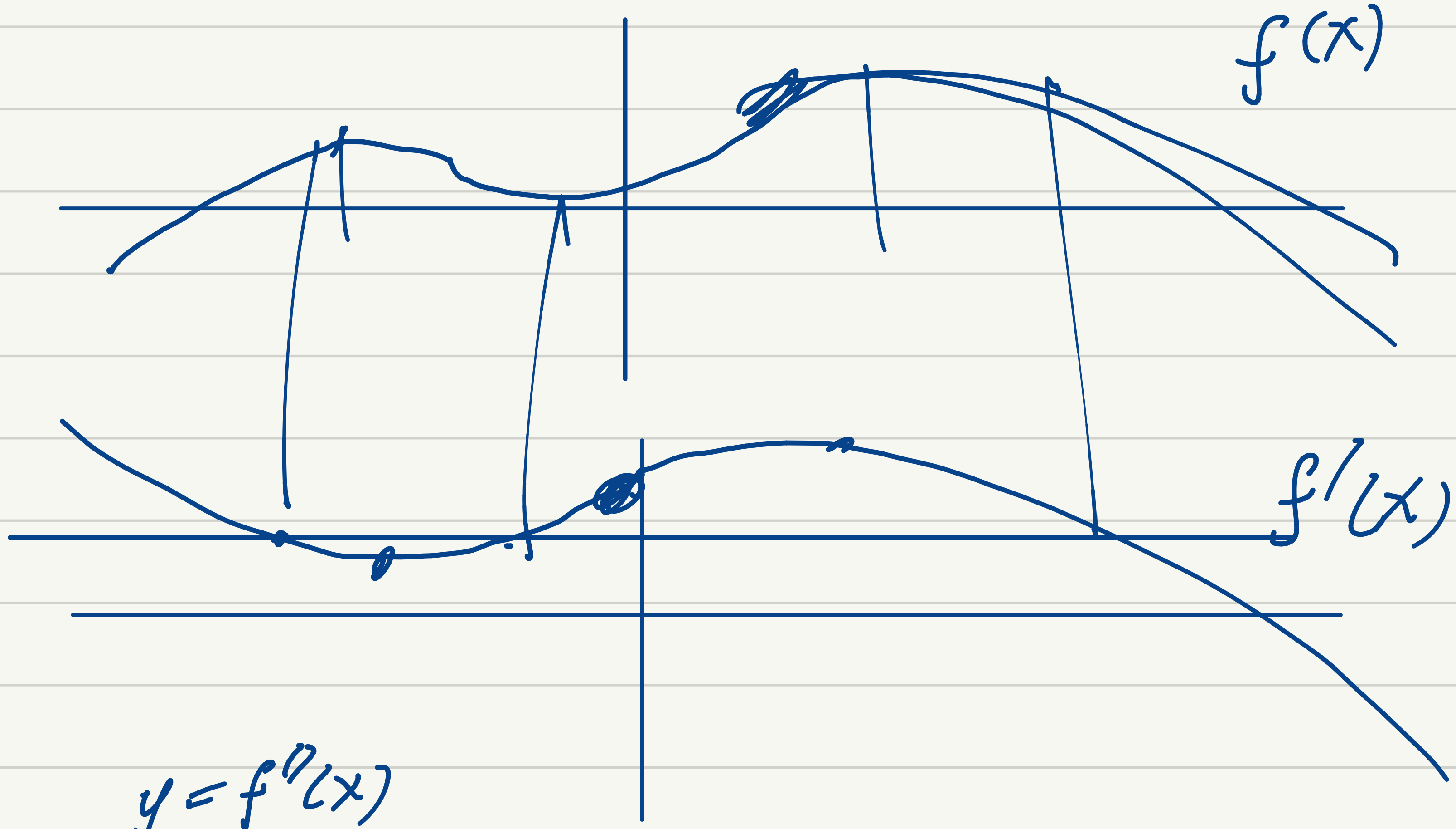
MATH 462 09-10 (F) Lecture 4

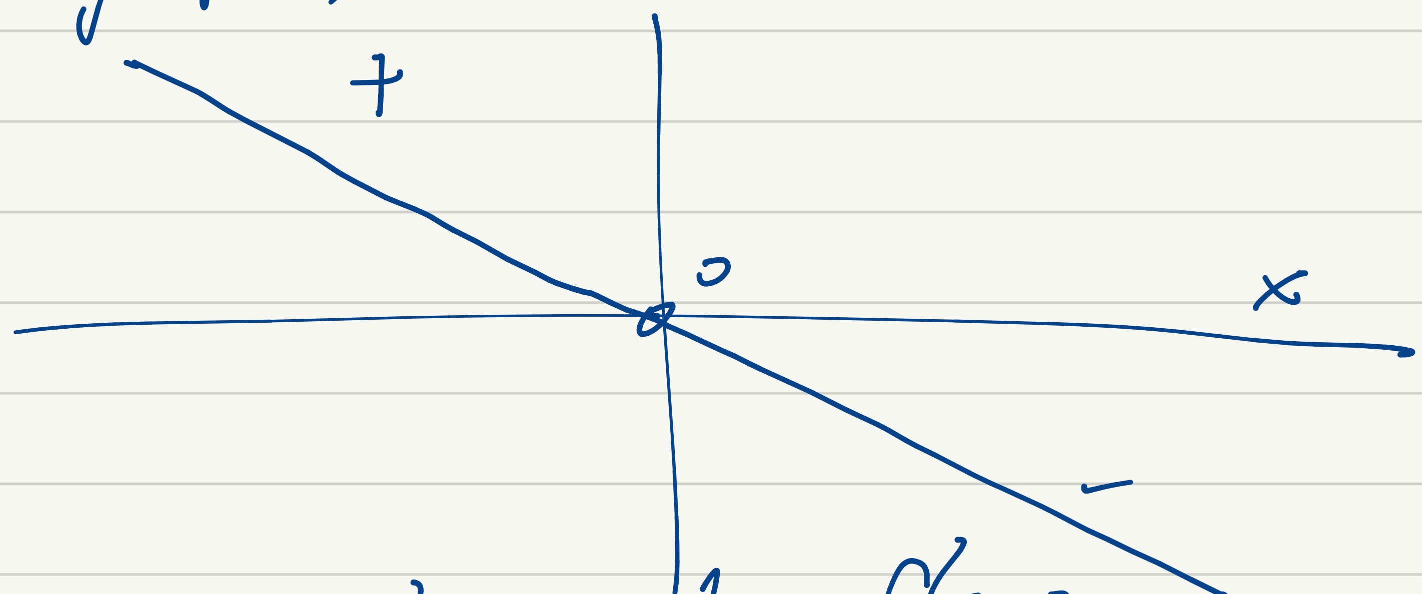
Biz:
- Lecture Recordings up
- Course Web page, adam-oberman.github.io

Today: Calculus Review, Vector Calc Review
gradient of EL (different cases)
(Tutorial Style)

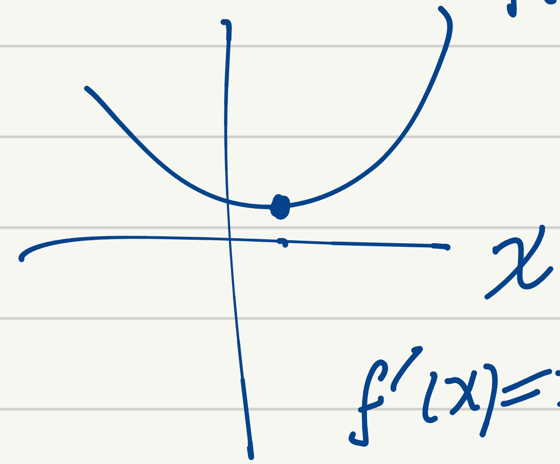
See also: typed notes.

Calc Review





$f(x) = \frac{x^2}{2}$



$f'(x) = x$

crit point $f'(x) = 0$
$x = 0$

$f''(x) \geq 0$
convex.

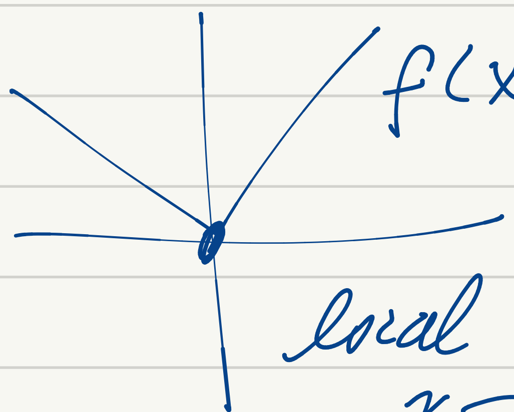
crit points $f'(x) = 0$

saddle local min ($f'' \geq 0$) local max ($f'' \leq 0$)

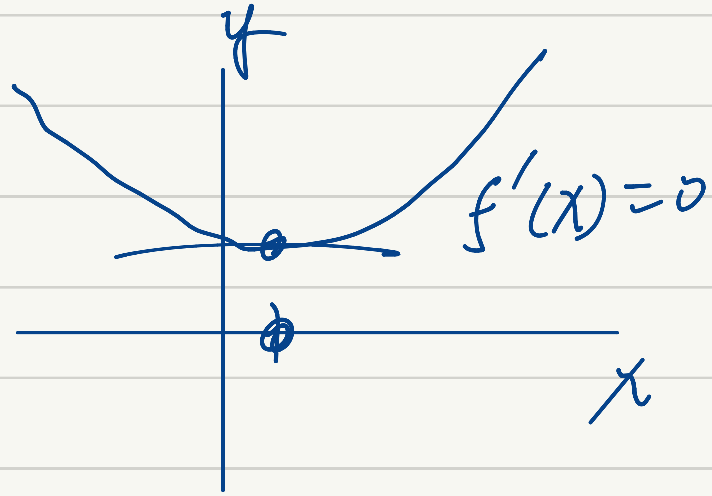
Function $f(x)$ convex

(suff cond $f''(x) \geq 0$ for all x)

$\Rightarrow$ every crit point $f'(x)=0$ is a global minimum)



local min at $x=0$ is global min.



$f''(x) \geq 0$

want $\min_{x} f(x)$ when $f(x)$ convex

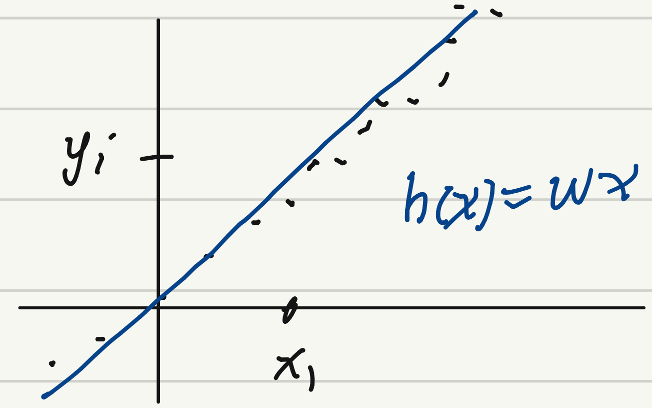
enough to find critical point $f'(x)=0$

EXAMPLE

$$f(w) = \hat{L}(w) = \frac{1}{m}\sum_{i=1}^{m} q(wx_i - y_i)$$

$w \in \mathbb{R}$

here
- model $y = wx$
- learn w
- data (x_i, y_i)
- $\hat{L}(w)$ EL.



solve $\min_{w} f(w) = \hat{L}(w)$

find $f'(w) = 0.$

Chain Rule $f(w) = q(g(w))$

$$f'(w) = q'(g(w))\, g'(w)$$

Apply $g_i(w) = wx_i - y_i$

$$g_i'(w) = x_i$$

$$f'(w) = \frac{1}{m}\sum_{i=1}^{m} q'(\underbrace{wx_i - y_i})\, x_i$$

$$f'(w) = \frac{1}{m} \sum_{i=1}^{m} q'(\underbrace{wx_i - y_i}) x_i$$

case $\quad q(e) = e^2/2$

$$f'(w) = \frac{1}{m} \sum_{i=1}^{m} (wx_i - y_i) x_i$$

$$q'(e) = e$$

$$0 = f'(w) = \hat{L}(w)$$

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

$$\Rightarrow \quad w \sum_{i=1}^{m} x_i^2 = \sum_{i=1}^{m} x_i y_i$$

$$\sum x_i^2 = x^T x = x \cdot x$$

vector notation

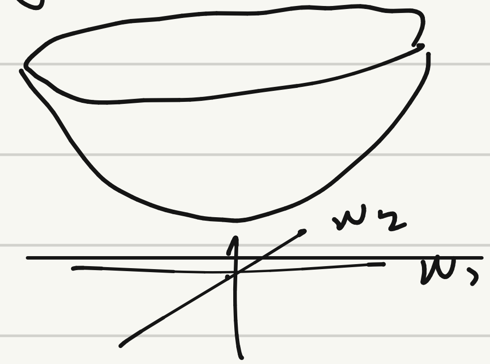
$$\sum x_i y_i = x^T y = x \cdot y$$

$$(x^T x) w = x^T y$$

Now $x, w \in \mathbb{R}^d$

$$f(w) : \mathbb{R}^d \to \mathbb{R}$$

$y = f(w_1, w_2)$

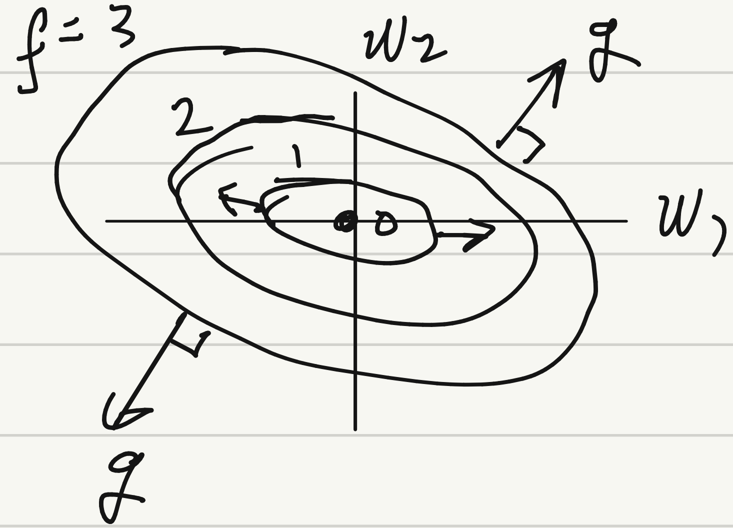


write $g(w) = \text{grad}\, f(w)$

$$= \nabla f(w)$$

$$g : \mathbb{R}^d \to \mathbb{R}^d$$

$$g_j(w) = \frac{\partial f(w)}{\partial w_j}$$



EX $f(w_1, w_2) = (3w_1 - 2w_2 - 5)^2/2$

$$\frac{\partial f}{\partial w_1} = (3w_1 - 2w_2 - 5)\,3$$

$$\frac{\partial f}{\partial w_2} = (3w_1 - 2w_2 - 5)(-2)$$

$$h(x) = \frac{1}{2}(3x - 2c - 5)^2$$

$$h'(x) = \frac{2}{2}(3x - 2c - 5)^1\, 3$$

$$= 3(3x - 2c - 5)$$

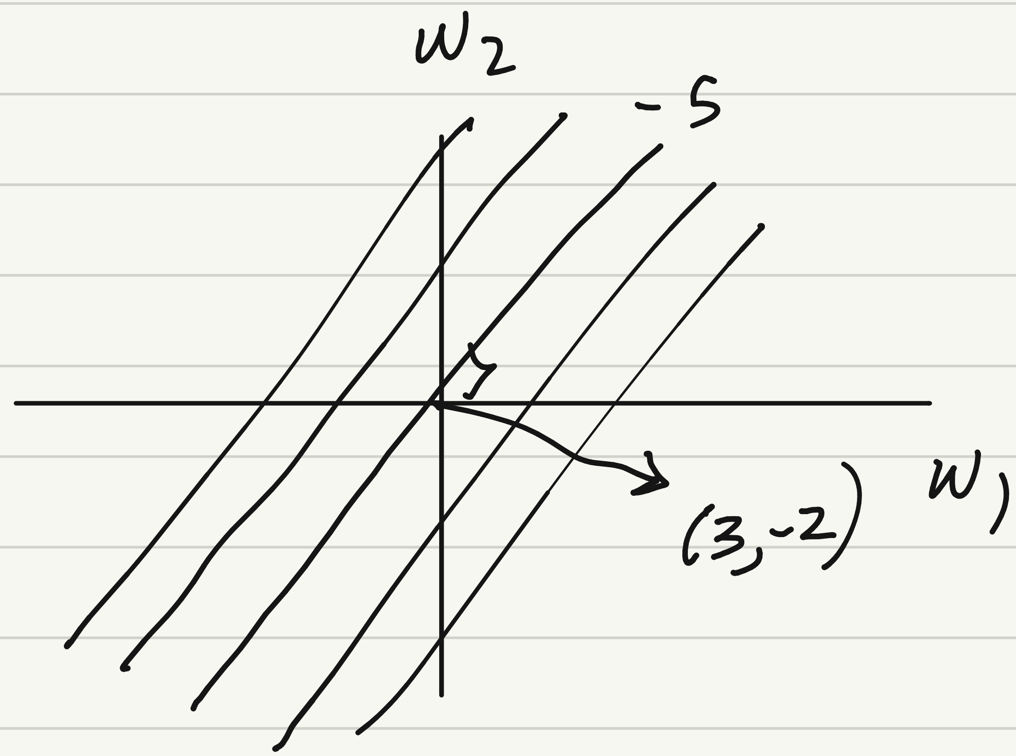
Ex $f(w_1, w_2) = (3w_1 - 2w_2 - 5)^2/2$

$$\frac{\partial f}{\partial w_1} = (3w_1 - 2w_2 - 5)3$$

$$\frac{\partial f}{\partial w_2} = (3w_1 - 2w_2 - 5)(-2)$$

$$\nabla f(w) = (3w_1 - 2w_2 - 5)\begin{bmatrix} 3 \\ -2 \end{bmatrix}$$

crit point $\nabla f(w) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

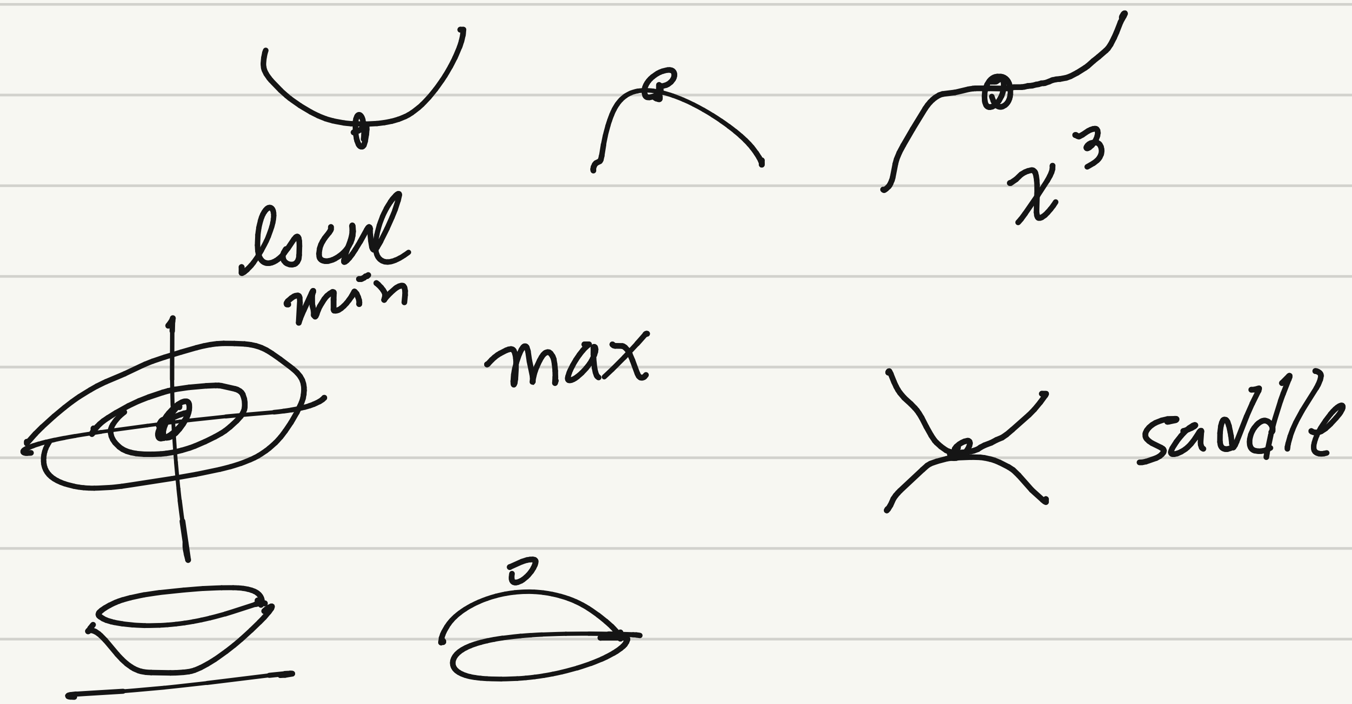


$$\begin{bmatrix} 3 \\ -2 \end{bmatrix} \cdot [w_1, w_2] - 5 = 0$$

$(2, -3)$

1d $f'(x)=0$

2d $\nabla f(x)=0$

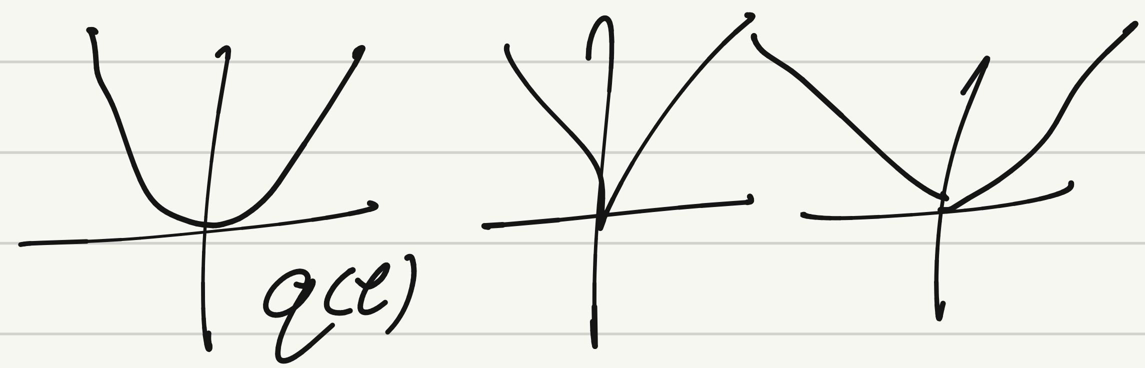


convexity $f(x)$ convex $\Rightarrow$ every critical point is a global minimum.

FACT (EL) when $q(e)$ convex

$h_w(x) = w \cdot x$

then $\hat{L}(w)$ convex.



need $\min_w \hat{L}(w)$ enough to find $\nabla \hat{L}(w) = 0$.

Clarification

$$\begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = 1 \cdot 4 + 2 \cdot 5 + 3 \cdot 6 = 4 + 10 + 18 = 32$$

$$\begin{bmatrix} 3 \\ 4 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 \cdot 1 & 3 \cdot 2 \\ 4 \cdot 1 & 4 \cdot 2 \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ 4 & 8 \end{bmatrix}$$

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \qquad x^T x = \sum x_i^2$$

$$x x^T = \text{matrix} \quad (x x^T)_{ij} = x_i x_j .$$

$$w \in \mathbb{R}^d$$

$$\hat{L}(w) = \frac{1}{m} \sum_{i=1}^{m} q(w \cdot x_i - y_i)$$

WANT $\nabla \hat{L}(w)$.

use chain rule

$e_i : \mathbb{R}^d \to \mathbb{R}$

function of w

$$e_i(w) = w \cdot x_i - y_i$$

$$= \sum_{k=1}^{d} w_k x_{ik} - y_i$$

$$x_i = \begin{bmatrix} x_{i1} \\ \vdots \\ x_{id} \end{bmatrix} \in \mathbb{R}^d$$

$$= w_1 x_{i1} + w_2 x_{i2} + \dots w_d x_{id} - y_i$$

$$\frac{\partial e_i}{\partial w_j}(w) = x_{ij}$$

$$\nabla e_i(w) = x_i$$

Chain Rule

$$\frac{\partial}{\partial w_j} q(e_i(w)) = q'(e_i(w)) \underbrace{\frac{\partial}{\partial w_j} e_i(w)}_{x_{ij}}$$

$$\nabla q(e_i(w))$$

$$= q'(e_i(w)) x_i$$

$w \in \mathbb{R}^d$

$$\hat{L}(w) = \frac{1}{m} \sum_{i=1}^{m} q(\underbrace{w \cdot x_i - y_i}_{e_i})$$

WANT $\nabla \hat{L}(w)$.

$$\frac{\partial}{\partial w_j} \hat{L}(w) = ? \quad \text{using}$$

$$\frac{\partial}{\partial w_j} q(e_i(w)) = q'(e_i(w))\, x_{ij}$$

$$\nabla q(e_i(w)) = q'(e_i)\, x_i$$

$$\Rightarrow \boxed{\nabla \hat{L}(w) = \frac{1}{m} \sum_{i=1}^{m} q'(e_i(w))\, x_i} \quad \text{Thm}$$

lots case

$$\Rightarrow \boxed{\nabla \hat{L}(w) = \frac{1}{m} \sum_{i=1}^{m} q'(e_i(w)) x_i}$$

$w \cdot x_i$

$\nabla_w (w \cdot x_i) = x_i$

$\nabla_w h_w(x_i)$

$$q(e) = e^2/2 \qquad q'(e) = e$$

$$\Rightarrow \nabla \hat{L}(w) = \frac{1}{m} \sum_{i=1}^{m} \underbrace{e_i(w)}_{(w \cdot x_i - y_i)} x_i$$

$\Rightarrow$ crit point

$$0 = \nabla \hat{L}(w) = \frac{1}{m} \sum_{i=1}^{m} (w \cdot x_i - y_i) x_i$$

$$\Rightarrow \boxed{\sum_{i=1}^{m} (w \cdot x_i) x_i = \sum_{i=1}^{m} x_i y_i}$$

$$X^T X w = X^T y$$

$$\Rightarrow \boxed{\sum_{i=1}^{m} (w \cdot x_i) x_i = \sum_{i=1}^{m} x_i y_i}$$

$$X^T X w = X^T y$$

$m = 3 \qquad d = 2$

$$X = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 0 & 1 \end{bmatrix} \qquad y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \qquad w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

$$\underbrace{\begin{bmatrix} 1 & 3 & 0 \\ 2 & 4 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 0 & 1 \end{bmatrix}}_{\begin{bmatrix} 10 & 14 \\ 14 & 21 \end{bmatrix}} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 3 & 0 \\ 2 & 4 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}}_{\begin{bmatrix} y_1 + 3y_2 \\ 2y_1 + 4y_2 + y_3 \end{bmatrix}} \rightarrow \begin{pmatrix} 1 \\ 2 \end{pmatrix} y_1 + \begin{pmatrix} 3 \\ 4 \end{pmatrix} y_2 + \begin{pmatrix} 0 \\ 1 \end{pmatrix} y_3$$

$$\begin{bmatrix} 10 & 14 \\ 14 & 21 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

$$(X^T X) w$$