

CSCI 567 - LINEAR ALGEBRA DISCUSSION II

① DERIVATIVES, GRADIENTS AND JACOBIAN

$$f: \mathbb{R}^1 \rightarrow \mathbb{R}^1$$

e.g. $f(x)$

$$\nabla f = \frac{df}{dx} = \frac{\partial f}{\partial x} = f_x$$

1×1 $\underbrace{\hspace{10em}}$
derivative

$$f: \mathbb{R}^2 \rightarrow \mathbb{R}^1$$

e.g. $f(x, y)$

$$\nabla f = \left[\begin{array}{cc} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{array} \right]_{1 \times 2}$$

$\underbrace{\hspace{10em}}$
gradient (row vector)

$$\vec{f}: \mathbb{R}^1 \rightarrow \mathbb{R}^2$$

e.g. $\begin{bmatrix} a(x) \\ b(x) \end{bmatrix}$

$$\nabla f = \begin{bmatrix} \frac{\partial a}{\partial x} \\ \frac{\partial b}{\partial x} \end{bmatrix}_{2 \times 1} = \begin{bmatrix} a_x \\ b_x \end{bmatrix}$$

$\underbrace{\hspace{10em}}$
Jacobian

$$\vec{f}: \mathbb{R}^2 \rightarrow \mathbb{R}^3$$

e.g. $\begin{bmatrix} a(x, y) \\ b(x, y) \\ c(x, y) \end{bmatrix}$

$$\nabla f = \begin{bmatrix} \frac{\partial a}{\partial x} & \frac{\partial a}{\partial y} \\ \frac{\partial b}{\partial x} & \frac{\partial b}{\partial y} \\ \frac{\partial c}{\partial x} & \frac{\partial c}{\partial y} \end{bmatrix}_{3 \times 2}$$

$\underbrace{\hspace{10em}}$
Jacobian

functions
are
always
column
vectors.

* The usual derivative is a special case of gradients which itself is a special case of jacobians.

* We always represent gradients as row vectors.
This is the numerator layout.

Exercise: Find the dimension of the jacobian/gradient

- i) $f: \mathbb{R} \rightarrow \mathbb{R}$
- ii) $f: \mathbb{R}^5 \rightarrow \mathbb{R}$
- iii) $\vec{f}: \mathbb{R} \rightarrow \mathbb{R}^3$
- iv) $\vec{f}: \mathbb{R}^7 \rightarrow \mathbb{R}^{10}$

② CHAIN RULE

a) $f: \mathbb{R} \rightarrow \mathbb{R}$

$g: \mathbb{R} \rightarrow \mathbb{R}$

$h = f \circ g : \mathbb{R} \rightarrow \mathbb{R}$

$h(x) = f(g(x))$

} Single variable

$$\nabla h = \frac{dh}{dx} = \frac{df}{dg} \cdot \frac{dg}{dx} = f'(g) g'$$

$$\nabla h = \nabla f \cdot \nabla_x g$$

$$\textcircled{b} \quad f: \mathbb{R} \rightarrow \mathbb{R}$$

$$g: \mathbb{R}^2 \rightarrow \mathbb{R} \quad \text{e.g. } g(x, y)$$

$$h: f \circ g: \mathbb{R}^2 \rightarrow \mathbb{R}$$

$$h(x, y) = f(g(x, y))$$

$$\nabla h = [h_x \ h_y] = \left[\frac{\partial h}{\partial x} \quad \frac{\partial h}{\partial y} \right]$$

$$\frac{\partial h}{\partial x} = \frac{\partial f}{\partial g} \cdot \frac{\partial g}{\partial x} \quad \text{and} \quad \frac{\partial h}{\partial y} = \frac{\partial f}{\partial g} \cdot \frac{\partial g}{\partial y}$$

$$\nabla h = \begin{bmatrix} \frac{\partial f}{\partial g} & \frac{\partial g}{\partial x} \\ \frac{\partial f}{\partial g} & \frac{\partial g}{\partial y} \end{bmatrix}_{1 \times 2} = \begin{bmatrix} \frac{\partial f}{\partial g} \end{bmatrix}_{1 \times 1} \begin{bmatrix} \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{bmatrix}_{1 \times 2}$$

$$\nabla h = \begin{bmatrix} \nabla_g f \\ 1 \times 2 \end{bmatrix} \cdot \begin{bmatrix} \nabla_{(x,y)} g \\ 1 \times 1 \end{bmatrix}_{1 \times 2}$$

$$\textcircled{c} \quad \vec{f}: \mathbb{R} \rightarrow \mathbb{R}^2 \quad \text{e.g. } \vec{f}(x) = \begin{bmatrix} a(x) \\ b(x) \end{bmatrix}$$

$$g: \mathbb{R} \rightarrow \mathbb{R} \quad \text{e.g. } g(x)$$

$$h = f \circ g: \mathbb{R} \rightarrow \mathbb{R}^2$$

$$h(x) = f(g(x)) = \begin{bmatrix} a(g(x)) \\ b(g(x)) \end{bmatrix}$$

$$\nabla h = \begin{bmatrix} \frac{\partial a}{\partial x} \\ \frac{\partial b}{\partial x} \end{bmatrix}_{2 \times 1} = \begin{bmatrix} \frac{\partial a}{\partial g} & \frac{\partial a}{\partial x} \\ \frac{\partial b}{\partial g} & \frac{\partial b}{\partial x} \end{bmatrix}_{2 \times 1}$$

$$= \begin{bmatrix} \frac{\partial a}{\partial g} \\ \frac{\partial b}{\partial g} \end{bmatrix}_{2 \times 1} \begin{bmatrix} \frac{\partial g}{\partial x} \end{bmatrix}_{1 \times 1}$$

$$\nabla h = \begin{bmatrix} \nabla_g f \\ \nabla_x g \end{bmatrix}_{2 \times 1} \cdot \begin{bmatrix} \nabla_x g \end{bmatrix}_{1 \times 1}$$

d) $f: \mathbb{R}^3 \rightarrow \mathbb{R}$

$$\vec{g}: \mathbb{R}^2 \rightarrow \mathbb{R}^3$$

$$h = f \circ g : \mathbb{R}^2 \rightarrow \mathbb{R}$$

$$f = f(p, q, r) \quad \vec{g} = g(x, y) = \begin{bmatrix} p(x, y) \\ q(x, y) \\ r(x, y) \end{bmatrix}$$

$$\frac{\partial h}{\partial x} = \frac{\partial f}{\partial p} \frac{\partial p}{\partial x} + \frac{\partial f}{\partial q} \frac{\partial q}{\partial x} + \frac{\partial f}{\partial r} \frac{\partial r}{\partial x}$$

$$\frac{\partial h}{\partial y} = \frac{\partial f}{\partial p} \frac{\partial p}{\partial y} + \frac{\partial f}{\partial q} \frac{\partial q}{\partial y} + \frac{\partial f}{\partial r} \frac{\partial r}{\partial y}$$

$$\nabla h = \begin{bmatrix} \frac{\partial h}{\partial x} \\ \frac{\partial h}{\partial y} \end{bmatrix}_{1 \times 2}$$

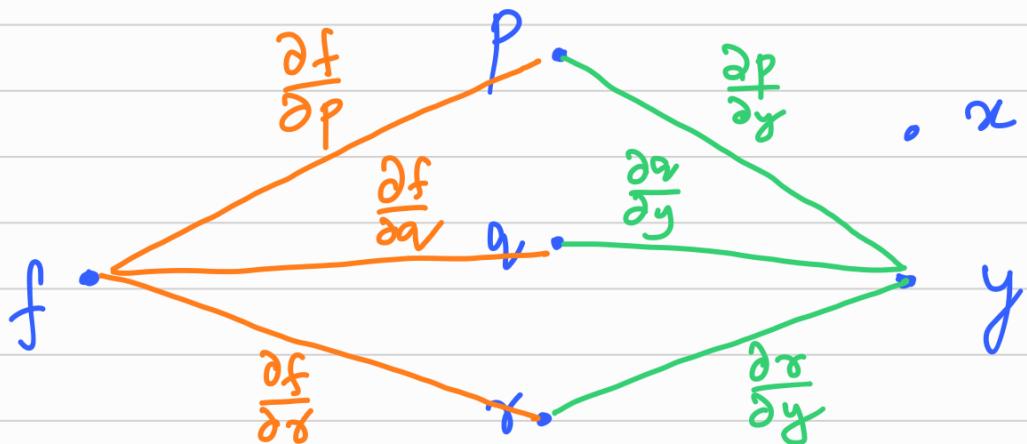
$$\begin{bmatrix} \frac{\partial f}{\partial p} \frac{\partial p}{\partial x} + \frac{\partial f}{\partial v} \frac{\partial v}{\partial x} & \frac{\partial f}{\partial p} \frac{\partial p}{\partial y} + \frac{\partial f}{\partial v} \frac{\partial v}{\partial y} + \frac{\partial f}{\partial x} \frac{\partial x}{\partial y} \end{bmatrix}_{1 \times 2}$$

$$= \begin{bmatrix} \frac{\partial f}{\partial p} & \frac{\partial f}{\partial v} & \frac{\partial f}{\partial x} \end{bmatrix}_{1 \times 3} \begin{bmatrix} \frac{\partial p}{\partial x} & \frac{\partial p}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \\ \frac{\partial x}{\partial x} & \frac{\partial x}{\partial y} \end{bmatrix}_{3 \times 2}$$

3×2

$$\nabla h = \nabla f(p, v) \cdot \nabla \vec{g}(u, y)$$

$1 \times 3 \quad 3 \times 2$



$$\frac{\partial h}{\partial y} = \frac{\partial f}{\partial p} \frac{\partial p}{\partial y} + \frac{\partial f}{\partial q} \frac{\partial q}{\partial y} + \frac{\partial f}{\partial s} \frac{\partial s}{\partial y}$$

(e) $\vec{f}: \mathbb{R}^2 \rightarrow \mathbb{R}^3,$

$$\vec{g}: \mathbb{R}^3 \rightarrow \mathbb{R}^2$$

$$\vec{h} = \vec{f} \circ \vec{g}: \mathbb{R}^3 \rightarrow \mathbb{R}^3$$

e.g. $\vec{f} = f(p, q) = \begin{bmatrix} a(p, q) \\ b(p, q) \\ c(p, q) \end{bmatrix}$

$$\vec{g} = g(x, y, z) = \begin{bmatrix} p(x, y, z) \\ q(x, y, z) \end{bmatrix}$$

$$h = \begin{bmatrix} a(p(x, y, z), q(x, y, z)) \\ b(p(x, y, z), q(x, y, z)) \\ c(p(x, y, z), q(x, y, z)) \end{bmatrix}$$

$$\nabla h(x, y, z) = \begin{bmatrix} \frac{\partial a}{\partial x} & \frac{\partial a}{\partial y} & \frac{\partial a}{\partial z} \\ \frac{\partial b}{\partial x} & \frac{\partial b}{\partial y} & \frac{\partial b}{\partial z} \\ \frac{\partial c}{\partial x} & \frac{\partial c}{\partial y} & \frac{\partial c}{\partial z} \end{bmatrix}_{3 \times 3}$$

$$= \begin{bmatrix} \frac{\partial a}{\partial p} \frac{\partial p}{\partial x} + \frac{\partial a}{\partial q} \frac{\partial q}{\partial x} & \frac{\partial a}{\partial p} \frac{\partial p}{\partial y} + \frac{\partial a}{\partial q} \frac{\partial q}{\partial y} & \frac{\partial a}{\partial p} \frac{\partial p}{\partial z} + \frac{\partial a}{\partial q} \frac{\partial q}{\partial z} \\ \frac{\partial b}{\partial p} \frac{\partial p}{\partial x} + \frac{\partial b}{\partial q} \frac{\partial q}{\partial x} & \frac{\partial b}{\partial p} \frac{\partial p}{\partial y} + \frac{\partial b}{\partial q} \frac{\partial q}{\partial y} & \frac{\partial b}{\partial p} \frac{\partial p}{\partial z} + \frac{\partial b}{\partial q} \frac{\partial q}{\partial z} \\ \frac{\partial c}{\partial p} \frac{\partial p}{\partial x} + \frac{\partial c}{\partial q} \frac{\partial q}{\partial x} & \frac{\partial c}{\partial p} \frac{\partial p}{\partial y} + \frac{\partial c}{\partial q} \frac{\partial q}{\partial y} & \frac{\partial c}{\partial p} \frac{\partial p}{\partial z} + \frac{\partial c}{\partial q} \frac{\partial q}{\partial z} \end{bmatrix}_{3 \times 3}$$

$$= \begin{bmatrix} \frac{\partial a}{\partial p} & \frac{\partial a}{\partial q} \\ \frac{\partial b}{\partial p} & \frac{\partial b}{\partial q} \\ \frac{\partial c}{\partial p} & \frac{\partial c}{\partial q} \end{bmatrix}_{3 \times 2} \begin{bmatrix} \frac{\partial p}{\partial x} & \frac{\partial p}{\partial y} & \frac{\partial p}{\partial z} \\ \frac{\partial q}{\partial x} & \frac{\partial q}{\partial y} & \frac{\partial q}{\partial z} \end{bmatrix}_{2 \times 3}$$

$$\Rightarrow \vec{\nabla} h(x, y, z) = \vec{\nabla} f(p, q) \cdot \vec{\nabla} g(x, y, z)$$

3×3 3×2 2×3

5.5 Consider the following functions:

$$\begin{aligned}f_1(\mathbf{x}) &= \sin(x_1) \cos(x_2), \quad \mathbf{x} \in \mathbb{R}^2 \\f_2(\mathbf{x}, \mathbf{y}) &= \mathbf{x}^\top \mathbf{y}, \quad \mathbf{x}, \mathbf{y} \in \mathbb{R}^n \\f_3(\mathbf{x}) &= \mathbf{x} \mathbf{x}^\top, \quad \mathbf{x} \in \mathbb{R}^n\end{aligned}$$

Draft (2024-01-15) of “Mathematics for Machine Learning”. Feedback: <https://mml-book.com>.

Exercises

171

- a. What are the dimensions of $\frac{\partial f_i}{\partial \mathbf{x}}$?
 - b. Compute the Jacobians.

$$\textcircled{a} \quad f_j : \mathbb{R}^2 \rightarrow \mathbb{R}^1, \nabla f_j \in \mathbb{R}^{1 \times 2}$$

$$f_2: \mathbb{R}^{2n} \rightarrow \mathbb{R}^1, \quad \nabla f_2 \in \mathbb{R}^{1 \times 2n}$$

$$f_3 : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}, \quad \nabla f_3 \in \mathbb{R}^{(n \times n) \times n}$$

3D matrix OR
3rd order tensor

$$\textcircled{b} \quad f_1(x_1, x_2) = \sin(x_1) \cos(x_2)$$

$$\nabla f_1 = \left[\frac{\partial f_1}{\partial x_1} \quad \frac{\partial f_1}{\partial x_2} \right]$$

$$= \begin{bmatrix} \cos(x_1) \cos(x_2) & -\sin(x_1) \sin(x_2) \end{bmatrix}$$

$$f_2 = \mathbf{x}^T \mathbf{y} . \quad \text{Let } \mathbf{x}, \mathbf{y} \in \mathbb{R}^2$$

$$f_2(x_1, x_2, y_1, y_2) = [x_1, x_2] \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

$$= x_1 y_1 + x_2 y_2$$

$$\nabla f_2 = \left[\frac{\partial f_2}{\partial x_1} \quad \frac{\partial f_2}{\partial x_2} \quad \frac{\partial f_2}{\partial y_1} \quad \frac{\partial f_2}{\partial y_2} \right]$$

$$= \begin{bmatrix} y_1 & y_2 & x_1 & x_2 \end{bmatrix}$$

\therefore in general, $\nabla f_2 = [\vec{y} : \vec{x}]$

$f_3 = \mathbf{x}\mathbf{x}^T$, let $\mathbf{x} \in \mathbb{R}^2$

$$f_3 = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} [x_1 \ x_2]$$

$$= \begin{bmatrix} x_1^2 & x_1 x_2 \\ x_2 x_1 & x_2^2 \end{bmatrix}$$

$$\nabla f_3 = \begin{bmatrix} \frac{\partial x_1^2}{\partial x_2} & \frac{\partial (x_1 x_2)}{\partial x_2} \\ \frac{\partial (x_2 x_1)}{\partial x_1} & \frac{\partial x_2^2}{\partial x_1} \end{bmatrix}$$

$$= \begin{bmatrix} 0 & x_1 \\ x_1 & 2x_2 \end{bmatrix}$$
$$\begin{bmatrix} 2x_1 & x_2 \\ x_2 & 0 \end{bmatrix}$$

Can be generalized when $x \in \mathbb{R}^n$
(we will skip this)

RESULTS TO REMEMBER

① $f(\vec{x}) = \vec{a}^T \vec{x}$ $\nabla f(x) = \underline{\vec{a}^T}$
 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ gradient always a
 $\vec{x} \in \mathbb{R}^n$ row vector



② $f(\vec{x}) = \vec{x}^T \vec{a}$ $\nabla f(x) = \vec{a}^T$
 $f: \mathbb{R}^n \rightarrow \mathbb{R}$
 $\vec{x} \in \mathbb{R}^n$

③ $\vec{f}(\vec{x}) = A\vec{x}$ $\nabla f = A$

$$A \in \mathbb{R}^{m \times n}$$
$$\vec{x} \in \mathbb{R}^n$$

$$f: \mathbb{R}^n \rightarrow \mathbb{R}^m$$

Ex: $A = \begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix}$ $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$$f(\mathbf{x}) = A \mathbf{x} = \begin{bmatrix} ax_1 + bx_2 \\ cx_1 + dx_2 \\ ex_1 + fx_2 \end{bmatrix}$$

$$\nabla f = \begin{bmatrix} \frac{\partial}{\partial x_1} (ax_1 + bx_2) & b \\ c & \frac{\partial}{\partial x_2} (cx_1 + dx_2) \\ e & f \end{bmatrix}$$

$$= A$$

$$\text{COROLLARY} \quad \bar{f}(\vec{x}) = \vec{x} = \mathbf{I} \vec{x}$$

$$\text{then } \nabla f = \mathbf{I}$$

$$\textcircled{4} \quad \vec{f}(x) = x^T A \text{ is a row vector}$$

But vector functions are always

column vectors by convention.

So we convert $\vec{f}(x)$ into a column vector first.

$$\vec{f}(x) = (x^T A)^T = A^T x$$

$$\nabla f = A^T$$

$$\textcircled{5} \quad f(x) = x^T A x$$

$$\text{Let } g(x) = Ax, \quad \nabla g = A$$

$$f(x) = x^T g$$

$$\nabla f = g^T + x^T \nabla g \quad (\text{product rule})$$

$$= (Ax)^T + x^T A$$

$$= x^T A^T + x^T A$$

$$= x^T (A^T + A) \leftarrow \text{row vector}$$

COROLLARY

$$\|x\|^2 = x^T x = x^T I x$$

$$\nabla \|x\|^2 = \nabla (x^T (I + I^T)) = 2x^T$$

Q3 $f(\omega) = \|x\omega - y\|_2^2 + \omega^T M \omega, \omega \in \mathbb{R}^d$

Find $\omega^* = \arg \min f(\omega)$

Solⁿ Set $\nabla f = 0$

$$\nabla f = \nabla \|z\|^2 + \nabla (\omega^T M \omega)$$

$$= 2z^T \nabla z + \omega^T (M + M^T)$$

$$= 2z^T x + \omega^T (M + M^T)$$

$$= 2(x\omega - y)^T x + \omega^T (M + M^T)$$

$$= 2(\omega^T x^T x - y^T x) + \omega^T (M + M^T)$$

$$\nabla f = 0 \Rightarrow \omega^T (2x^T x + M + M^T) = 2y^T x$$

$$\Rightarrow (2x^T x + M^T + M) \omega = 2x^T y$$

$$\Rightarrow (x^T x + \frac{M^T + M}{2}) \omega = x^T y$$

$$\Rightarrow \omega^* = (x^T x + \frac{M^T + M}{2})^{-1} x^T y$$

solⁿ²

$$f(\omega) = \|x\omega - y\|_2^2 + \omega^T M \omega$$

$$\nabla f = \nabla \|x\omega - y\|_2^2 + \nabla (\omega^T M \omega)$$

$$= \nabla \|x\omega - y\|_2^2 + \omega^T (M + M^T)$$

$$\|x\omega - y\|_2^2 = (x\omega - y)^T (x\omega - y)$$

$$= (\omega^T x^T - y^T)(x\omega - y)$$

$$= \omega^T x^T x\omega - \omega^T x^T y - y^T x\omega + y^T y$$

$$\Rightarrow \nabla \|x\omega - y\|_2^2 = \omega^T (x^T x + x^T x) - y^T x - y^T x + 0$$

$$= 2\omega^T x^T x - 2y^T x$$

$$\therefore \nabla f = 2\omega^T x^T x - 2y^T x + \omega^T (M + M^T)$$

$$\Rightarrow \nabla f = 0 \Rightarrow 2\omega^T (x^T x + \frac{M + M^T}{2}) = 2y^T x$$

$$\Rightarrow \left(x^T x + \frac{M^T + M}{2}\right) \omega = x^T y$$

$$\Rightarrow \omega^* = \left(x^T x + \frac{M^T + M}{2}\right)^{-1} x^T y$$

COROLLARY 1: LINEAR REGRESSION

(ORDINARY LEAST SQUARES)

In linear regression, $\hat{y} = X\omega$ and

$$\text{error} = \|\hat{y} - y\|_2^2 = \|X\omega - y\|_2^2$$

Here $M = 0$

then $f(\omega) = \|X\omega - y\|_2^2$ = error in linear regression

$$\omega^* = \left(X^T X + \frac{M + M^T}{2} \right)^{-1} X^T y$$

$$\boxed{\omega^* = (X^T X)^{-1} X^T y}$$

ω^* = Solution of best parameters for the linear regression model $\hat{y} = X\omega$.

Note: $X^T X \omega = X^T y$ is called the normal equation.

ω^* is a solution of the normal equation.

There could be multiple solutions based on whether $X^T X$ is invertible or not.

COROLLARY 2 : Regularization

If $M = \lambda I$

$$\text{Then } f(w) = \|Xw - y\|_2^2 + w^T(\lambda I)w$$

$$= \|Xw - y\|_2^2 + \lambda w^T I w$$

$$= \|Xw - y\|_2^2 + \lambda w^T w$$

$$= \|Xw - y\|_2^2 + \lambda \|w\|^2$$

\rightsquigarrow Regularizer
to control model
complexity.

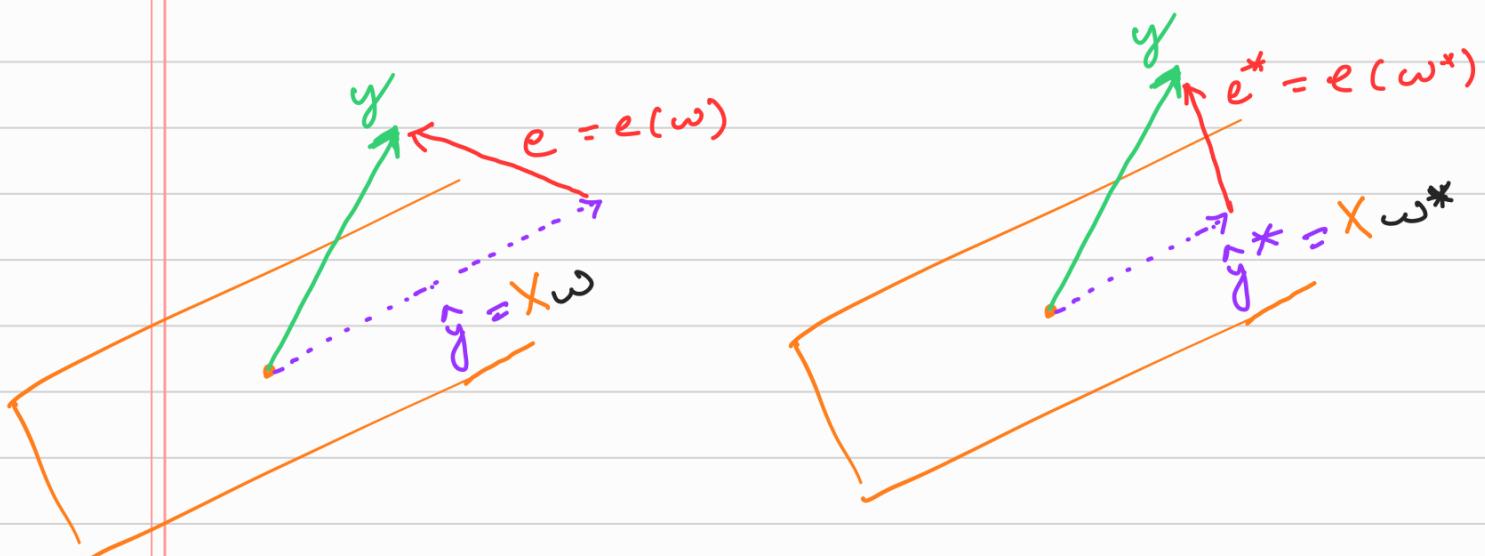
$$\begin{aligned} \text{Then } w^* &= \left[X^T X + \frac{\lambda I + (\lambda I)^T}{2} \right]^{-1} X^T y \\ &= (X^T X + \lambda I)^{-1} X^T y \end{aligned}$$

GEOMETRIC INTERPRETATION OF OLS / LINEAR REGRESSION SOLUTION

$$\hat{y} = X\omega$$

$$\vec{e} = y - \hat{y} = y - X\omega$$

$$\omega^* = \operatorname{arg\,min} \|X\omega - y\|_2^2$$



Clearly, $e^* \perp$ Subspace (X) $(X^T X)^{-1} X^T y$

$\Rightarrow y - \hat{y}^* \perp$ each column of X

$$\Rightarrow X^T (y - \hat{y}^*) = 0$$

$$\Rightarrow X^T (y - X\omega^*) = 0$$

$$\Rightarrow X^T y - X^T X \omega^* = 0$$

$$\Rightarrow X^T X \omega^* = X^T y \quad (\text{Normal eqn})$$

$$\Rightarrow w^* = (X^\top X)^{-1} X^\top y$$