CSCI 567 - LINEAR ALGEBRA DISCUSSION II

① DERIVATIVES, GRADIENTS AND JACOBIAN

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

e.g. $f(x)$

$$\nabla f = \underbrace{\frac{df}{dx} = \frac{\partial f}{\partial x} = f_x}_{\text{derivative}}$$

1×1

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

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

$$\nabla f = \underbrace{\begin{array}{cc} x & y \\ \left[\frac{\partial f}{\partial x} \right. & \left. \frac{\partial f}{\partial y} \right]_{1 \times 2} \end{array}}_{\text{gradient (row vector)}}$$

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

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

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

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

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

$$\nabla f = \underbrace{\begin{array}{cc} x & y \\ \multicolumn{2}{c}{\begin{bmatrix} \frac{\partial a}{\partial x} & \frac{\partial a}{\partial y} \\ \frac{\partial b}{\partial x} & \frac{\partial b}{\partial y} \\ \frac{\partial c}{\partial x} & \frac{\partial c}{\partial y} \end{bmatrix}_{3 \times 2}} \end{array}}_{\text{Jacobian}}$$

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) $f: \mathbb{R} \longrightarrow \mathbb{R}$ $\quad 1 \times 1$

ii) $f: \mathbb{R}^5 \longrightarrow \mathbb{R}$ $\quad 1 \times 5$

iii) $\vec{f}: \mathbb{R} \longrightarrow \mathbb{R}^3$ $\quad 3 \times 1$

iv) $\vec{f}: \mathbb{R}^7 \longrightarrow \mathbb{R}^{10}$ $\quad 10 \times 7$

② CHAIN RULE

ⓐ
$$\left.\begin{aligned} &f: \mathbb{R} \longrightarrow \mathbb{R} \\ &g: \mathbb{R} \longrightarrow \mathbb{R} \\ &h = f \circ g : \mathbb{R} \longrightarrow \mathbb{R} \end{aligned}\right\} \text{Single variable}$$

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

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

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

(b) $f: \mathbb{R} \rightarrow \mathbb{R}$ $\longrightarrow$ 1×1

$g: \mathbb{R}^2 \rightarrow \mathbb{R}$ e.g. $g(x,y)$ $\longrightarrow$ 1×2

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

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

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

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

$$\nabla h = \underbrace{\left[\frac{\partial f}{\partial g}\frac{\partial g}{\partial x} \quad \frac{\partial f}{\partial g}\frac{\partial g}{\partial y} \right]}_{1\times2} = \underbrace{\left[\frac{\partial f}{\partial g} \right]}_{1\times1} \underbrace{\left[\frac{\partial g}{\partial x} \quad \frac{\partial g}{\partial y} \right]}_{1\times2}$$

$$\rightarrow \underbrace{\nabla h}_{1\times2} = \underbrace{\nabla_g f}_{1\times1} \cdot \underbrace{\nabla_{(x,y)} g}_{1\times2}$$

$h' = f'(g)\, g'$

$\partial f \quad \nabla g$

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

$g: \mathbb{R} \longrightarrow \mathbb{R}$ e.g. $g(x)$

$h = f \circ g : \mathbb{R} \longrightarrow \mathbb{R}^2$ $\longrightarrow$ $\underline{2\times1}$

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

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

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

$$\underset{2\times 1}{\nabla h} = \underset{2\times 1}{\nabla_g f} \cdot \underset{1\times 1}{\nabla_x g}$$

(d)

$$f : \mathbb{R}^3 \longrightarrow \mathbb{R} \rightarrow 1\times 3$$
$$\vec{g} : \mathbb{R}^2 \longrightarrow \mathbb{R}^3 \rightarrow 3\times 2$$
$$h = f \circ g : \mathbb{R}^2 \longrightarrow \mathbb{R} \rightarrow \underline{1\times 2}$$

$$f : \underline{f(p,q,r)} \text{ \& } \vec{g} = \underline{g(x,y)} = \begin{bmatrix} p(x,y) \\ q(x,y) \\ r(x,y) \end{bmatrix}$$

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

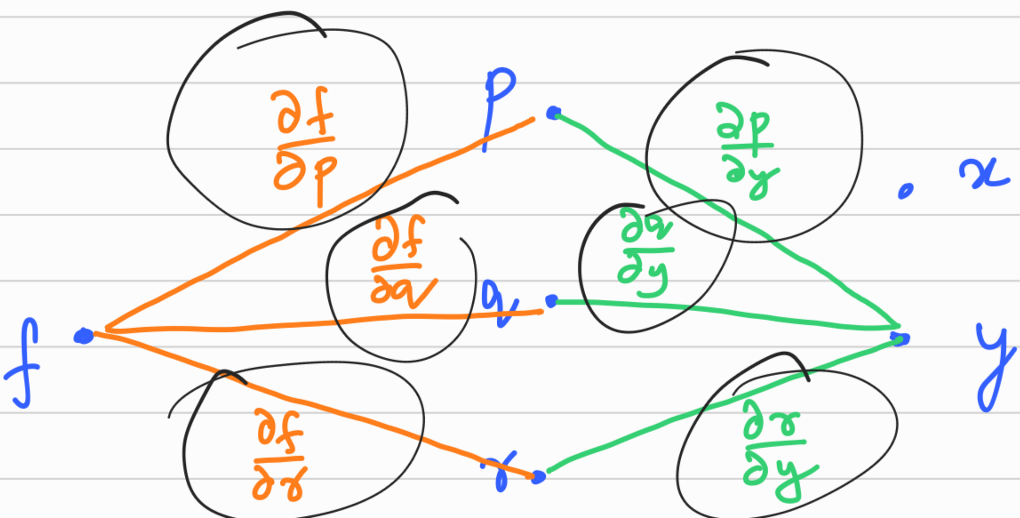
$$\frac{\partial h}{\partial y} = \frac{\partial f}{\partial p}\frac{\partial p}{\partial y} + \frac{\partial f}{\partial q}\frac{\partial q}{\partial y} + \frac{\partial f}{\partial r}\frac{\partial r}{\partial y}$$

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

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

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

$$\nabla h = \underbrace{\nabla f(p,q)}_{1\times 3} \cdot \underbrace{\nabla \vec{g}(x,y)}_{3\times 2}$$



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

(e)

$$\vec{f} : \mathbb{R}^2 \to \mathbb{R}^3, \to 3\times 2$$

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

$$3\times 3$$

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

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

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

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

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

3×3

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

3×3

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

3×2 $\qquad$ 2×3

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

3×3 $\quad$ 3×2 $\quad$ 2×3

5.5 Consider the following functions:

$$f_1(\boldsymbol{x}) = \sin(x_1)\cos(x_2)\,, \quad \boldsymbol{x} \in \mathbb{R}^2 \leftarrow 1\times 2$$

$$f_2(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{x}^\top \boldsymbol{y}\,, \quad \boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^n \quad 1\times 2n$$

$$\underline{f_3(\boldsymbol{x})} = \underline{\boldsymbol{x}\boldsymbol{x}^\top}\,, \quad \boldsymbol{x} \in \mathbb{R}^n \quad (n\times n)\times \underline{\underline{n}}$$

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

ⓐ $f_1 : \mathbb{R}^2 \to \mathbb{R}^1\,, \quad \nabla f_1 \in \mathbb{R}^{1\times 2}$

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

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

3D matrix OR
3rd order tensor

(b) $f_1(x_1, x_2) = \sin(x_1)\cos(x_2)$

$$\nabla f_1 = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \end{bmatrix}$$

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

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

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

$$= x_1 y_1 + x_2 y_2$$

$$\nabla f_2 = \begin{bmatrix} \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \frac{\partial f_2}{\partial y_1} & \frac{\partial f_2}{\partial y_2} \end{bmatrix}$$

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

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

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

$$f_3 = \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{2\times 1} \begin{bmatrix} x_1 , x_2 \end{bmatrix}_{1\times 2} = [\ ]_{2\times 2}$$

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

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

$\underline{2\times 2\times 2}$

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

Can be generalized when $x \in \mathbb{R}^n$

(we will skip this)

RESULTS TO REMEMBER

① $f(\vec{x}) = \underline{a^T x}$ $\qquad \nabla f(x) = a^T$

$f: \mathbb{R}^n \to \mathbb{R}$

$x \in \mathbb{R}^n$

gradient always a row vector

$f(x) = [a_1 \;\; a_1] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = a_1 x_1 + a_2 x_2$ $\qquad [a_1 \;\; a_2]$

$\parallel$

a^T

② $f(\vec{x}) = \underline{x^T a}$ $\qquad \nabla f(x) = \underline{a^T}$

$f: \mathbb{R}^n \to \mathbb{R}$

$x \in \mathbb{R}^n$

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

$A \in \mathbb{R}^{m \times n}$ $\qquad f: \mathbb{R}^n \to \mathbb{R}^m$

$x \in \mathbb{R}^n$

Ex: $A = \begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix}_{3\times 2}$ $\qquad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_{2\times 1}$

$$f(x) = Ax = \begin{bmatrix} ax_1 + bx_2 \\ cx_1 + dx_2 \\ ex_1 + fx_2 \end{bmatrix}_{3 \times 1}$$

$$\nabla f = \begin{bmatrix} \frac{\partial}{\partial x_1}(ax_1 + bx_2) & b \\ c & \frac{\partial}{\partial x_2}(cx_1 + dx_2) \\ e & f \end{bmatrix}_{3 \times 2}$$

$$= A$$

<u>COROLLARY</u> $\vec{f}(\vec{x}) = \vec{x} = \mathbf{I}\vec{x}$

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

$x = \mathbb{I} x$

$A = \mathbb{I}$

$Ax \rightarrow A = \mathbb{I}$

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

But vector functions are always column vectors by convention.

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

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

$$\nabla f = A^T$$

$x^T a \longrightarrow a^T$

$x^T A x = A^T$

5 $f(x) = x^T A x$ ← Quadratic Form

Let $g(x) = Ax$, $\nabla g = A$

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

$$\nabla f = g^T + x^T \nabla g$$ (product rule)

$$= (Ax)^T + x^T A$$ $(fg)' = f'.g + f.g'$ ✓

$f'g + g'f$ ✗

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

$$= x^T (A^T + A)$$ ← row vector

COROLLARY

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

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

Q3 $f(w) = \|\underbrace{Xw - y}_{z}\|_2^2 + w^T M w, \quad w \in \mathbb{R}^d$

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

$f \in \mathbb{R}^d \to \mathbb{R}^1$

Sol^n Set $\nabla f = 0$

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

$$\nabla f = \nabla \|Z\|^2 + \nabla (w^T M w)$$

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

$$= 2z^T X + w^T (M + M^T)$$

$\|z\|^2 \to 2z^T \nabla_x z$

$$= 2(Xw - y)^T X + w^T (M + M^T)$$

$$= 2(w^T X^T X - y^T X) + w^T (M + M^T)$$

$$\nabla f = 0 \Rightarrow w^T (2X^T X + M + M^T) = 2y^T X$$

$$\Rightarrow (2X^T X + M^T + M) w = 2X^T y$$

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

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

$\underline{\underline{\text{Sol}^n\ 2}}$

$$f(w) = \underline{\|Xw - y\|_2^2} + w^T M w$$

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

$$= \nabla \|Xw - y\|_2^2 + w^T (M + M^T)$$

$\|z\|^2 = z^T z$

$$\|Xw - y\|_2^2 = \underline{(Xw - y)^T (Xw - y)}$$

$$= (w^T X^T - y^T)(Xw - y)$$

$$= w^T X^T X w - w^T X^T y - y^T X w + y^T y$$

$$\Rightarrow \nabla \|Xw - y\|_2^2 = w^T (X^T X + X^T X) - y^T X - y^T X + 0$$

$$= 2 w^T X^T X - 2 y^T X$$

$$\therefore \nabla f = 2 w^T X^T X - 2 y^T X + w^T (M + M^T)$$

$$\Rightarrow \nabla f = 0 \Rightarrow 2 w^T \left(X^T X + \frac{M + M^T}{2} \right) = 2 y^T X$$

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

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

COROLLARY 1: LINEAR REGRESSION (ORDINARY LEAST SQUARES)

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

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

Here $M = 0$

$M = 0$

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

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

0

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

$x = A^{-1} y$

$Ax = y$

w^* = Solution of best parameters for the linear regression model $\hat{y} = Xw$.

Note : $\boxed{X^T X w = X^T y}$ is called the normal equation.

w^* 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 $M = \lambda \mathbb{I}$

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

$$= \|Xw - y\|_2^2 + \lambda\, w^T \mathbb{I} w$$

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

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

Regularizere to control model Complexity.

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

$$= (X^T X + \lambda \mathbb{I})^{-1} X^T y$$

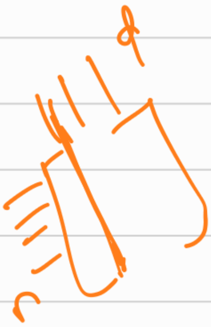
Geometric Interpretation of OLS / Linear Regression Solution

$\hat{y} = Xw$ $\qquad$ $\vec{e} = y - \hat{y} = y - Xw$

$$w^* = \arg\min \|Xw - y\|_2^2 = \|e\|^2$$

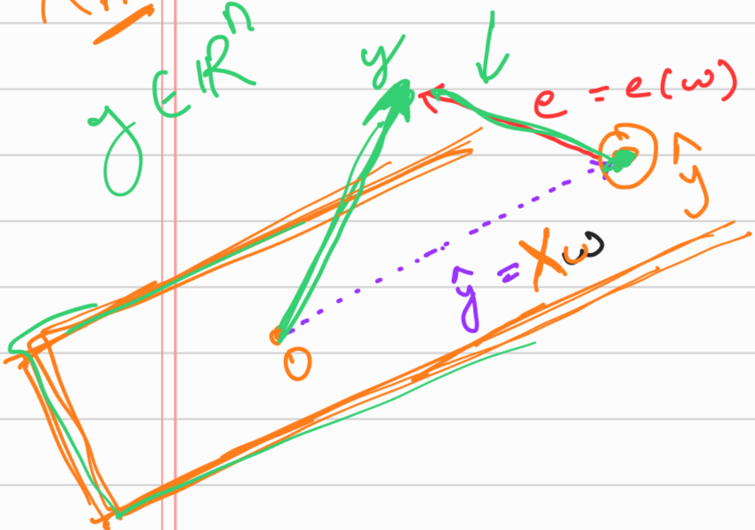
$Xw \to$

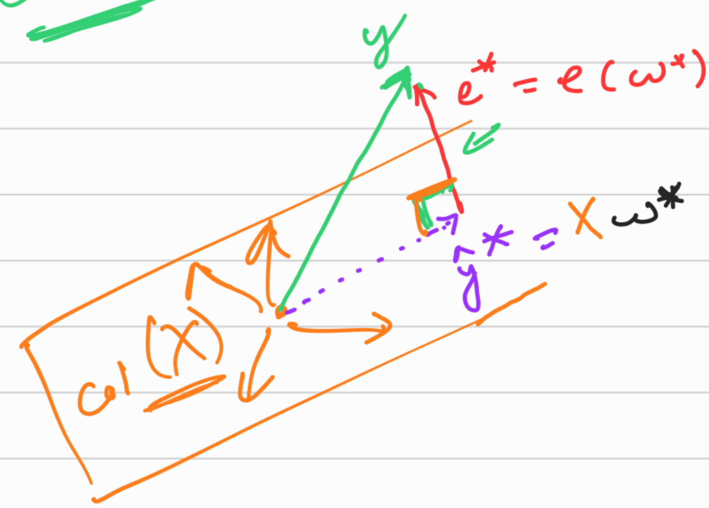
$X_{n \times d}$



$y \in \mathbb{R}^n$

$$X = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \qquad y = \begin{bmatrix} 10 \\ 11 \\ 12 \end{bmatrix}$$





Clearly, $e^* \perp \text{ColSpace}(X)$ $\qquad$ $(X^TX)^{-1}X^Ty$

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

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

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

$\Rightarrow X^Ty - X^TXw^* = 0$

$\Rightarrow \boxed{X^TXw^* = X^Ty}$ (normal eq^n)

$$\Rightarrow \quad \boxed{w^* = (X^T X)^{-1} X^T y}$$