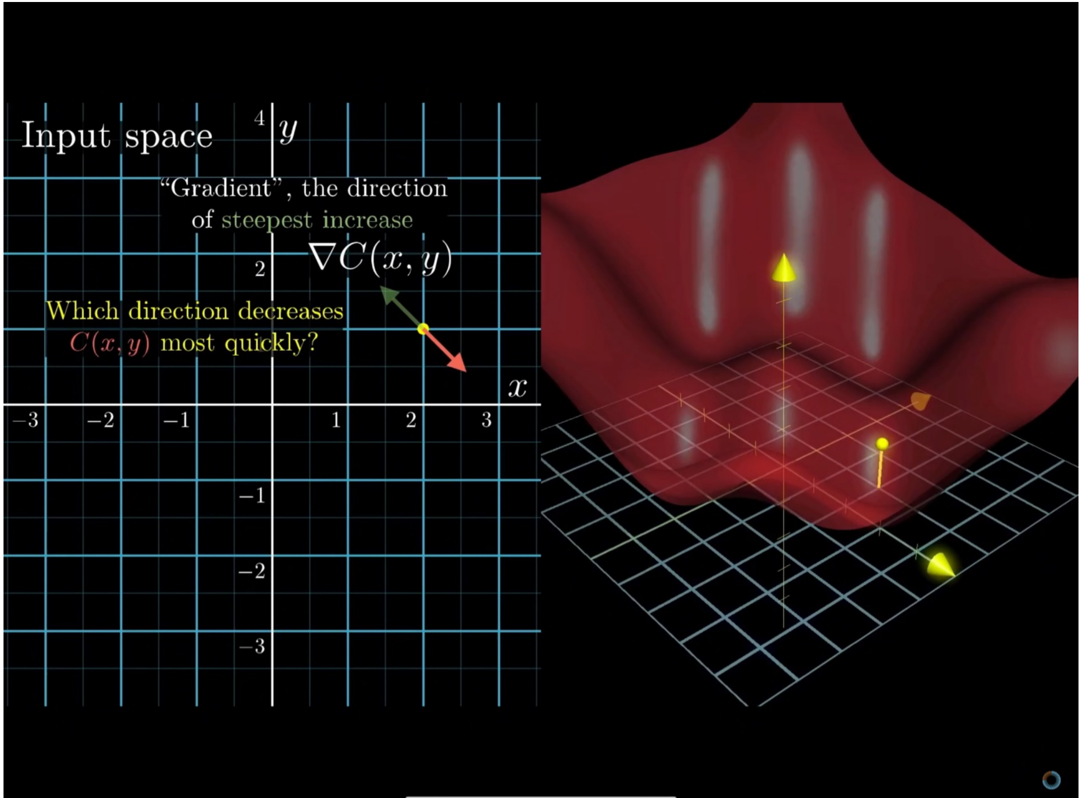
at x_0, v 방향에 따른 f의 변화율.

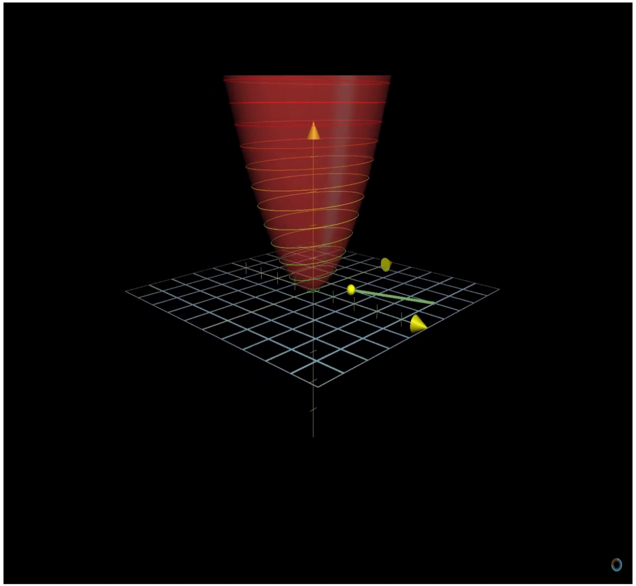
$$\boxed{\langle \nabla f(x_0), v \rangle = \frac{\partial f}{\partial x}(x_0)\, v_1 + \frac{\partial f}{\partial y}(x_0)\, v_2 + \frac{\partial f}{\partial z}(x_0)\, v_3.}$$

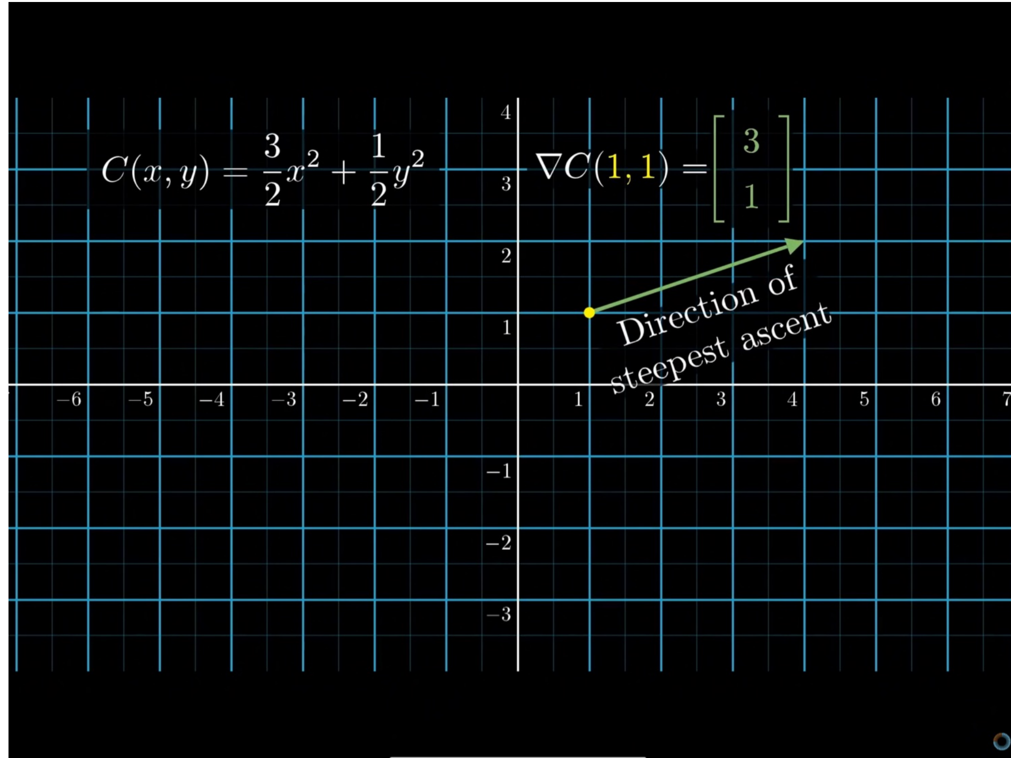
∇f_v. (방향도함수).

Th). $$\langle \nabla f, u \rangle = \underbrace{\|\nabla f\| \cdot \cos\theta}_{\text{maximum when } \theta = 0 \ (\nabla f \,//\, u)}.$$

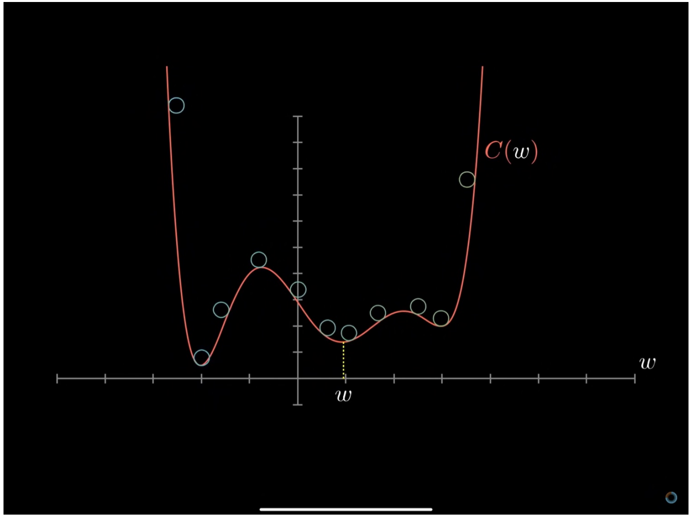
$\Rightarrow$ Gradient 방향이 변화율 최대.

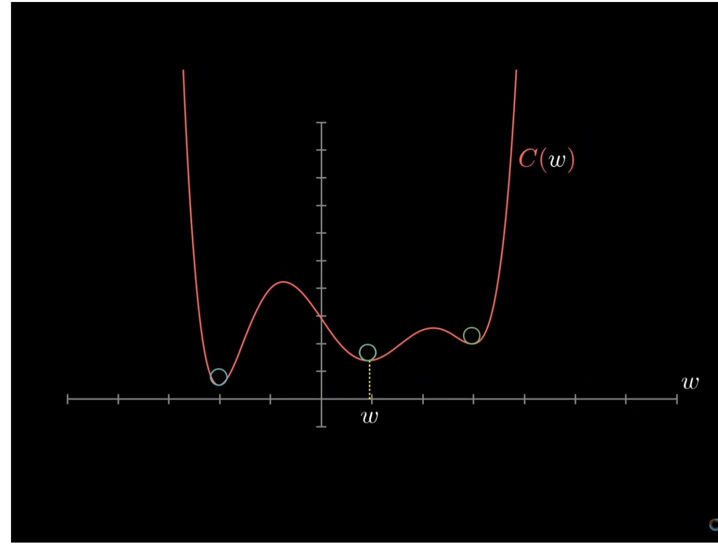

Input space
"Gradient", the direction
of steepest increase
$\nabla C(x, y)$
Which direction decreases
$C(x, y)$ most quickly?
x
y
4
2
−1
−2
−3
−3
−2
−1
1
2
3



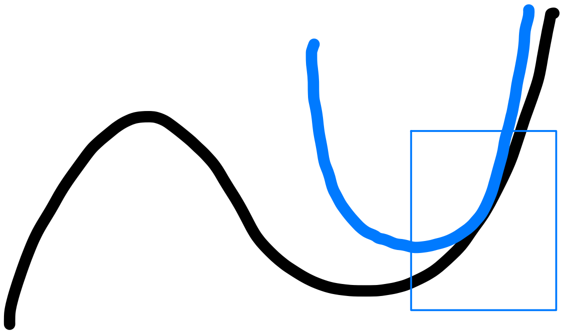


$C(x,y) = \frac{3}{2}x^2 + \frac{1}{2}y^2$
$\nabla C(1,1) = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$
Direction of
steepest ascent


$C(w)$
w
w

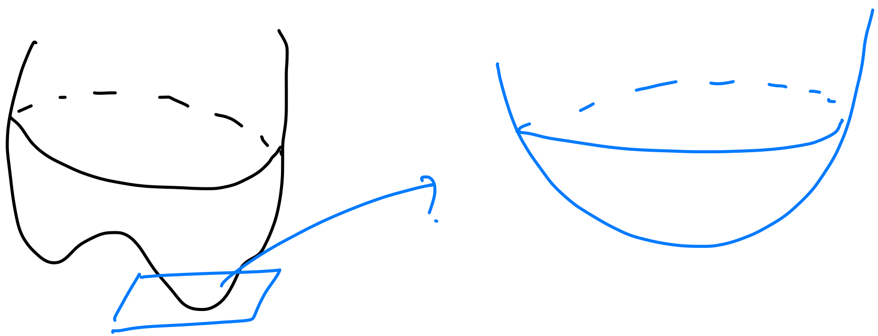

$C(w)$
w
w


Local minimum = Doable
Global minimum = Crazy hard



2차 곡면으로 근사.

=> 최솟값 존재.



$$f(x_0) \Rightarrow \frac{1}{2} f''(x_0) + f'(x_0) + f(x_0).$$

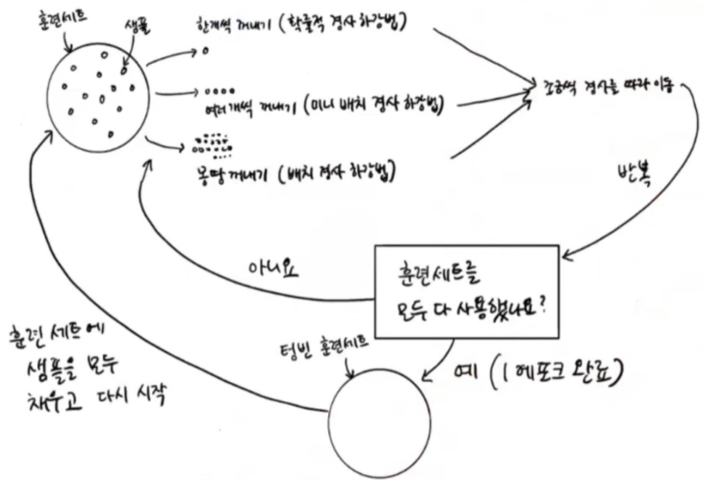
in 2D

$$F(x) = F(0) + x^T (\text{grad } F) + \frac{1}{2} x^T A x.$$

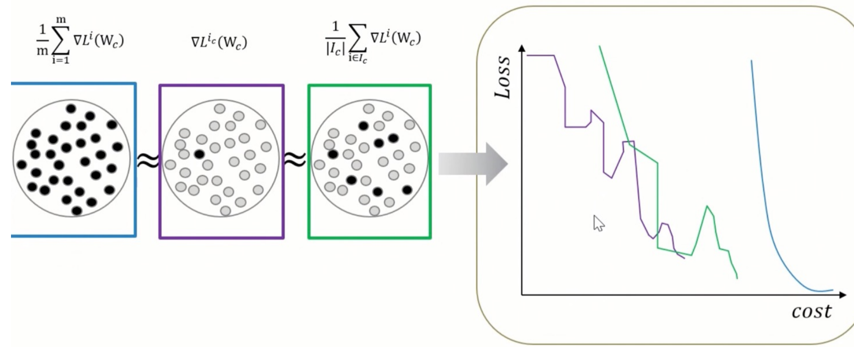
$$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

Symmetric

& semi-positive definite.


훈련세트
샘플
한개씩 꺼내기 (확률적 경사 하강법)
여러개씩 꺼내기 (미니 배치 경사 하강법)
몽땅 꺼내기 (배치 경사 하강법)
조금씩 경사를 따라 이동
반복
훈련세트를 모두 다 사용했나요?
아니요
예 (1 에포크 완료)
텅빈 훈련세트
훈련 세트에 샘플을 모두 채우고 다시 시작

Stochastic


$\frac{1}{m}\sum_{i=1}^{m}\nabla L^{i}(W_c)$
$\nabla L^{i_c}(W_c)$
$\frac{1}{|I_c|}\sum_{i\in I_c}\nabla L^{i}(W_c)$
Loss
cost